

PREDICCIÓN DE DATOS FALTANTES EN SERIES TEMPORALES

PROPUESTA DE PREDICCIÓN DE DATOS FALTANTES EN SERIES TEMPORALES

USADAS EN EL MODELAMIENTO DE LA PRODUCCIÓN EN UN CAMPO

PETROLERO MEDIANTE TÉCNICAS GEOESTADÍSTICAS

MAIKA GAMBÚS ORDAZ

**Trabajo de Grado presentado como requisito para optar el título de Especialista en
Estadística**

Directora

EDDY JOHANNA FAJARDO ORTIZ

Doctora en Estadística

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS

ESCUELA DE MATEMÁTICAS

ESPECIALIZACIÓN EN ESTADÍSTICA

BUCARAMANGA

2020

DEDICATORIA

A Dios, mi todo, por ofrecerme en cada respiro la oportunidad de vivir,

A mi madre Angela Graciela, que desde el cielo me cuida,

A mi padre Jorge Eliécer, que desde la tierra me acompaña,

A mi compañero de vivencias, Wilson Stive por el apoyo,

A mi familia, ejemplo continuo para mi existencia.

AGRADECIMIENTOS

A la Universidad Industrial de Santander por brindarme la posibilidad de aprender bajo sus aulas.

A la Escuela de Matemáticas por ofrecer el programa de Especialización en Estadística y nutrirlo con excelentes docentes, que me permitieron conocer un poco más acerca de esta disciplina tan interesante y con gran aplicación en todas las áreas de nuestra vida.

Al profesor Germán Moreno Arenas quien me introdujo al mundo R y por quien guardo un sincero respeto y admiración. Paz siempre a su alma.

Al profesor Giampaolo Orlandoni Merli, sus clases despertaron el interés por realizar este trabajo de investigación aplicada.

A la doctora Eddy Fajardo, directora de este trabajo por su continuo acompañamiento y orientación.

A todos los profesores del programa, especialmente a su coordinadora profesora Tulia Rivera, y a la profesional Claudia Garavito.

A todos mis compañeros que semana a semana transitamos por esta experiencia de estudio y amistad; especialmente a Elda Carolina García Tobo, a quien le deseo el mayor éxito en todo lo que emprenda.

TABLA DE CONTENIDO

	Pág.
Introducción	11
CAPÍTULO 1: El problema	13
1.1. Planteamiento del problema	13
1.2. Objetivos de la investigación	15
1.2.1. Objetivo General.....	15
1.2.2. Objetivos Específicos	15
CAPÍTULO 2: Marco Referencial.....	16
2.1. Antecedentes	16
2.2. Marco Teorico	18
2.2.1. Declinación de Producción.....	18
2.2.2. Series de Tiempo.....	21
2.2.2.1. Análisis de las series de tiempo	21
2.2.2.2. Inferencia Estadística	22
2.2.2.3. Modelos de Series de Tiempo.....	24
2.2.2.4. Criterios de selección de modelos ARIMA.....	26
2.2.2.5. Función de autocovarianza.....	27
2.2.2.6. Función de autocorrelación simple	27
2.2.2.7. Función de autocorrelación parcial.....	29
2.2.3. Metodología Box-Jenkins-Reinsel	30
2.2.4. Geoestadística	30
2.2.4.1. Variograma	32

2.2.4.2. Criterio de selección para modelos Bayesianos	33
2.2.4.3. Kriging	34
CAPÍTULO 3: Marco Metodológico.....	36
3.1. Tipo de investigación	36
3.2 Alcance de la investigación..	36
3.3 Técnica Estadística	36
3.4 Fuente de información y procesamiento de los datos..	37
CAPÍTULO 4: Análisis de los Resultados.....	39
4.1. Fase Descriptiva	39
4.2. Fase Inferencial	44
4.2.1. Identificación del Modelo	46
4.2.2. Estimación del Modelo	51
4.2.3. Evaluación y Diagnóstico del Modelo	52
4.2.4. Predicción de la serie temporal	57
4.3. Fase Predictiva	58
4.3.1. Estudio de la estructura de autocorrelación a través del variograma	58
4.3.2. Comparación predictiva entre la metodología BJR de la serie temporal y la técnica geoestadística Kriging	64
4.3.3. Predicción de los valores ausentes mediante el uso de la técnica geoestadística Kriging	68
5. Conclusiones.....	73
6. Recomendaciones	75
Referencias bibliográficas	76

LISTA DE FIGURAS

	Pág.
Figura 1. Modelos teóricos para el análisis de las curvas de declinación de producción	21
Figura 2. Correlograma simple una serie de tiempo.	29
Figura 3. Correlograma parcial de la serie de tiempo original	30
Figura 4. Fases generales del marco estadístico de la investigación	38
Figura 5. Mapa de burbujas de la producción total de petróleo por pozo para el área del campo de estudio	40
Figura 6. Histórico de producción de petróleo, gas y agua para el yacimiento del campo de estudio	41
Figura 7. Histogramas de la producción histórica de hidrocarburos total para el yacimiento del campo de estudio.....	42
Figura 8. Gráficos de densidad de las variables de producción histórica total	42
Figura 9. Descomposición de la serie de tiempo original de la producción total de aceite.	46
Figura 10. Metodología Box Jenkins y Reinsel.	46
Figura 11. Serie de tiempo original de la producción total de petróleo de un campo en el periodo 01/1996 a 07/2008.....	48
Figura 12. Correlograma simple de la serie de tiempo original.....	48
Figura 13. Correlograma parcial de la serie de tiempo original.	49
Figura 14. Serie de tiempo transformada de la producción total de aceite.	50
Figura 15. Correlograma simple de la primera diferencia de la serie de tiempo original.....	51
Figura 16. Correlograma parcial de la primera diferencia de la serie de tiempo	51

Figura 17. Tdisplay de la serie de tiempo de la producción total de aceite con la primera diferencia de la parte regular.....	52
Figura 18. Análisis de los residuales del modelo SARIMA (0,1,1) (1,0,0) ₁₂ empleado para la serie de tiempo de la producción total de petróleo.	54
Figura 19. Análisis del periodograma de los residuales del modelo SARIMA (0,1,2) (0,0,1) ₁₂ . ..	55
Figura 20. Análisis exploratorio de los residuales para el modelo SARIMA (0,1,2) (0,0,1) ₁₂ . ..	56
Figura 21. Boxplots comparativos entre (1) la serie original y (2) la variabilidad de los residuos del modelo SARIMA (0,1,1) (1,0,0) ₁₂	57
Figura 22. Predicciones de la serie de tiempo producción total de petróleo.....	58
Figura 23. Modelos de variograma para el histórico de la producción total de petróleo.....	61
Figura 24. Gráficos de Diagnóstico para el coeficiente alpha (α) del modelo teórico exponencial.	63
Figura 25. Gráficos de Diagnóstico para el coeficiente beta (β) del modelo teórico exponencial.	64
Figura 26. Gráficos de Diagnóstico para el coeficiente theta (θ) del modelo teórico exponencial.	64
Figura 27. Semivariograma exponencial definido para la predicción a futuro de la producción total de petróleo del campo de estudio (año 2008).	67
Figura 28. Predicción a futuro de la producción total de petróleo del campo de estudio para el año 2008.....	68
Figura 29. Semivariograma esférico definido para la predicción de valores faltantes de la producción total de petróleo del campo de estudio (periodo 01/01/2001 – 01/12/2002)	71

