

Modelo en series de tiempo para la tasa de penetración de un pozo de petróleo de referencia: Caso Puerto Boyacá - Colombia

Henry Daniel Hernández Martínez ¹y Diego Fernando Lemus Polanía ²

Recepción: 18-12-2014 | Aceptación: 06-04-2015 | En línea: 31-07-2015

MSC: 62F86, 62L12, 62M10, 62P30, 80M50

doi:10.17230/ingciencia.11.22.7

Resumen

En este trabajo se identificó un modelo en series de tiempo para el control de la tasa de penetración (ROP) en un pozo de referencia denominado V_{***} que pertenece al campo en desarrollo VEL que está ubicado en la cuenca del Valle del Magdalena Medio (VMM), puntualmente en el municipio de Puerto Boyacá -Colombia. Los valores ajustados por el modelo identificado y su intervalo de confianza del 95 % pueden ser empleados como guía en la perforación de pozos vecinos dentro del mismo campo en desarrollo para disminuir la incertidumbre en los tiempos de operación del proyecto. Se hace necesario hacer un análisis comparativo entre la ROP de diferentes procesos de perforación para verificar si la metodología presentada en este trabajo puede ser empleada en el campo completo.

Palabras clave: tasa de penetración; proceso de memoria larga; modelo ARFIMA; optimización

¹ Universidad Santo Tomás, Bogotá, Colombia, henryhernandezm@usantotomas.edu.co.

² Universidad Santo Tomás, Bogotá, Colombia, diego.lemus@usantotomas.edu.co.

Time Series Model for the Rate of Penetration in a Reference Oil Well: Puerto Boyacá - Colombia Case

Abstract

In this work, a time series control model was identified to describe the rate of penetration in a reference oil drilling well named V_{***} . The oil field development plan was named VEL (located in the Magdalena Medio Valley, municipality of Puerto Boyacá - Colombia). The fitted values of the identified model and its 95% confidence interval can be used to guide the wells drilling process for the same field in order to reduce uncertainty in operation times. It is necessary to make a comparative analysis between different ROP drilling processes to verify that the presented methodology can be used in the entire field.

Key words: penetration rate; long memory process; ARFIMA model; optimization

1 Introducción

La tasa de penetración (ROP) en la industria del petróleo se define como la velocidad a la cual la sarta o broca del equipo perfora la roca subyacente para profundizar al pozo en busca del yacimiento estimado [1]. Perforar un pozo de petróleo depende del tiempo de renta de los equipos usados en la operación y cualquier propuesta metodológica que optimice tiempos de operación sin comprometer la seguridad del personal, la vida útil de la broca, la estabilidad del pozo o la trayectoria al objetivo tendrá un efecto positivo en la disminución de los costos de operación [2].

Actualmente en Colombia, los equipos de perforación no son automáticos y por esto otro aspecto importante a tener en cuenta durante la perforación de un pozo de petróleo es el factor humano, el cual tiene considerable incidencia sobre la ROP. En estos equipos el perforador controla manualmente parámetros como las revoluciones y el peso sobre la broca generando una tasa de perforación de acuerdo a una prognosis con alta incertidumbre -Ver [3]-. De lo expresado en el párrafo anterior se puede deducir que un modelo de control en un pozo de referencia contribuye a que el operador tenga una guía para la perforación de pozos vecinos dentro de un campo en desarrollo.

Se considera un campo en desarrollo como el área prospectiva previamente caracterizada para la explotación de hidrocarburos, la validación del proceso de exploración geofísica se consigue a través de la perforación de varios pozos colindantes que usualmente presentan condiciones geológicas semejantes debido a que se encuentran en una misma cuenca geológica. El presente trabajo tiene como objetivo identificar un modelo de control para un pozo de referencia denominado V_{***} , ubicado en el campo en desarrollo VEL, de la cuenca del Valle Magdalena Medio (VMM) en el municipio de Puerto Boyacá - Colombia.

El documento tiene la siguiente estructura: en la segunda sección se presenta la teoría estadística considerada para determinar el modelo más adecuado para explicar la tasa de penetración (ROP) en la industria del petróleo. En la tercera sección se presentan algunas de las pruebas de raíz unitaria más tradicionales en la literatura relacionada consideradas bajo el marco de la metodología propuesta por Box-Jenkins [4] para la identificación de procesos ARIMA(p,d,q), la cual fue diseñada exclusivamente para el análisis de series temporales con órdenes de integración enteros y no fraccionales. También se presentan algunas de las pruebas de raíz fraccional estacionaria más tradicionales en la literatura relacionada bajo el marco de la metodología propuesta por [5] para la identificación de procesos ARFIMA(p,d,q) la cual fue diseñada exclusivamente para el análisis de series temporales con órdenes de integración fraccionales. En la cuarta sección se presentan los resultados del proceso de identificación del modelo de pronóstico en series de tiempo, la validación de los supuestos del modelo identificado y el ajuste del mismo a la secuencia de observaciones de interés. Finalmente se presentan las conclusiones y recomendaciones para trabajos futuros.

2 Metodología

En el presente trabajo se implementarán algunos conceptos de series de tiempo con la finalidad de ajustar un modelo de control para la tasa de penetración ROP en un pozo de referencia que pueda ser empleado por los operarios en procesos de perforación posteriores. Puntualmente se busca controlar la tasa de penetración de un proceso de perforación y así reducir tiempos de operación innecesarios en el alquiler del equipo, factor que

influye de forma directa en el costo general del proyecto. A continuación se presentan algunas definiciones que van a ser empleadas a lo largo del documento.