Figura 30. Predicción de los valores faltantes de la producción total de petróleo del campo de estudio para el periodo 01/01/2001 – 01/12/2002.....	72
Figura 31. Distribución de los volúmenes de producción total de petróleo anual (X) y mensual (Y) en el campo de estudio.....	72
Figura 32. Análisis exploratorio de los residuales para las predicciones realizadas mediante estimación geoestadística de la producción total de petróleo en el periodo 01/01/2001 – 01/12/2002.	73

LISTA DE TABLAS

	Pág.
Tabla 1. Estadísticas descriptivas de las variables del estudio: De izquierda a derecha petróleo, gas y agua	43
Tabla 2. Pruebas de normalidad aplicadas a las variables de la producción histórica total.....	44
Tabla 3. Comparación de la bondad de ajuste de distintos modelos predictivos.	53
Tabla 4. Características del modelo seleccionado.	53
Tabla 5. Estadísticas descriptivas de los errores residuales del modelo seleccionado para la ST diferenciada del estudio	57
Tabla 6. Coeficientes de modelos teóricos relacionados con los parámetros de variabilidad del análisis estructural de los datos.....	61
Tabla 7- Parámetros de variabilidad de acuerdo al análisis estructural de los datos	62
Tabla 8. Resumen de las desviaciones para los modelos de inferencia Bayesiana ajustados a los datos de la producción total de petróleo del campo de estudio, en el periodo 01/01/1996 – 01/07/2008.	65
Tabla 9. Resumen de las desviaciones para los modelos de inferencia Bayesiana ajustados a los datos de la producción total de petróleo del campo de estudio, en el periodo 01/01/1996 – 01/12/2007.	67
Tabla 10. Errores estimados de la predicción geoestadística vía Kriging de la producción total de petróleo en el periodo 01/01/2008 – 01/07/2008	69
Tabla 11. Resumen de las desviaciones para los modelos de inferencia Bayesiana ajustados a los datos de la producción total de petróleo del campo de estudio, con los datos faltantes del periodo 01/01/2001 – 01/12/2002.	70

RESUMEN

TITULO: Propuesta de predicción de datos faltantes en series temporales usadas en el modelamiento de la producción en un campo petrolero mediante técnicas geoestadísticas*

AUTOR: Maika Karen Gambús Ordaz**

PALABRAS CLAVES: Producción de hidrocarburos, Serie Temporal, Valores Faltantes, Predicción, Geoestadística.

DESCRIPCIÓN:

En el análisis del comportamiento de la producción de hidrocarburos, técnicas empíricas han sido aplicadas a lo largo del tiempo; donde se extrapola el comportamiento histórico de las observaciones bajo la premisa de que el pasado, presente y futuro continúan bajo una misma tendencia, no sujeta a intervenciones.

Con el fin de extraer la mayor cantidad de información de esta serie temporal, este trabajo aplicó la metodología de Box Jenkins y Reinsel, para descomponer su tendencia, en elementos adicionales como estacionalidad y aleatoriedad, una vez lograda la condición de estacionaridad. Un modelo SARIMA (0,1,1) (1,0,0)₁₂, capturó el 70.22% de la información de los datos, sin embargo, los residuos resultaron no ser normales; y el componente regular de la serie no pudo ser modelado autorregresivamente; sólo como una función lineal de sus errores actuales y previos ($p=0$ y $q=1$). Aunque las predicciones de este modelo fueron lineales, in-sesgadas y de varianza mínima, no se ajustaron con la naturaleza de la serie temporal.

Finalmente, se demostró la conveniencia del uso de técnicas geoestadísticas. La etapa predictiva comprendió desde el 01/01/2008 hasta el 01/12/2008, y considerando que se disponía de datos observados desde enero hasta julio de ese mismo año, estos sirvieron como comparación estimando un error porcentual promedio absoluto de las estimaciones (MAPE) del 5%. Adicionalmente, se usó la técnica geoestadística para predecir valores faltantes, orientando las estimaciones hacia la estructura interna de la serie. Del histórico de la producción total de petróleo fueron eliminados las observaciones de dos ciclos estacionales (01/01/2001 – 01/12/2002). Los resultados fueron satisfactorios reportando un MAPE del 7.15%; y más importante honrando la estacionalidad y la tendencia de la serie temporal. Los errores a través de esta técnica resultaron normales, idénticamente distribuidos e incorrelacionados.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Matemáticas. Especialización en Estadística. Director: Eddy Johanna Fajardo Ortíz. Doctora en Estadística.

ABSTRACT

TITULO: Proposal for predicting missing data in time series used in modeling production in an oil field using geostatistical techniques *

AUTHOR: Maika Karen Gambús Ordaz **

KEY WORDS: Production Decline, Time Series, Missing Data, Prediction, Geostatistics.

DESCRIPTION:

In the analysis of the hydrocarbon production performance, empirical techniques have been applied over time; extrapolating the historical behavior of the observations under the assumption that the past, present and future continue under the same trend, not subject to interventions.

In order to extract the greatest amount of information from this time serie, this work applied the Box Jenkins and Reinsel methodology, to decompose its trend, into additional elements, such as seasonality and randomness, once the stationarity condition has been achieved. A model representing the 70.22% of the information was defined as SARIMA(0,1,1)(1,0,0)₁₂. Its predictions were linear, unbiased, and of minimal variance, however, this model did not match the nature of the time serie, due to the fact that the regular component of the serie could not be modeled as autoregressive. In contrast, this was only modeled as a linear function of the actual and previous errors ($p = 0$ and $q = 1$), and the computed residuals were not normal.

Finally, the convenience of the use of geostatistical techniques was demonstrated. A forecast step comprised from 01/01/2008 to 12/01/2008, and considering that data was observed from January to July of that same year, these served as a comparison estimating an absolute average percentage error of the estimates (MAPE) of 5%. On the same way, this technique was applied to predict missing values from the internal structure of the time serie, along two seasonal cycles. Observations between 01/01/2001 and 12/01/2002 were removed from the total oil production historical data. Results showed a MAPE of 7.15%, not only demostrating the efficiency of the geostatistical technique but also, that the methodology honored the seasonality and trend of the time serie. Errors were normal, identically distributed, and uncorrelated.

* Degree Work

** Faculty of Sciences. School of Mathematics. Specialization in Statistics. Director: Eddy Johanna Fajardo Ortíz. PhD in Statistics.

INTRODUCCIÓN

El petróleo convencional se acumula a través de largos procesos geológicos en formaciones subterráneas conocidas como yacimientos. Estos depósitos consisten en rocas porosas, como areniscas o carbonatos, donde el petróleo reside en los pequeños espacios vacíos entre los granos de roca, que denominamos porosidad. Un campo petrolero puede consistir de uno o varios yacimientos accesibles desde la superficie mediante la perforación de pozos, y un campo maduro es aquel que se ha desarrollado al punto máximo de producción y se encuentra actualmente declinando debido al agotamiento de su energía natural.

Actualmente, la mayoría de los campos se encuentran en su etapa de producción madura, lo que se traduce en una rápida pérdida de energía y una acelerada e inevitable declinación de producción de fluidos (Handayani y Simamora, 2012). El conocimiento del comportamiento productivo del yacimiento permite pronosticar a futuro las reservas remanentes de hidrocarburos en el subsuelo. En este sentido, diferentes metodologías han sido empleadas por investigadores, entre estas, la metodología de Box, Jenkins y Reinsel (2008) que permite considerar el modelamiento por regresión del historial de producción de un campo, al considerarlo como una serie temporal.

No obstante, los modelos se construyen con base a la información disponible y la omisión de cualquier dato de entrada, genera incertidumbre en los resultados de sus predicciones.

Para los campos maduros, existe una gran cantidad de mediciones históricas de producción de petróleo, gas y agua; sin embargo, existe la posibilidad de que sus registros sean inexactos o no dispersos, grabados en frecuencias irregulares, o simplemente estén extraviados (se resalta que en Colombia la explotación de hidrocarburos posee más de cien años). Como lo refiere Montes (2010), la presencia de valores irregulares es normal en cualquier proceso productivo y para el

caso de fallas en las mediciones de la producción de hidrocarburos, se pueden destacar: cortes eléctricos, paradas programadas de producción o, simplemente, la no medición de la producción en la fecha establecida.

Debido a que es imposible regresar en el tiempo para disponer de las mediciones perdidas (i.e., producción temprana, datos de inyección, propiedades iniciales de los fluidos, composiciones producidas, entre otros.); técnicas especiales se hacen necesarias para determinar los valores faltantes (Blaskovich, 2000). Entre estas técnicas se encuentra la imputación de datos, pero como lo refieren Medina y Galván (2007), aplicar métodos de imputación inapropiados puede generar más problemas de los que resuelve y en su trabajo concluyeron que no existe el método de imputación ideal.

El presente estudio planteó la aplicación de la geoestadística temporal como propuesta alternativa para el análisis de la declinación de producción de un campo petrolero, ante la ausencia de información en algunos periodos de tiempo, que pudieron ser predichos utilizando la técnica definida como Kriging. Esto, aunado a la aplicación de la metodología de Box, Jenkins y Reinsel (2008) para series temporales permitió el análisis exhaustivo de la información, reducir la incertidumbre de las proyecciones futuras y evaluar la contribución de la geoestadística sobre el caso de estudio.

Así, el primer capítulo del documento establece la generalidad del problema que se aborda, enunciando los objetivos que se trazaron a lo largo del desarrollo de la investigación. En el segundo capítulo se define el marco teórico conceptual, resaltando los antecedentes de la problemática planteada, los conceptos, definiciones y el desarrollo de las metodologías implementadas. El tercer capítulo se establece el marco metodológico de la investigación, donde se describen las fases y técnicas utilizadas. Finalmente, el cuarto capítulo presenta el análisis de los resultados obtenidos y define el aporte de la propuesta de trabajo desarrollada.

CAPITULO 1: EL PROBLEMA

1.1. Planteamiento del problema

Ante la acelerada e inevitable declinación de producción de fluidos, la industria de los hidrocarburos se interesa por modelar el comportamiento productivo de los yacimientos, con el objeto de predecir los volúmenes remanentes en el subsuelo. Los modelos se construyen con base a información y en el caso de algunos campos petroleros maduros, existe una gran cantidad de mediciones históricas de producción de petróleo inexactas o, que no han sido registradas en la frecuencia requerida, o simplemente se han extraviado. Debido a que no se puede regresar en el tiempo para disponer de las mediciones perdidas, técnicas especiales se hacen necesarias para determinar los valores faltantes (Blaskovich, 2000).

Entre estas técnicas se encuentra la imputación de datos, pero como lo refieren Medina y Galván (2007), aplicar métodos de imputación inapropiados puede generar más problemas de los que resuelve. Estos investigadores demostraron que no existe el método de imputación ideal.

Otras técnicas empíricas hacen uso de la extrapolación bajo el supuesto de que el pasado, presente y futuro de los datos observados continúan bajo una misma tendencia, no sujeta a intervenciones.

Ante la necesidad de modelar los datos históricos de la producción total de petróleo de un yacimiento, el presente trabajo planteó la aplicación de la metodología de series de tiempo buscando extraer de las observaciones, elementos adicionales (tendencia, estacionalidad, aleatoriedad, estacionaridad). Sin embargo, el modelo predictivo que se genera a partir de esta metodología no se considera una solución al problema planteado, ya que esta analiza el comportamiento de los datos observados en el tiempo; y considerando la premisa del problema

acerca de los valores ausentes, se encuentran limitaciones de predicción en su modelo. De esta manera, ante la ausencia de información en algunos periodos de tiempo, se propone la conveniencia del uso de técnicas geoestadísticas, área de la ingeniería de petróleos muy poco explorada en el dominio temporal como alternativa para la predicción de datos faltantes en la serie temporal de la declinación de producción de fluidos en el yacimiento de un campo petrolero.

1.2. Objetivos de la Investigación

1.2.1. Objetivo General

Proponer un método para predecir datos faltantes en series temporales usadas en el modelamiento de la producción en un campo petrolero a través de técnicas geoestadísticas.

1.2.2. Objetivos Específicos

Identificar la naturaleza del fenómeno representado por la secuencia de observaciones

Describir la serie de tiempo relacionada con la declinación de la producción de aceite de un campo petrolero

Estimar los modelos compatibles con la estructura de los datos

Aplicar técnicas geoestadísticas para predecir datos faltantes en la serie temporal usada en el modelamiento de la declinación de la producción de aceite en un campo petrolero.

CAPITULO 2: MARCO REFERENCIAL

2.1. Antecedentes

Originalmente, el modelamiento de la declinación de la producción de aceite y/o gas, fué considerado una práctica completamente empírica. Fetkovich (1994) y otros han colocado al análisis de curvas de declinación en una fundación semi-teórica. El análisis de la declinación de la producción permite predecir cuanta tasa de aceite, agua y gas se obtendrá en el futuro, con el fin de estimar reservas remanentes y planificar operaciones a largo-plazo.

Considerando la importancia que el modelamiento de la declinación de la producción de aceite y/o gas tiene en la estimación de reservas remanentes de hidrocarburos en el subsuelo, diferentes metodologías han sido estudiadas por investigadores y propuestas han surgido para realizar esta proyección. Por ejemplo, en el trabajo presentado por Montes, Calvete, y Mantilla (2016) se aplicó la metodología de series de tiempo en la realización de pronósticos de producción, considerando diferentes tendencias en la declinación de producción a las ya tradicionalmente conocidas. Esta metodología propuesta por Box, Jenkins y Reinsel (2008) marcó un hito en el análisis de las series temporales, al presentar una propuesta unificada para estudiar series estacionarias y no estacionarias, estacionales o no, y aplicar estos modelos en la práctica (Peña, 2010).

Sin embargo, en las mediciones de producción de aceite, agua y gas de los campos petroleros maduros frecuentemente existe una gran cantidad de datos históricos faltantes. Esto, aumenta la incertidumbre en los resultados que las predicciones de los modelos arrojan; y genera decisiones erróneas en cuanto al volumen real de las reservas remanentes de hidrocarburos en los yacimientos.

La imputación de datos es una técnica que usualmente se utiliza para determinar los valores faltantes (Blaskovich, 2000); sin embargo, tal y como lo refieren Medina y Galván (2007), aplicar

métodos de imputación inapropiados puede generar más problemas de los que resuelve y en su trabajo concluyeron que no existe el método de imputación ideal.

En este sentido, De Iaco, Palma, y Donato (2013) mostraron como a pesar de que el análisis exploratorio de datos y la predicción en el modelado de series temporales no son típicamente basado en técnicas geoestadísticas, estas técnicas pueden ser convenientes. Para resaltar el papel del variograma para el modelado y la predicción, la interpolación de valores perdidos, predicción temporal, estimación no paramétrica y sus problemas computacionales, fueron considerados en un extenso estudio sobre una serie temporal ambiental (Baca, 2016).