Según [5] y [6], la propiedad de memoria larga en una serie temporal suele entenderse como la persistencia que muestran las autocorrelaciones muestrales de ciertas series de tiempo estacionarias las cuales decrecen a un ritmo muy lento pero finalmente convergen hacia cero. [7] afirma que este comportamiento no es compatible con la función de autocorrelación (ACF) de los procesos estacionarios autorregresivos y de medias móviles $ARMA(p,q)$, que imponen un decrecimiento exponencial en las autocorrelaciones, ni con el decrecimiento casi nulo de los modelos integrados no estacionarios $ARIMA(p,d,q)$.

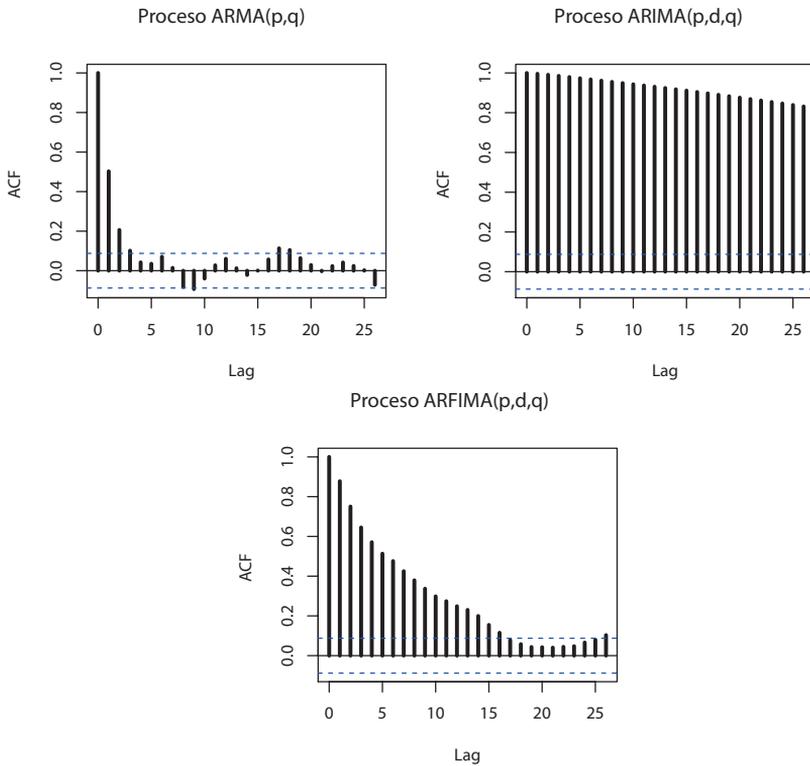


Figura 1: Comparación de las funciones de autocorrelación

En la Figura 1 se presenta la diferencia en el comportamiento de la ACF para algunos procesos ARMA(p,q), ARIMA(p,d,q) y ARFIMA(p,d,q) simulados.

2.1 Modelos ARIMA(p,d,q)

Según [8],[9],[10], un proceso estocástico $\{Z_t : t \in Z\}$ sigue un proceso autorregresivo integrado de media móvil ARIMA(p,d,q) si es una solución a la ecuación:

$$\phi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)a_t \quad (1)$$

donde $\phi_p(B)$ es el polinomio autorregresivo estacionario de orden p y $\theta_q(B)$ es el polinomio de media móvil invertible de orden q, ambos en términos del operador de retardos B y además no tienen raíces comunes entre sí. Si $d = 0$ el proceso en la ecuación (1) corresponde a un proceso ARMA(p,q) estacionario e invertible, en este caso θ_0 está relacionado con la media del proceso. Si $d \in \mathbb{Z}^+$ el proceso en la ecuación (1) corresponde a un proceso no estacionario homogéneo con d raíces unitarias, en este caso θ_0 es llamado el término de tendencia determinística del proceso. El proceso a_t es una sucesión de variables aleatorias no observables, con media cero y varianza finita σ_a^2 .

2.2 Modelos ARFIMA(p,d,q)

Según [6],[7] y [11], un proceso estocástico $\{Z_t : t \in Z\}$ sigue un proceso ARFIMA(p,d,q) si es una solución a la ecuación:

$$\phi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)a_t \quad (2)$$

En el modelo autorregresivo y de media móvil fraccionalmente integrado ARFIMA(p,d,q) el grado de memoria y de estacionaridad del proceso está definido por el parámetro de diferenciación fraccional d , el cual puede tomar cualquier valor reales. En estos modelos $(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k$ es el

operador de diferencia fraccional. θ_0 , $\phi_p(B)$, $\theta_q(B)$ y a_t tienen la misma definición que presentada en la subsección anterior.

En [12] se determinó para el proceso ARFIMA(0,d,0) que:

- Si $d > -0.5$, entonces Z_t es un proceso invertible.
- Si $d < 0.5$, entonces Z_t es un proceso estacionario.

Según [7],[6] y [13] el comportamiento asintótico de la función de autocorrelación del proceso ARFIMA(p,d,q) general es similar al obtenido para el proceso ARFIMA(0,d,0), ya que para observaciones muy distantes los efectos de los parámetros ARMA son prácticamente despreciables.