El estudio planteó el análisis y la predicción de la serie temporal calidad de aire usando el método geoestadístico en la región del Valle de Aburrá en Colombia. Los resultados satisfactorios presentados en este estudio local, motivaron el uso de semivariogramas y de la técnica de Kriging para la estimación de datos faltantes en el análisis de la declinación de producción de hidrocarburos.

Otro caso de estudio donde la metodología propuesta por De Iaco, Palma, y Donato (2013) fue aplicada, en combinación con la metodología de Box, Jenkins y Reinsel (2008) para el análisis de series temporales, lo representa el trabajo de Giraldo, Pacheco y Orozco (2017). Estos investigadores mostraron un procedimiento basado en el modelamiento del variograma y el estudio de la estructura de autocorrelación de series temporales autorregresivas.

De esta manera, la revisión bibliográfica orientada hacia el desarrollo de la presente investigación, permitió conocer la viabilidad teórica de plantear la estimación de modelos de series de tiempo para analizar la declinación de la producción de aceite de un yacimiento ubicado en un campo petrolero colombiano, a partir de sus datos históricos. Asimismo, la ausencia de información en algunos periodos de tiempo, que podría conducir al modelamiento inadecuado del

comportamiento histórico de producción del yacimiento representó el problema a solucionar en este trabajo.

2.2. Marco Teórico

El análisis de las curvas de declinación de la producción, balance de materiales y simulación numérica son metodologías que se usan para predecir el comportamiento de yacimientos. Sin embargo, estos métodos tienen fortalezas y limitaciones muy particulares. Mientras los primeros están basados en teorías y observaciones analíticas; el último es un proceso puramente estocástico (Olominu y Sulaimon, 2014).

2.2.1. Declinación de la Producción

El análisis de declinación de la producción de pozos se establece como una estrategia para la estimación de reservas remanentes de hidrocarburos en un yacimiento convencional. Esta estrategia comúnmente se desarrolla a partir de curvas o modelos de declinación.

Una curva de declinación tiene dos parámetros que la describe. El primero, la tasa de declinación (D), que refleja el ritmo al cual disminuye la producción de un pozo en términos de porcentaje diario, mensual o anual, y que se calcula mediante la ecuación (1), donde q corresponde a la tasa de producción del pozo medida en el tiempo t .

$$D = -\frac{1}{q} \frac{dq}{dt} \quad (1)$$

El segundo de estos parámetros está representado por el exponente de declinación (b) que se ve reflejado en la tendencia de la curva de declinación (ver figura 1). Ambos parámetros se relacionan mediante la ecuación (2), donde D_i es la tasa de declinación inicial, y q_i es la tasa de producción al inicio de la declinación

$$D = D_i \left(\frac{q}{q_i} \right)^b \quad (2)$$

Al igualar las ecuaciones (1) y (2), se obtiene la ecuación (3):

$$D_i \left(\frac{q}{q_i} \right)^b = - \frac{1}{q} \frac{dq}{dt} \quad (3)$$

Posteriormente al separar las variables, e integrar estableciendo los limites, se obtiene la ecuación (4):

$$\frac{D_i}{q_i^b} \int_{t=0}^t dt = - \int_{q_i}^q \frac{dq}{q^{b+1}} \quad (4)$$

Al evaluar la integración en (4), se obtiene la ecuación (5):

$$\frac{D_i t}{q_i^b} = \frac{1}{b} \left(\frac{1}{q^b} - \frac{1}{q_i^b} \right) \quad (5)$$

Finalmente, de la ecuación (5) se resuelve en función de (q) , resultando la ecuación (6):

$$q(t) = \frac{q_i}{(1 + b D_i t)^{1/b}} \quad (6)$$

La ecuación (6) describe la tasa de producción (q) como una función del tiempo (t) . Existen tres modelos teóricos de curvas de declinación: Hiperbólico, Armónico y Exponencial (Arps, 1944). Estos modelos están definidos por la ecuación (6) y los valores que asuma el exponente de declinación (b) . De esta manera, si $b=1$, la expresión para calcular la tasa de producción en función de su modelo de declinación armónico se transforma en la ecuación (7):

$$q(t) = \frac{q_i}{(1 + D_i t)} \quad (7)$$

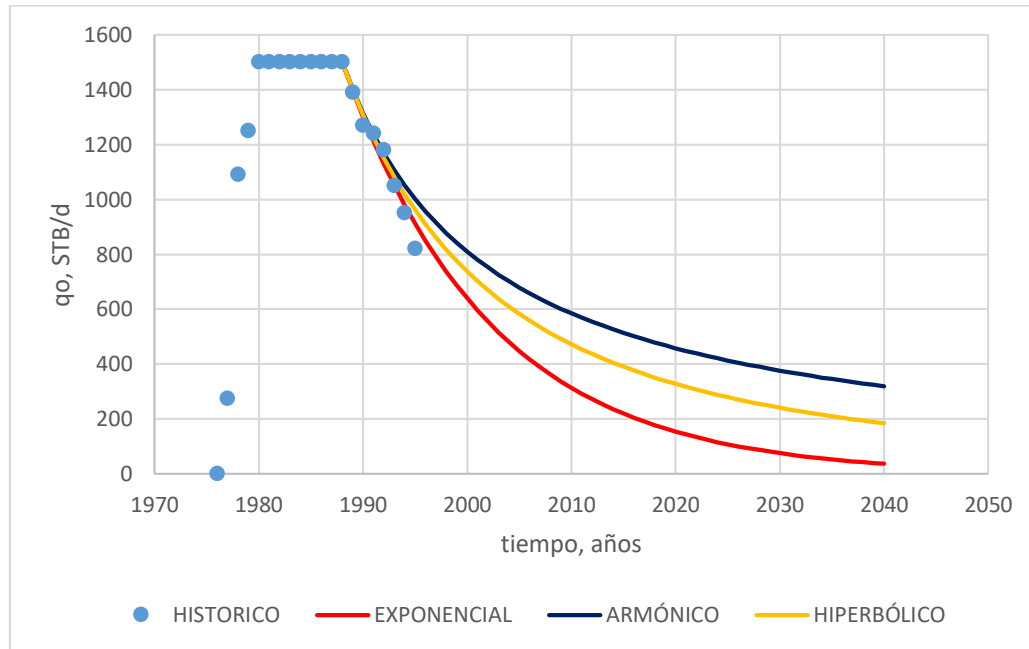
Si $b=0$, la tasa de producción corresponde al modelo exponencial, y se expresa en los términos de la ecuación (8):

$$q(t) = q_i e^{-D_i t} \quad (8)$$

Para $b \neq 0$ y $b \neq 1$, el modelo de declinación a aplicar se llama hiperbólico y la tasa de producción (q), se obtiene a través de la ecuación (6).

Figura 1

Modelos teóricos para el análisis de las curvas de declinación de producción



La figura 1 muestra datos históricos de producción de aceite en puntos de color azul-celeste, mientras las líneas reflejan la respuesta de la tasa de producción en el tiempo usando los modelos: a) exponencial (línea roja, $b=0$); b) Hiperbólico (línea amarilla; $b \neq 0$ y $b \neq 1$) y c) Armónico (línea azul, $b=1$). Los datos mostrados corresponden a valores funcionales conocidos sólo en puntos discretos; es decir, las mediciones de tasas de producción en campo son medidas en intervalos de tiempo fijos, y como se observa en la figura para ciertos periodos de tiempo estos datos no se ven reflejados.

En la industria de los hidrocarburos, el monitoreo del desempeño del yacimiento es una de las actividades críticas en todo estudio, ya que se busca obtener el mayor factor de recobro a costos reducidos.

El análisis de la declinación de producción, a partir de los modelos o curvas referidos previamente, debe ser realizado cuando exista abundante y precisa información de las mediciones de producción en los pozos que han sido perforados y completados en el yacimiento de estudio, sobre todo cuando estos experimenten merma de fluidos o depletamiento; convirtiéndose en una limitación propia de la metodología.

2.2.2. Series de Tiempo (ST).

El análisis de los datos de campo que ha sido observado en diferentes periodos de tiempo, orienta hacia problemas de modelamiento e inferencia estadística. La correlación introducida por el muestreo medido a distintos registros de tiempo, puede restringir la aplicabilidad de métodos estadísticos tradicionales que dependen de la premisa de que las observaciones son independientes e idénticamente distribuidas. La aproximación sistemática por la cual se responde a preguntas que plantean correlaciones con el tiempo es comúnmente referida como análisis de series temporales.

2.2.2.1. Análisis de las Series de Tiempo. Una serie temporal (ST) es un conjunto de observaciones de una variable que está registrado secuencialmente en el tiempo (Peña, 2010).

Este análisis se inicia mediante representaciones gráficas que pueden revelar patrones aleatorios, tendencias, cambios de nivel, ciclos, u observaciones inusuales.

En primer lugar, se debe graficar la serie temporal (ST). Con esta representación se pueden identificar componentes adicionales a la tendencia de la serie y cualquier posible anomalía. Entre estos elementos se pueden citar:

Outliers o puntos extremos: Estos se definen como observaciones anormales del fenómeno que representa a la serie de tiempo, o también puede indicar presencia de errores de medición. Que un dato atípico represente o no el fenómeno de estudio es primordial en el análisis de la serie, ya que de no representarla se debe omitir o reemplazar por otro valor.

Tendencia: Esta representa el comportamiento predominante de la serie. Puede ser definida como el cambio en la media de la serie a lo largo de un periodo de tiempo.

Variación Estacional: Se define por los movimientos periódicos o ciclos de la serie de tiempo. La frecuencia del periodo (trimestre, mes, día, semana) es inferior a la unidad temporal (generalmente se toma el año).

Variaciones Irregulares: También llamados componentes aleatorios de una serie de tiempo, son elementos adicionales a los componentes ya nombrados y que pueden o no contener información significativa del conjunto de observaciones que representan a la serie temporal.

2.2.2.2. Inferencia Estadística. Para realizar inferencias estadísticas acerca de la estructura de un proceso estocástico con base al registro de observaciones, usualmente se considera la estacionaridad. Esta idea consiste en considerar que las leyes de la probabilidad que gobiernan el proceso no cambian en el tiempo (Cryer y Chan, 2008).

Box, Jenkins y Reinsel (2008) propusieron una metodología que planteó el desarrollo de modelos sistemáticos integrados llamados autorregresivos de promedio móvil (ARIMA) para trabajar con modelamiento y predicción correlacionado con el tiempo. El método ARIMA es

apropiado solamente cuando las series de tiempo cumplen la condición de estacionaridad, lo que indica que la media, la varianza y la autocorrelación de la serie son aproximadamente constantes en el tiempo. También asume que los valores de los parámetros estimados son constantes a lo largo de la serie.

Una de las estrategias para verificar el cumplimiento de la condición de estacionaridad de la ST, es evaluar su tendencia. Así, se busca visualizar que el proceso tiene comportamiento estacionario alrededor de la tendencia.

$$x_t = \mu_t + y_t \quad (9)$$

En la ecuación (9), x_t son las observaciones; μ_t representa la tendencia y y_t es un proceso estacionario.

El proceso es llamado estrictamente estacionario si la distribución conjunta de $x_{t_1}, x_{t_2}, \dots, x_{t_n}$ es la misma distribución conjunta de $x_{t_1-h}, x_{t_2-h}, \dots, x_{t_n-h}$, para todos los puntos de tiempo $t_1, t_2, \dots, t_n$ y todos los retardos h .

A menudo, tendencias fuertes limitan encontrar el proceso estacionario; así en ocasiones remover la tendencia es el primer paso en el análisis exploratorio de la serie temporal. Otra de las estrategias para hacer que datos no estacionarios aparezcan como estacionarios es la diferenciación.

La ecuación (10) expresa matemáticamente la diferenciación de primer orden de una ST de x_t observaciones.

$$\nabla x_t = x_t - x_{t-1} \quad (10)$$

La ecuación (11) expresa matemáticamente la diferenciación de segundo orden de una ST de x_t observaciones.

$$\nabla^2 x_t = x_t - 2x_{t-1} + x_{t-2} \quad (11)$$

La diferenciación no requiere la estimación de parámetros, sólo implica remover valores previos de la ST, de manera de generar que los datos sean más estacionarios. El orden de la diferenciación corresponde al número de veces que la estrategia debe aplicarse para cumplir la condición de estacionaridad de la serie.

2.2.2.3. Modelos de Series de Tiempo. Los modelos lineales de ST se caracterizan porque tratan de explicar la evolución de las series en función de sus propios valores históricos y de sumas ponderadas de perturbaciones aleatorias actuales y pasadas. Los principales modelos lineales son el autorregresivo, el de promedios móviles y el modelo mixto (Peña, 2010).

Modelo Autorregresivo (Autorregressive, AR por sus siglas en inglés). Una serie temporal se dice se modela autorregresivamente (AR) con orden p , si esta puede ser expresada como una función lineal de sus valores previos $(x_{t-1}, x_{t-2}, x_{t-p})$ más un error aleatorio (ϵ_t) en el periodo t . La ecuación (12) muestra un modelo general autorregresivo de orden p , $AR(p)$, con parámetros (ϕ_1, ϕ_2, ϕ_p) .

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t \quad (12)$$

Modelo Promedio Móvil (Movil Average, MA por sus siglas en inglés). Este modelo describe una serie temporal como una función lineal de los errores actuales y previos $(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-q})$, también llamados residuales o un proceso ruido blanco. Un ejemplo importante de un proceso estacionario lo constituye el proceso de ruido blanco, el cual está definido como una secuencia de variables aleatorias independientes e idénticamente distribuidas.

La ecuación (13) muestra una serie de tiempo x_t que sigue un modelo promedio móvil de orden q , MA(q), con parámetros $(\theta_1, \theta_2, \theta_q)$.

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (13)$$

Modelo Autorregresivo de Promedio Móvil (ARMA por sus siglas en inglés). Estos modelos contienen los procesos anteriores combinados entre sí. La serie debe ser estacionaria, esto quiere decir, que los valores esperados de la serie y su función de autocovarianza son independientes del tiempo. Adicionalmente, la serie debe tener una varianza estable y media constante. De esta manera, se dice que una serie temporal sigue un modelo ARMA(p, q) de ordenes p y q ; si se cumple la ecuación (14):

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \cdots - \phi_p x_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (14)$$

Modelo Autorregresivo Integrado de Promedio Móvil (ARIMA por sus siglas en inglés). La mayoría de las series temporales no son estacionarias, pero algunas pueden ser transformadas a series estacionarias mediante la diferenciación. Este proceso es usado para remover tendencias, estacionalidad, y variaciones periódicas de las series. De esta manera, la serie temporal diferenciada puede ser modelada como cualquier otra serie estacionaria y posteriormente su resultado transformado a la serie original; al revertir el orden de la diferenciación.