3 Pruebas de raíz unitaria y pruebas de raíz fraccional

3.1 Prueba de raíz unitaria de Dickey-Füller aumentada (ADF)

En la práctica no siempre es posible representar una serie de tiempo empleando alguno de los modelos donde no hay correlación serial en el término de error -Ver [14] -. En este caso Dickey-Fuller proponen extender la prueba considerando los siguientes modelos:

- Modelo 1: $\Delta Z_t = \gamma Z_{t-1} + \sum_{i=1}^p \delta_i \Delta Z_{t-i} + a_t$
- Modelo 2: $\Delta Z_t = \beta_0 + \gamma Z_{t-1} + \sum_{i=1}^p \delta_i \Delta Z_{t-i} + a_t$
- Modelo 3: $\Delta Z_t = \beta_0 + \beta_1 t + \gamma Z_{t-1} + \sum_{i=1}^p \delta_i \Delta Z_{t-i} + a_t$

donde a_t es un ruido blanco gaussiano de media cero y varianza constante $-a_t \sim RBN(0, \sigma_a^2)$ -. Con la finalidad de probar la existencia de una raíz unitaria se deben contrastar las siguientes hipótesis según el modelo considerado en el estudio:

- $H_0 : \gamma = 0$. Bajo esta hipótesis la serie tiene raíz unitaria, lo que implica que el proceso es no estacionario.
- $H_1 : \gamma < 0$ la serie no tiene raíz unitaria y por tanto la serie es estacionaria

El estadístico de prueba para cada una de las pruebas es $\tau = \frac{\hat{\gamma}}{se(\hat{\gamma})}$ donde $\hat{\gamma}$ es el estimador de mínimos cuadrados de γ y $se(\hat{\gamma})$ es el error estándar en el modelo bajo estudio. La distribución de τ bajo la hipótesis nula no es la t-Student tradicional, pertenece a una clase de distribuciones no estándar y se encuentra tabulada para diferentes valores de n y niveles de significancia α -Ver [14],[15] y [16] -. Si $\tau(\alpha, n)$ es el percentil α -inferior de dicha distribución se debe rechazar la hipótesis nula si el valor observado de $\tau < \tau(\alpha, n)$.

3.2 Prueba de raíz unitaria de Phillips-Perron

En [17] se propone una modificación de test de Dickey-Fuller que difiere de este en cómo tratar la correlación serial y la heterocedasticidad en los errores. Para implementar esta prueba se considera uno de los siguientes modelos

- Modelo 1: $Z_t = \gamma Z_{t-1} + u_t$
- Modelo 2: $Z_t = \beta_0 + \gamma Z_{t-1} + u_t$
- Modelo 3: $Z_t = \beta_0 + \beta_1 t + \gamma Z_{t-1} + u_t$

donde u_t es un proceso sin raíz unitaria $I(0)$ y puede ser heterocedástico. La prueba de Phillips y Perron (P-P) realiza una corrección no paramétrica de la correlación y la heterocedasticidad presentes en el término de error u_t de los modelos enunciados. Los estadísticos de prueba son:

- $Z_\rho = n(\hat{\rho}_n - 1) - \frac{1}{2} \frac{n^2 \hat{\sigma}^2}{s_n^2} (\hat{\lambda}_n^2 - \hat{\gamma}_{0,n})$
- $Z_\tau = \frac{\hat{\rho}_n - 1}{\hat{\sigma}} \sqrt{\frac{\hat{\gamma}_{0,n}}{\hat{\lambda}_n^2}} - \frac{1}{2} (\hat{\lambda}_n^2 - \hat{\gamma}_{0,n}) \frac{1}{\hat{\lambda}_n} \frac{n \hat{\sigma}}{s_n}$

Bajo la hipótesis nula $H_0 : \gamma = 0$ (equivalente a $\rho = 1$), los estadísticos de la prueba de Phillips - Perron Z_ρ y Z_τ tienen la misma distribución asintótica que el estadístico de la prueba ADF y sesgo estadístico normalizado.

3.3 Prueba de raíz fraccional estacionaria de Geweke y Porter-Hudak

El contraste semiparamétrico propuesto en [18] se basa en la inferencia sobre el estimador del parámetro de diferenciación fraccional d en la ecuación de regresión de mínimos cuadrados en el dominio de la frecuencia dada por

$$\log I_z(w_j) = c + d \log \left(4 \sin^2 \left(\frac{w_j}{2} \right) \right) + \nu_j \quad (3)$$

donde $I_z(w_j)$ es la j -ésima ordenada del periodograma de Z_t , $w_j = \frac{2\pi j}{T}$ es la j -ésima frecuencia de Fourier, ν_j es el término de error en la regresión, el cual se asume i.i.d ¹ de media cero y varianza constante $\frac{\pi}{6}$, con $j = 1, \dots, m$ siendo m el mayor entero que es menor o igual a $T^{\frac{1}{2}}$. Los autores afirman que cuando $-0.5 < d < 0$ y se cumplen algunas condiciones adicionales, entonces el estimador obtenido es consistente y asintóticamente normal

$$\frac{\hat{d}_{GPH} - d}{\sqrt{Var(\hat{d})}} \xrightarrow{d} N(0, 1) \quad (4)$$

donde $\sqrt{Var(\hat{d})} = \frac{\pi^2}{6} (\sum_{j=1}^m R_j - \bar{R})^{-1}$ con $R_j = \log(4 \sin^2(w_j/2))$.

3.4 Prueba de raíz fraccional estacionaria de Robinson

En [19] se propone una versión modificada y más eficiente del estimador d_{GPH} a partir de la siguiente regresión:

$$\log I_z(w_j) = \beta_0 + 2d \log(|w_j|) + \nu_j \quad (5)$$

donde $I_z(w_j)$ es la j -ésima ordenada del periodograma de Z_t , $w_j = \frac{2\pi j}{T}$ es la j -ésima frecuencia de Fourier, ν_j es el término de error en la regresión, el cual se asume i.i.d de media cero y varianza constante $\frac{\pi}{6}$. También determinó que el estimador obtenido es consistente y asintóticamente normal bajo la hipótesis de normalidad para el proceso ARFIMA(p,d,q)

¹independiente e idénticamente distribuido

$$\sqrt{m}(\hat{d}_{rob} - d) \xrightarrow{d} N\left(0, \frac{\pi^2}{24}\right)$$

El contraste se realiza por medio de una prueba de dos colas para determinar la significancia del coeficiente \hat{d}_{rob} en 5. Algunos desarrollos más recientes en estimación semiparamétrica se presentan en [20],[21],[22],[23] y [24].