Un modelo ARIMA predice un valor en respuesta a modelar la serie de tiempo diferenciada como una combinación lineal de sus valores previos, errores pasados y valores y errores actuales. Estos modelos en presencia de un componente periódico, son llamados Modelos Estacionales Autorregresivos Integrados de Promedio Móvil (SARIMA, por sus siglas en inglés).

El orden de un modelo ARIMA se denota como ARIMA(p,d,q); donde p es el orden del componente autorregresivo, d es el orden de diferenciación y q es el orden del proceso de promedio móvil. Así, los modelos presentados anteriormente son consecuencia de los modelos ARIMA.

La ecuación (15) muestra la expresión matemática de un modelo ARIMA

$$w_t = \mu + \frac{\theta(B)}{\phi(B)} \varepsilon_t \quad (15)$$

Donde,

w_t , es la serie de tiempo diferenciada

μ , una constante o intercepto

B , es el operador regresivo (i.e. $Bx_t = x_{t-1}$)

$\phi(B)$, es el operador autorregresivo, representado como un polinomio en el operador regresivo:

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B_p \quad (16)$$

$\theta(B)$, es el operador de promedio móvil, representado como un polinomio en el operador regresivo:

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B_q \quad (17)$$

ε_t , es el error aleatorio o ruido blanco

2.2.2.4. Criterios de selección de modelos ARIMA. Entre los criterios para seleccionar modelos lineales de series temporales se encuentran los criterios de información de Akaike y Bayesiano (AIC y BIC; respectivamente por sus siglas en inglés). Para usar estos criterios, se estima por máxima verosimilitud un conjunto de potenciales modelos ARIMA y para cada uno de estos se calcula el AIC y BIC, respectivamente.

Los parámetros AIC y BIC dependen la suma de errores cuadrados medios entre las observaciones de la serie temporal y los predictores estimados a partir de los modelos. El modelo con el menor valor AIC es el que más se adecúa a la ST.

2.2.2.5. Función de Autocovarianza (ACV). Hay dos formas equivalentes de analizar las series de tiempo: Dominio del tiempo y dominio de las frecuencias. El primer enfoque se apoya en la función de autocovarianza y autocorrelación (Peña, 2010).

La función ACV define la covarianza entre observaciones de la serie temporal, separadas por h retardos: $x_t, x_{t-h}; \forall t, h$

$$C(x_t, x_{t-h}) = E[(x_t - \mu)(x_{t-h} - \mu)] \quad (18)$$

O simplemente,

$$C(t_1, t_2) = E\{[x_{t_1} - E(x_{t_1})][x_{t_2} - E(x_{t_2})]\}, t_1, t_2 \in T \quad (19)$$

La autocovarianza muestral es un estimador sesgado de la ACV poblacional. Pero si el proceso estocástico es estacionario y ergódico, entonces la ACV muestral es asintóticamente insesgada.

2.2.2.6. Función de Autocorrelación Simple (ACF). Generalmente una serie temporal estacionaria puede describirse mediante sus primeros momentos: media aritmética, varianza, funciones de autocorrelación simple y parcial.

La función ACF define la correlación entre observaciones de la serie temporal separadas por h retardos $x_t, x_{t-h}; h = 1, 2, \dots$. Esta función permite conocer si el proceso estocástico es estacionario. Las ecuaciones (20) y (21) definen matemáticamente a esta función, mientras la representación gráfica de la ACF es un correlograma (ver figura 2).

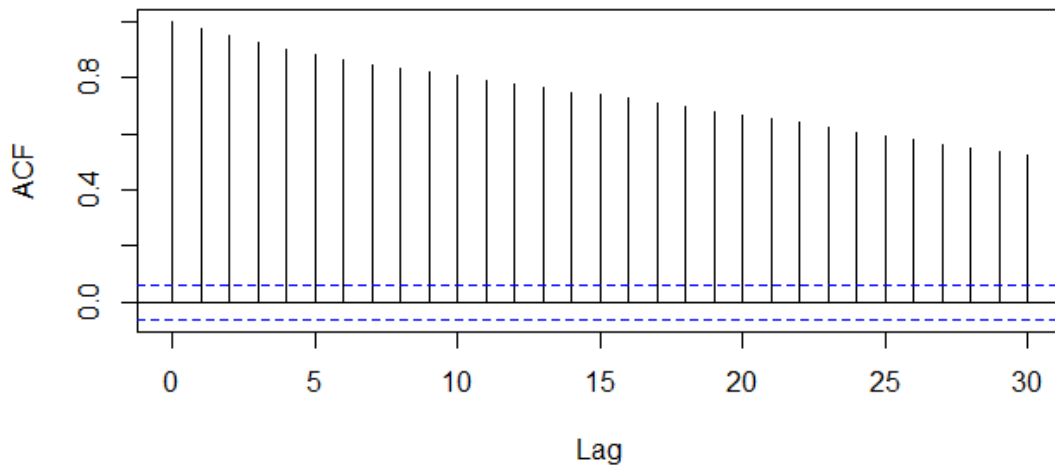
$$\rho_h = Cor(x_t, x_{t-h}) = \frac{C(x_t, x_{t-h})}{\sigma_{x_t} \sigma_{x_{t-h}}} \quad (20)$$

Esto es,

$$\rho(t_1, t_2) = \frac{C(t_1, t_2)}{\sqrt{Var(x_{t_1})Var(x_{t_2})}}, \quad t_1, t_2 \in T \quad (21)$$

Figura 2

Correlograma simple una serie de tiempo.



Un correlograma es una representación gráfica de la correlación de estadísticas. En la figura 2 se observa los valores de la función de autocorrelación simple (ACF), donde toda la información que queda por fuera de las líneas azules (intervalo de confianza del 95%) es significativo. Entre las características más resaltantes de este correlograma se pueden resumir:

- a) La serie no es estacionaria.
- b) La serie no es homogénea porque no cae el correlograma rápidamente a cero, y en ese caso la diferenciación se hace necesaria.
- c) En la medida que el lag (retardo) aumenta la correlación debe tender a cero, al no evidenciarse esto, la serie temporal muestra la no ergodicidad.

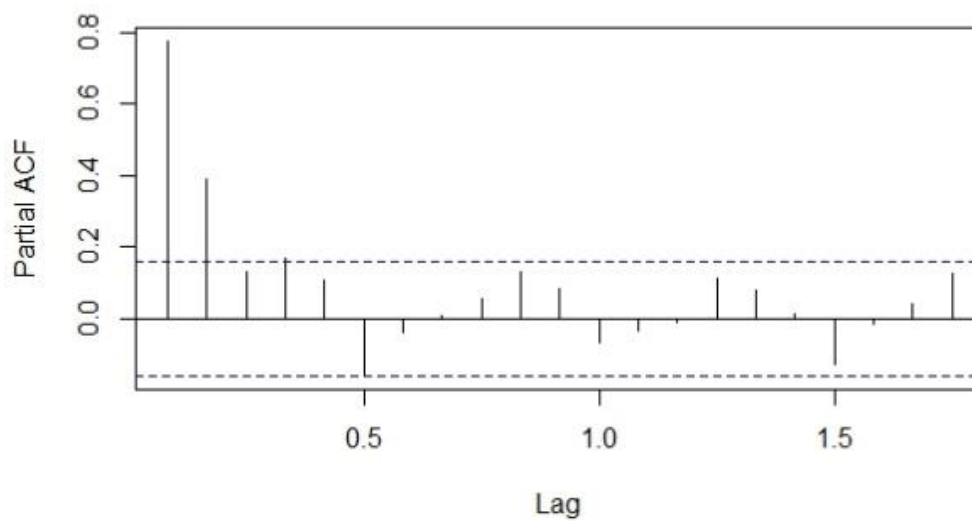
La ergodicidad requiere que las observaciones muy separadas en el tiempo estén poco correlacionadas (Peña, 2010).

2.2.2.7. Función de Autocorrelación Parcial (PACF). Mide la correlación lineal entre (x, x_{t-h}) , eliminada la dependencia lineal de observaciones intermedias $(x_{t+1}, x_{t+2}, \dots, x_{t+h-1})$. La ecuación (22) define matemáticamente a esta función como, mientras la representación gráfica de la PACF es un correlograma (ver figura 3).

$$\phi_{hh} = \text{Cor} \frac{(x_t, x_{t-h})}{(x_{t+1}, x_{t+2}, \dots, x_{t+h-1})} \quad (22)$$

Figura 3

Correlograma parcial de la serie de tiempo original



La figura 3 muestra las bandas de confianza para la PACF aproximadas por $p = \pm 1.96 \frac{1}{\sqrt{n}}$. Al contrario de la función de autocorrelación simple, que identifica la naturaleza del proceso generador de datos (i.e. AR), la función de autocorrelación parcial identifica con claridad el orden del modelo mediante el truncamiento del retardo. Por ejemplo, para la serie de tiempo visualizada en la figura 3 podría pensarse en un modelo autorregresivo de orden p igual a 2 (i.e. AR(2)).

2.2.3. Metodología Box, Jenkins y Reinsel.

Esta metodología es un procedimiento estadístico organizado que permite identificar y estimar el modelo ARIMA que mejor se ajusta a los datos históricos de la serie en estudio. El procedimiento consta de los siguientes pasos iterativos:

Identificación de la estructura del modelo: La condición de estacionaridad (regular, estacional) de la serie de tiempo es determinada a partir de los correlogramas simples y parciales; para posteriormente identificar el modelo lineal que capture la información de la serie.

Mediante el uso Máxima Verosimilitud, estimar los parámetros del modelo identificado.

Evaluación y diagnóstico del modelo ajustado: Se debe verificar la existencia de ruido blanco, buscando que en los residuos no se observe ningún patrón de comportamiento (aleatoriedad). El análisis de residuos se desarrolló a través de las funciones de autocorrelación simple y parcial de los residuos y el análisis de aleatoriedad de los residuos con la prueba Q-Ljung-Box.

Predicción. El objetivo de este paso es el de obtener predicciones lineales, insesgadas y de varianza mínima para la serie temporal; estimaciones futuras de la serie temporal.

2.2.4. Geoestadística.

En el análisis de las series temporales se observa la función de autocorrelación (ACF) y la función parcial de autocorrelación (PACF), para posteriormente inferir un modelo que permita predicciones del comportamiento futuro de estas series (Box, Jenkins, y Reinsel; 2008). Sin embargo, la geoestadística omite el paso del modelamiento; considerando al variograma como la herramienta básica para inferir y realizar estimaciones sobre series temporales estacionarias y no estacionarias (Haslett (1997), Ma (2004)).

Cressie (1993) menciona que los dos pilares básicos de la geoestadística, son el estudio de la estructura de autocorrelación de datos (espacial o temporal) a través de la función de variograma, y su uso en la predicción de los valores de la variable en sitios o periodos de tiempo no observados. Este investigador proporcionó un procedimiento gráfico para determinar la no estacionaridad en el análisis de las series de tiempo; mientras Khachatryan y Bisgaard (2009) propusieron al variograma como una herramienta gráfica para evaluar la estacionaridad.

Existen otras diferencias entre la geoestadística y la aproximación en el dominio de tiempo (series temporales) que pueden resaltarse; tales como: a) la serie temporal trabaja en una única dimensión (tiempo); mientras la geoestadística aplica a múltiples dimensiones; b) la aproximación por series de tiempo está basada en el análisis de los datos regularmente espaciados, mientras la geoestadística evita este problema al tratar cualquier proceso como continuo, modelando el variograma como una función continua. Aunque la metodología de Box, Jenkins y Reinsel (2008) hace posible modelar el proceso temporal como continuo, debe usarse una apropiada ecuación lineal diferencial (Iaco, Palma y Posa, 2013).

Para describir el proceso como estocástico, la geoestadística emplea la función continua de covarianza (ACV) o el variograma. En el análisis estructural, la estimación y modelamiento del variograma para propósitos predictivos es crucial. Similarmente al análisis espacial, el modelo del variograma puede ser usado para predicciones temporales mediante el uso de interpoladores (i.e. Kriging).

2.2.4.1 Variograma. El variograma es una herramienta gráfica para evaluar estacionaridad y periodicidad en el análisis de las series de tiempo. Este comparado con la covarianza, es una medida general de correlación.

Considere $\{x_t, t \in T\}$ un proceso estocástico² de valor real sobre el dominio de tiempo $T \subseteq \mathbb{R}$, el variograma en su expresión general puede expresarse matemáticamente por la ecuación (23):

$$\gamma(t_1, t_2) = 0.5Var[x_{t_1} - x_{t_2}], t_1, t_2 \in T \quad (23)$$

En la aproximación de dominio temporal, el ACF muestral es comúnmente la herramienta estándar exploratoria para identificar la estructura del modelo; así como en Geoestadística, la covarianza de la muestra y el variograma son las herramientas básicas usadas para el modelamiento y la predicción con la técnica Kriging. Sin embargo, el uso del variograma como estimador es más conveniente.

El variograma a diferencia de la covarianza, no requiere del conocimiento del valor esperado asociado al proceso estocástico. Es por esto, que algunos investigadores como Yaglom (1987), definen al variograma como una función de análisis estructural de los datos.

Por ejemplo, sobre estacionaridad de segundo orden, si el valor esperado es desconocido y estimado de los datos, el estimador de covarianza será sesgado; esto a pesar, de que la serie temporal disponga de mucha información (gran número de datos). En contraste, el variograma no es afectado por este problema, ya que automáticamente filtra el valor esperado y su estimador garantiza la propiedad insesgada en el proceso (Matheron, 1962).

Así, considerando un conjunto H de datos a diferentes puntos de tiempo (no necesariamente igualmente espaciados), la expansión de la forma general del variograma puede expresarse por la ecuación (24) como:

² Un proceso estocástico es un concepto matemático que permite que la naturaleza posiblemente impredecible de futuras observaciones, sean supuestas considerando que cada observación x_t es una realización de cierta variable aleatoria X_t (Brockwell y Davis, 2006).

$$\hat{\gamma}(r_t) = \frac{1}{2|N(r_t)|} \sum_{N(r_t)} [x_{t+h} - x_t]^2 \quad (24)$$

Donde r_t es el retardo temporal, $N(r_t) = \{t + h \in H \text{ y } t \in H \text{ tal que } \|r_t - h\| < \delta_t\}$, δ_t es la tolerancia, y $|N(r_t)|$ es la cardinalidad del conjunto de datos; es decir, la cantidad de valores únicos contenidos en una columna o campo particular de la base de datos.

Es de resaltar que, en la mayoría de los procesos no estacionarios, el uso del ACF, o de la covarianza puede ser erróneo, ya que una serie temporal no estacionaria está caracterizada por media y varianza no constante; y estas variables al considerar las autocorrelaciones que dependen de los puntos de tiempo pueden reflejar resultados inexactos y sesgados.