3.5 Prueba de raíz fraccional estacionaria de Castaño, Gómez y Gallón

En [5] consideran un proceso ARFIMA(p,d,q) invertible y para cierto orden p^* suficientemente grande aproximan el comportamiento de corto plazo del modelo 2 por medio del siguiente $(1 - B)^d \pi^*(B) Z_t = a_t$ donde $\pi^*(B) = -\pi_1^* B - \pi_2^* B^2 - \dots - \pi_{p^*}^* B^{p^*}$. Bajo en el modelo ARFIMA($p^*, d, 0$) aproximado el procedimiento sugiere contrastar la hipótesis nula de memoria corta, $H_0 : d = 0$, con la alternativa de memoria larga $H_1 : d > 0$ empleando los resultados asintóticos del estimador máximo verosímil del parámetro de diferenciación fraccional.

Por lo tanto, el estadístico de prueba empleado en el contraste es $t_d = \frac{\hat{d}}{se(\hat{d})}$, donde \hat{d} y $se(\hat{d})$ son respectivamente el estimador máximo verosímil de d y la estimación de su error estándar. Los autores demostraron que bajo la hipótesis nula estadístico t_d tiene una distribución límite normal estándar ($t_d \xrightarrow{d} N(0, 1)$).

4 Pruebas y resultados

4.1 Análisis descriptivo

Según [25] los métodos de prospección geofísica permiten hacer inferencia sobre la configuración de los estratos subyacentes, la composición, forma y existencia o no de hidrocarburos bajo el manto terrestre a traspasar. Existe un grado de incertidumbre asociado con la calidad de la información registrada durante el proceso de exploración sísmica debido a las ambigüedades que se puedan presentar en su interpretación. Por ser una técnica

no invasiva no se tiene evidencia física que permita caracterizar la roca a ser perforada -Ver [26]-. No tener una perspectiva precisa del subsuelo (composición y propiedades físicas de la roca) produce una ROP con gran variabilidad entre mediciones; lo anterior puede implicar un aumento considerable en los tiempos de perforación.

Considerando lo descrito en el párrafo anterior en la presente investigación se trabajó con la tasa de penetración (ROP) obtenida en el proceso de perforación de un pozo de referencia denominado V_{***} , el cual se encuentra dentro del campo en desarrollo denominado VEL sobre la cuenca del Valle Magdalena Medio (VMM) en Puerto Boyacá, Colombia. La ROP fue medida cada 3 pies, comenzando el registro a una profundidad de 2132 pies y terminando a una profundidad de 5492 pies para un total de 1121 observaciones del proceso bajo estudio. Los gráficos y resultados que se presentan a continuación fueron obtenidos empleando la versión 3.1.3 del software estadístico *R* y sus paquetes *strucchange*, *car*, *urca*, *fracdiff*, *arfima*, *TSA*, *pracma*, *fArma*, *lmtest* y *fBasics*. En la Figura 2 se presenta la secuencia de observaciones de la tasa de perforación del pozo (V_{***}).

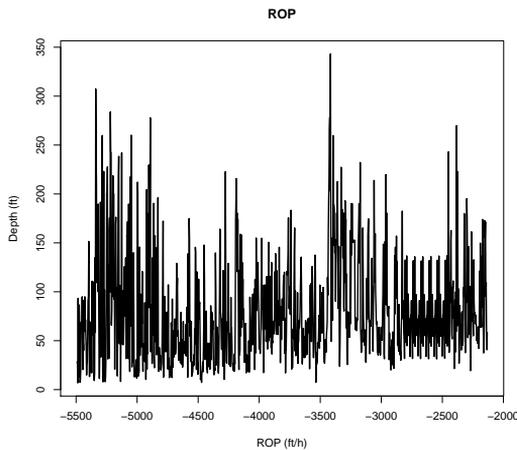


Figura 2: Serie de la tasa de perforación del pozo

En el proceso bajo estudio no se observa ninguna tendencia determinística evidente en la serie de la ROP. Por otra parte, la varianza incondicional

de la serie parece cambiar según el nivel del proceso lo que posiblemente indica que se deba realizar una transformación de potencia que permita estabilizar la varianza mencionada.

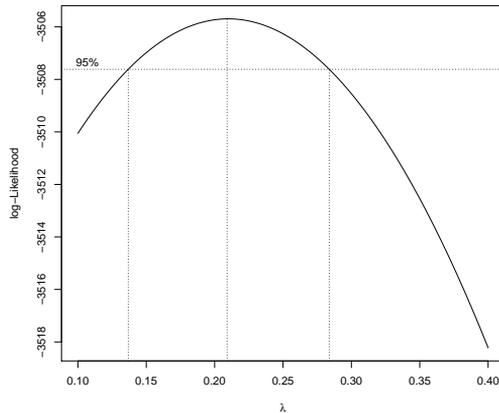


Figura 3: Curva de comparación de posibles transformaciones de potencia

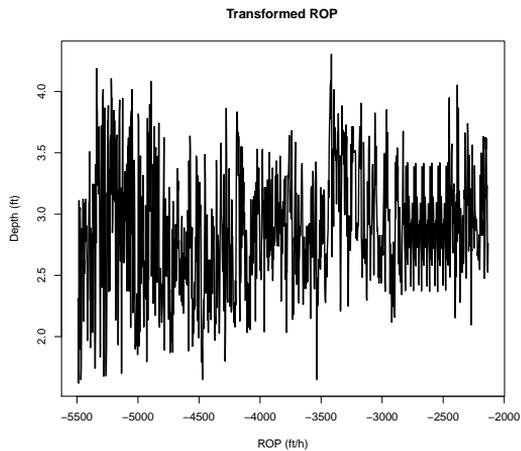


Figura 4: Serie transformada de la tasa de perforación

De la Figura 3 y de la prueba de transformación de potencia se identificó que la transformación Box-Cox más adecuada para estabilizar la varianza incondicional del proceso es la raíz cuarta. A partir de este momento se trabajará sobre la serie transformada: $X_t = \sqrt[4]{Z_t}$, la cual se presenta en la Figura 4.

En la Figura 5 se presenta la función de autocorrelación (denotada ACF) de la serie transformada $X_t = \sqrt[4]{Z_t}$

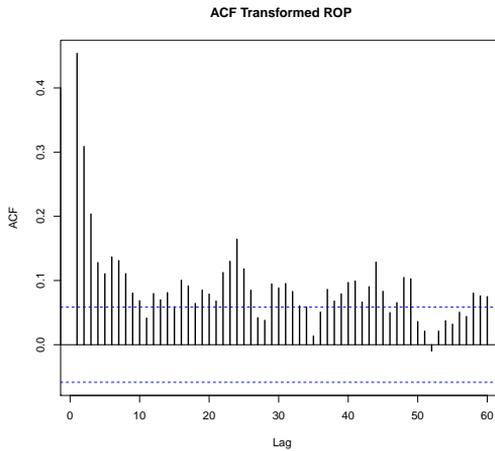


Figura 5: Correlograma de la serie transformada

En la Figura 5 no se observa el grado extremo de persistencia en las autocorrelaciones muestrales de los modelos $ARIMA(p,d,q)$ ni tampoco el patrón de decrecimiento exponencial que imponen los modelos $ARMA(p,q)$; gráficamente se evidencia una dependencia no despreciable entre observaciones separadas entre sí por una gran distancia. Lo descrito en el párrafo anterior parece indicar que la serie transformada tiene memoria larga y puede ser no estacionaria.

A continuación se determinará si el grado de interdependencia muestral del proceso corresponde al de un proceso estacionario o al de un proceso no estacionario homogéneo empleando diferentes pruebas de raíz unitaria siguiendo la metodología de Box-Jenkins para la construcción de modelos en series de tiempo univariadas [4].

4.2 Resultados de las pruebas de raíz unitaria

Al implementar la prueba ADF se consideró inicialmente la existencia de tendencia determinística y deriva. En la Tabla 1 se presentan los resultados obtenidos para esta prueba incluyendo un número máximo de retardos igual a la raíz cúbica de la longitud de la serie.

Tabla 1: Resultados prueba ADF - modelo 3

		Valores Críticos		
Estadísticos de Prueba	Valor Obtenido	1pct	5pct	10pct
tau3	-15.52	-3.96	-3.41	-3.12

En la Tabla 2 se presentan los resultados de la regresión ajustada en este caso.

Tabla 2: Regresión ajustada prueba ADF - modelo 3

Variable	Coefficiente	Error SD	Estadístico t	Valor P
Intercepto	1.473	0.09818	15.000	0.000
z.lag.1	-0.4898	0.03156	-15.520	0.000
tt	-0.0001117	0.00004	-2.752	0.006
z.diff.lag	-0.1248	0.02983	-4.182	0.000

Se puede observar que solamente el primer rezago de la prueba ADF resultó estadísticamente significativo. Igualmente se puede concluir que la tendencia determinística resulta significativa aunque el valor del coeficiente estimado para la pendiente está cercano a cero.

Considerando que la prueba anterior puede estar sobreparametrizada al incluir tendencia determinística (por su cercanía a cero), se especifica nuevamente la prueba ADF empleando el modelo 2, incluyendo un número máximo de retardos igual a la raíz cúbica de la longitud de la serie. En la Tabla 3 se presentan los resultados obtenidos.

Tabla 3: Resultados prueba ADF - modelo 2

		Valores Críticos		
Estadístico de Prueba	Valor Obtenido	1pct	5pct	10pct
tau2	-15.23	-3.43	-2.86	-2.57

En la Tabla 4 se presentan los resultados de la regresión ajustada en este caso.

Tabla 4: Regresión ajustada prueba ADF - modelo 2

Variable	Coefficiente	Error SD	Estadístico t	Valor P
Intercepto	1.3676	0.09073	15.073	0.000
z.lag.1	-0.4753	0.03121	-15.230	0.000
z.diff.lag	-0.1318	0.02981	-4.422	0.000

En la Tabla 4 se puede observar que solamente el primer rezago de la prueba aumentada resultó significativo. De la Tabla 1 y la Tabla 3 se concluye que hay evidencia suficiente para rechazar la hipótesis nula de raíz unitaria, ya que en ambas pruebas el estadístico es menor que el percentil α (nivel de significancia) inferior de la distribuciones $\tau_{\tau}(\alpha, n)$ y $\tau_{\mu}(\alpha, n)$ respectivamente, por tanto se concluye que no hay raíz unitaria en la serie de tasa de perforación del pozo V_{***} . La implementación de la prueba de raíz unitaria de Phillips-Perron conduce al mismo resultado de la prueba ADF.

Hasta el momento no existe una razón fuerte para afirmar que el proceso se comporta como un modelo ARMA(p,q) y tampoco como un modelo ARIMA (p,d,q); por un lado se debe a lo observado en la ACF y la PACF, y por otro lado los resultados de las pruebas efectuadas de raíz unitaria no tienen el suficiente peso para continuar con la metodología de Box-Jenkins. Se procede a evaluar la serie con un modelo intermedio ARFIMA(p,d,q) cuya raíz es fraccional está entre -0.5 y 0.5 utilizando la metodología propuesta en [5]. Cabe resaltar que la mayoría de los estadísticos existentes para contrastar la presencia de raíces unitarias, incluyendo los tests de Dickey-Fuller y Phillips-Perron tienen una potencia muy baja si las alternativas son de tipo fraccional -Ver [27] y [28]-.

4.3 Pruebas de raíz fraccional estacionarias

En la Tabla 5 se presentan los resultados de las pruebas de Geweke-Porter-Hudak y de Robinson empleando un ancho de banda de $\lfloor T^{1/2} \rfloor = \lfloor 33.48 \rfloor =$

Tabla 5: Pruebas de raíz fraccional estacionaria

Prueba	Valor d	Error SD	Estadístico de prueba	Valor Crítico
GPH	0.2879	0.0745	38.638	0.0003
ROB95	0.2293	0.1032	22.207	0.0157

Al observar los resultados de la Tabla 5, vemos que ambos contrastes rechazan la hipótesis nula de memoria corta ($H_0 : d = 0$), por lo cual se concluye que existe memoria larga (raíz fraccional positiva y estacionaria) en la serie transformada de la tasa de perforación del pozo de petróleo V_{***} en Puerto Boyacá-Colombia.

4.4 Procedimiento de identificación y estimación

A continuación se ilustra la implementación del procedimiento propuesto por [5] para la identificación y estimación de un modelo adecuado para la serie transformada de la tasa de perforación del pozo de petróleo V_{***} .

1. Inicialmente se ajusta el modelo aproximado ARFIMA($p^*, d, 0$) para determinar si el parámetro de diferenciación fraccional resulta significativo.

En la Tabla 6 se presentan los resultados del modelo autorregresivo aproximado de orden $p^* = T^{\frac{1}{4}} \approx 6$ ajustado empleando el paquete de R *arfima* (versión 1.2-7).

Tabla 6: Modelo ARFIMA(6,d,0) aproximado

Parámetro	Estimador	Error estándar	Estadístico de prueba	Valor p
ar(1)	0.1553	0.0806	1.9268	0.0540
ar(2)	0.0603	0.0397	1.5177	0.1291
ar(3)	-0.0030	0.0334	-0.0897	0.9285
ar(4)	-0.0534	0.0328	-1.6306	0.1030
ar(5)	-0.0316	0.0337	-0.9371	0.3487
ar(6)	0.0348	0.0337	1.0327	0.3018
d	0.2243	0.0743	3.0180	0.0025
Fitted mean	2.8770	0.0733	39.2635	<2.22e-16

En la Tabla 6 se puede observar que todos los coeficientes autorregresivos resultaron ser no significativos al 5% lo cual indica que el modelo

aproximado está sobreparametrizado. De acuerdo con lo anterior, no se requiere de ningún coeficiente autorregresivo en el modelo aproximado por lo cual se reespecificó este modelo a un ARFIMA(0,d,0) obteniendo los resultados que se presentan en la Tabla 7:

Tabla 7: Modelo ARFIMA(0,d,0) aproximado

	Estimador	Error estándar	Estadístico de prueba	Valor p
d	0.3399	0.0257	13.2225	<2.22e-16
Fitted mean	28.712	0.1606	17.8823	<2.22e-16

De la Tabla 7 y considerando un nivel de significancia del 5% se puede concluir que la estimación del parámetro de diferenciación fraccional d es estadísticamente significativa.

2. El segundo paso del proceso de identificación del modelo consiste en diferenciar la serie transformada empleando el valor del parámetro d estimado en el modelo aproximado. Por tanto se usará $d = 0.3399$ para diferenciar fraccionalmente la serie.

$$W_t = (1 - B)^{0.3399} X_t$$

Sobre la serie fraccionalmente diferenciada (W_t) se utilizan las técnicas tradicionales de identificación de un modelo ARMA(p,q) para la componente a corto plazo. Siguiendo el lineamiento de Tiao y Box (1981), Wei(2006) y Cryer (2008) para la identificación de los órdenes p y q de un modelo ARMA(p,q) se presenta el resultado de la función de autocorrelación extendida (ESACF) sobre la serie fraccionalmente diferenciada (Ver Tabla 8). En general, para un proceso ARMA(p,q) el vértice de un triángulo de ceros en la ESACF asintótica estará en la posición (p,q). En la Tabla 8 se observa el mencionado vértice en la posición (2,2) lo cual indica que proceso ARMA(2,2) es consistente para modelar el comportamiento de corto plazo de la serie transformada fraccionalmente diferenciada. Empleando los criterios de selección de modelos AIC y BIC se obtuvo el mismo modelo.

Tabla 8: EACF serie fraccionalmente diferenciada

AR/MA	0	1	2	3	4	5	6	7	8	9	10
0	0	0	0	x	x	0	0	0	0	0	x
1	x	0	0	0	x	0	0	0	0	0	x
2	x	x	0	0	0	0	0	0	0	0	0
3	x	x	x	0	0	0	0	0	0	0	0
4	x	x	x	x	0	0	0	0	0	0	0
5	x	x	x	x	x	0	0	0	0	0	0
6	x	x	0	x	x	x	0	0	0	0	0
7	x	x	0	x	x	x	x	0	0	0	0
8	x	x	x	x	0	x	x	0	0	0	0
9	x	x	x	x	0	x	x	x	0	0	0
10	x	x	x	x	0	x	0	x	x	0	0

3. El tercer paso del procedimiento consiste en estimar conjuntamente los parámetros del modelo identificado ARFIMA(2,d,2) en la serie transformada y posteriormente validar los supuestos del mismo.

En la Tabla 9 se resume el ajuste del modelo identificado y la estimación del parámetro de diferenciación fraccional d .

Tabla 9: Modelo identificado ARFIMA(2,d,2)

Parámetro	Estimador	Error estándar	Estadístico de prueba	Valor p
ar(1)	0.5150	0.1588	4.4157	0.0000
ar(2)	-0.6415	0.2199	-2.8603	0.0042
ma1	0.7012	0.1440	5.6595	0.0000
ma2	-0.6289	0.2103	-3.0498	0.0022
d	0.2537	0.0465	5.4610	0.0000
Media Ajus.	2.8765	0.0866	33.2211	0.0000

De la tabla anterior y considerando un nivel de significancia del 5% se puede concluir que la estimación del parámetro de diferenciación fraccional d y de los coeficientes del modelo de corto plazo del proceso en el modelo completo son significativos. A continuación se presenta el resultado de la prueba de Ljung-Box para los residuales del modelo identificado ARFIMA(2,d,2).

Tabla 10: Prueba de Ljung-Box para los residuales del modelo identificado

Q (Lag)	Estadístico	Valor P
Lag 10	1.6797	0.9983
Lag 15	7.3665	0.9467
Lag 20	10.8513	0.95

Los resultados presentados en la Tabla 10 permiten concluir que no hay correlación serial en los residuales del modelo ARIMA(2,d,2) identificado. Lo anterior indica que el modelo ARMA(2,2) es adecuado para capturar la correlación serial a corto plazo presente en la serie fraccionalmente diferenciada.

A continuación se presentan los resultados de algunas de las pruebas de normalidad sobre los residuales del modelo identificado.

Tabla 11: Prueba de Shapiro-Wilk

Estadístico	Valor P
W = 0.9978	0.1544

Tabla 12: Prueba de Jarque-Bera

Estadístico	Valor P
4.8788	0.0872

Tabla 13: Prueba de Kolmogorov-Smirnov

Estadístico	Valor P
0.0307	0.5279

Como el valor p de todas las pruebas consideradas es mayor al nivel de significancia nominal considerado ($\alpha = 0.05$) no se rechaza la hipótesis nula, lo que permite afirmar que los residuales del modelo identificado se ajustan a una distribución normal estándar. Los resultados obtenidos por las pruebas de correlación serial y de normalidad garantizan la validez del modelo ajustado.

4. Como resultado final se identificó el siguiente modelo ARFIMA(2,d,2) para el control de la ROP en el proceso de perforación de pozos de petróleo:

$$(1 - 0.5150B + 0.6414B^2)(1 - B)^{0.2537} \sqrt[4]{Z_t} = 2.8765 + (1 - 0.7012B + 0.6289B^2)a_t \quad (6)$$

donde B es el operador de retardos y a_t es un proceso ruido blanco gaussiano de media cero y varianza constante. En la Figura 6 se presentan los valores ajustados por modelo ajustado y su intervalo de confianza del 95% a la secuencia de observaciones de la tasa de penetración ROP del pozo V_{***} .

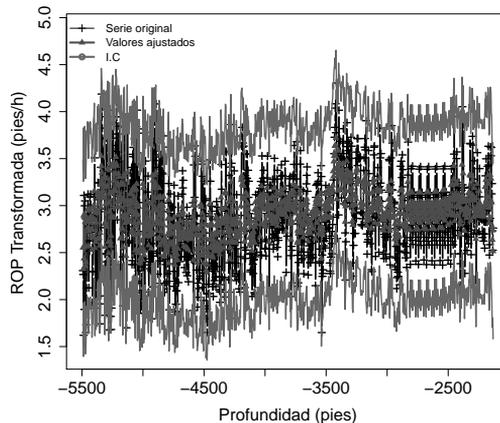


Figura 6: Ajuste de modelo ARFIMA(2,d,2) identificado

En este documento no se presentan los resultados obtenidos directamente del estudio de la serie original Z_t puesto que los residuales del modelo ajustado en este caso no se ajustan a la distribución normal y así la estimación máximo verosímil del parámetro de diferenciación fraccional y de los coeficientes del proceso ARMA(p,q) serían inapropiados.

5 Conclusiones

Se realizaron pruebas de raíz unitaria a la ROP del pozo de referencia V_{***} , obteniendo evidencia suficiente para rechazar la hipótesis nula que plantea la existencia de raíz unitaria en la serie. Se identificó que el modelo adecuado para controlar la tasa de penetración del pozo V_{***} es un ARFIMA(2,d,2) con $d = 0.2537$, lo que implica que el proceso bajo estudio es estacionario de memoria larga.

El identificar y entender el mecanismo que describe el comportamiento de la tasa de penetración en el pozo de referencia V_{***} permitirá utilizar los valores ajustados del modelo identificado y su intervalo de confianza del 95% como guía en la perforación de pozos vecinos dentro del mismo campo en desarrollo (asumiendo características geológicas y físicas semejantes entre sí) para disminuir la incertidumbre en los tiempos de operación del proyecto.

La metodología considerada para identificar un modelo de control para la tasa de penetración puede ser empleada en campos diferentes al de Puerto Boyacá solamente en un pozo de referencia. El modelo ARFIMA(2,d,2) identificado para el pozo de referencia V_{***} no puede ser empleado en un campo diferente al de Puerto Boyacá - Colombia.

Se hace necesario hacer un análisis comparativo entre la ROP de diferentes procesos de perforación para verificar si la metodología presentada en este trabajo puede ser empleada en un campo completo.

Agradecimientos

Los autores del artículo agradecen las valiosas contribuciones realizadas por los pares evaluadores asignados. De igual manera a los editores de la revista por su eficiente trabajo. Finalmente queremos agradecer a la Facultad de Estadística de la Universidad Santo Tomás por incentivar la investigación en su cuerpo docente y sus estudiantes.

Referencias

- [1] M. M. Moradi, V. Alvarado, and S. Huzurbazar, "Effect of salinity on water-in-crude oil emulsion: Evaluation through drop-size distribution proxy," *Energy and Fuels*, vol. 25, no. 1, pp. 260–268, 2011. 152
- [2] R. Caenn, H. C. H. Darley, and G. R. Gray, *Composition and Properties of Drilling and Completion Fluids*, 6th ed. 225 Wyman Street, Waltham, MA 02451, USA: Elsevier, 2011. 152
- [3] I. D. R. Bradford, W. A. Aldred, J. M. Cook, E. F. M. Elewaut, J. A. Fuller, T. G. Kristiansen, and T. R. Walsgrove, "When rock mechanics met drilling: Effective implementation of real-time wellbore stability control." Society of Petroleum Engineers. 152

-
- [4] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th ed. 111 River Street, Hoboken, NJ 07030, USA: John Wiley and Sons, 2008. 153, 162
- [5] E. Castaño, K. Gómez, and S. Gallón, “Una nueva prueba para el parámetro de diferenciación fraccional,” *Revista Colombiana de Estadística*, vol. 31, no. 1, pp. 67–84, 2008. [Online]. Available: <http://www.bdigital.unal.edu.co/30708/> 153, 154, 159, 164, 165
- [6] D. F. Lemus and E. Castaño, “Prueba de hipótesis sobre la existencia de una raíz fraccional en una serie de tiempo no estacionaria,” *Lecturas de economía*, vol. 78, no. 1, pp. 151–184, 2013. 154, 155, 156
- [7] J. Beran, *Statistics for long-memory processes*, 1st ed. One Penn Plaza New York, NY 10119: Chapman & Hall, 1994. 154, 155, 156
- [8] V. M. Guerrero, *Análisis estadístico de series de tiempo económicas*, 2nd ed. International Thomson Editores, S. A. de C. V., 2003. 155
- [9] W. W. S. Wei, *Time Series Analysis : Univariate and Multivariate Methods*, 2nd ed. Addison Wesley Pub Co Inc, 2005. 155
- [10] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. 233 Spring Street, New York, NY 10013, USA: Springer Science + Business Media, 2006. 155
- [11] J. a. Beran, “Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models,” *Journal of the Royal Statistical Society*, vol. 57, no. 4, pp. 659–672, 1995. 155
- [12] J. R. M. Hosking, “Fractional differencing,” *Biometrika*, vol. 68, no. 1, pp. 165–176, 1981. 156
- [13] W. Palma, *Long-Memory Time Series. Theory and Methods*, 1st ed. 111 River Street, Hoboken, NJ 07030, USA: Jhon Wiley and Sons, Inc., 2007. 156
- [14] D. A. Dickey and W. A. Fuller, “Distribution of the estimators for autoregressive time series with a unit root,” *Journal of the American Statistical Association*, vol. 74, no. 1, pp. 427–431, 1979. 156, 157
- [15] J. G. MacKinnon, “Critical values for cointegration tests,” 2010. 157
- [16] J. G. a. MacKinnon, “Approximate asymptotic distribution functions for unit root and cointegration tests,” *Journal of Business and Economic Statistics*, vol. 12, no. 2, pp. 167–176, 1994. 157

- [17] P. C. B. Phillips and P. Perron, "Testing for a unit root in time series regression," *Biometrika*, vol. 75, no. 2, pp. 335–346, 1988. 157
- [18] J. Geweke and S. Porter-Hudak, "The estimation and application of long-memory time series models," *Journal of Time Series Analysis*, vol. 4, no. 4, pp. 221–238, 1983. 158
- [19] P. M. Robinson, "Log-periodogram regression of time series with long range dependence," *The Annals of Statistics*, vol. 23, no. 3, pp. 1048–1072, 1995. 158
- [20] C. S. Kim and P. C. B. Phillips, "Log periodogram regression: The non-stationary case," University of Yale, Tech. Rep., 2006. 159
- [21] C. S. Kim, "Log periodogram estimation with nonstationary process," *Journal of Economic Theory and Econometrics*, vol. 19, no. 3, pp. 1–23, 2008. 159
- [22] P. C. B. Phillips and K. Shimotsu, "Pooled log-periodogram regression," *Journal of Time Series Analysis*, vol. 23, no. 1, pp. 57–93, 2002. 159
- [23] D. W. K. Andrews and P. Guggenberger, "A bias-reduced log-periodogram regression estimator for the long-memory parameter," *Econometrica*, vol. 71, no. 2, pp. 675–712, 2003. 159
- [24] C. Velasco, "Non-gaussian log-periodogram regression," *Econometric Theory*, vol. 16, no. 1, pp. 44–79, 2000. 159
- [25] M. S. Raymond and W. L. Leffler, *Oil and Gas Production in Nontechnical Language*, 1st ed. 1421 S Sheridan Rd, Tulsa, Oklahoma, Estados Unidos: PennWell Corporation, 2005. 159
- [26] C. J. Wright and R. A. Gallun, *Fundamentals of Oil and Gas Accounting*, 5th ed. 1421 S Sheridan Rd, Tulsa, Oklahoma, Estados Unidos: PennWell Corporation, 2008. 160
- [27] F. X. Diebold and G. Rudebusch, "On the power of dickey-fuller tests against fractional alternatives," *Economics Letters*, vol. 35, no. 2, pp. 155–160, 1991. 164
- [28] U. Hassler and J. Wolters, "On the power of unit root tests against fractional alternatives," *Economics Letters*, vol. 45, no. 1, pp. 1–5, 1994. 164