En contraste con el ACF y la covarianza, el variograma no es sólo una herramienta complementaria; sino que define bien procesos no estacionarios, incluyendo el modelamiento ARIMA (Haslett, 1997).

2.2.4.2. Criterio de Información para modelos Bayesianos. El criterio de información de la desviación (DIC), compara la media de las desviaciones posteriores, $\bar{D}$, considerada como una medida de ajuste de los modelos Bayesianos; con la complejidad del modelo, definida por el parámetro p_D , que establece el número efectivo de parámetros en el modelo.

$$DIC = \bar{D} + p_D \quad (25)$$

La inclusión de distribuciones iniciales (prior) induce una dependencia entre los parámetros que probablemente reducen la dimensionalidad efectiva, aunque el grado de la reducción depende de los datos disponibles. Spiegelhalter, Best, Carlin, y Van der Linde (2002) concluyeron que modelos que descarten información de distribuciones iniciales para las variables definidas, el criterio DIC será aproximadamente equivalente al criterio de información Akaike (AIC).

2.2.4.3. Kriging. En la predicción, el objetivo es estimar el valor desconocido x_t , al tiempo t , del proceso estocástico x , usando los datos históricos, y en el caso de interpolación, los datos observados después del tiempo t . Esto puede ser logrado en el análisis de series temporales, o sin el uso de un modelo explícito del proceso, como es el caso de la geoestadística lineal, donde sólo el conocimiento del variograma es necesario.

El krigeaje o krigeado es una técnica geoestadística de estimación de puntos, $\hat{x}_t$, que utiliza un modelo de variograma para la obtención de datos. Calcula los pesos $\lambda_i(t)$, que se darán a cada valor de las variables aleatorias x_{ti} , usados en los puntos de tiempo, t_i . Esta técnica de interpolación, expresada por la ecuación (25), se basa en la premisa de que la variación espacial y/o temporal continúa con el mismo patrón.

$$\hat{x}_t = \sum_{i=1}^n \lambda_i(t) x_{ti} \quad (26)$$

Los pesos $\lambda_i(t)$, $i=1,2, \dots, n$ en la ecuación (25), se obtienen de resolver el sistema presentado por (26); definido como Kriging Ordinario:

$$\begin{bmatrix} \gamma_{11} & \dots & \gamma_{1n} & -1 \\ \gamma_{21} & \dots & \gamma_{2n} & -1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \dots & \gamma_{nn} & -1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{bmatrix} \quad (27)$$

Donde:

γ_{ij} , es la función de variograma entre los valores de las variables aleatorias x_t , usados en los puntos de tiempo conocidos, t_i, t_j .

γ_{i0} , es la función de variograma entre los valores de las variables aleatorias x_t , usados en los puntos de tiempo conocidos, t_i y el tiempo donde se desea conocer al predictor.

μ , corresponde al multiplicador de Lagrange

El kriging ordinario es usado cuando el valor esperado del proceso es constante y desconocido. Este puede ser visto como un algoritmo para calcular el error mínimo de la varianza y se expresa en términos del variograma. El uso del kriging para estimar los valores faltantes en series temporales es conveniente por diferentes razones: a) Disponibilidad del error de la varianza en los estimados, b) Reconstrucción de los datos con respecto al espaciamiento regular o irregular, y c) Estimación del valor desconocido al tiempo t usando los datos observados antes y después del tiempo t , a pesar de que no exista un modelo explícito o la asunción de la distribución de probabilidad del proceso estocástico.

CAPÍTULO 3: MARCO METODOLÓGICO

El conocimiento puede generarse a partir de diversos procesos que implican una búsqueda o indagación por parte del investigador (Hurtado, 2000)

3.1. Tipo de Investigación

Según el propósito con el que se realiza la investigación, la propuesta se enmarcó dentro de la investigación de tipo aplicada (Hurtado, 2000). En su dimensión holística al tener como objetivo el diseñar o crear una propuesta dirigida a resolver un problema determinado sustentado en un proceso de investigación, se define dentro del tipo proyectivo (Hurtado, 2000).

3.2. Alcance de la Investigación

En este estudio se planteó la estimación de modelos de series de tiempo para analizar la declinación de la producción de un yacimiento a partir de datos históricos de un campo petrolero colombiano. Sin embargo, la ausencia de información en algunos periodos de tiempo, podría conducir al modelamiento inadecuado de su comportamiento histórico. Este hecho sugiere la posibilidad, de utilizar el análisis geoestadístico en la solución del problema propuesto.

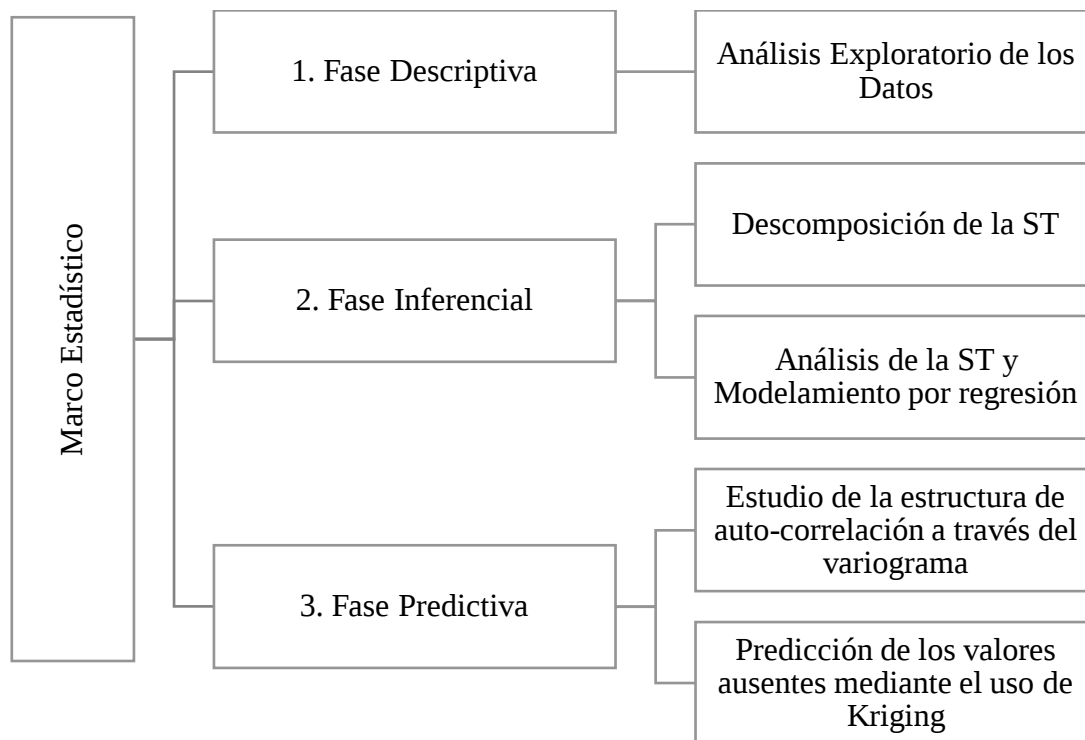
3.3. Técnica Estadística

El presente trabajo se enmarcó dentro de tres (3) fases generales: a) Fase Descriptiva donde se realiza un análisis exploratorio de los datos, b) Fase de estadística inferencial que considera el modelamiento por regresión y el análisis de series temporales (descomposición de la serie, en sus componentes: tendencia, estacionalidad y residual), y c) Fase predictiva, con el estudio de la

estructura de auto-correlación de los datos a través de la función de variograma, y su uso en la predicción de los valores de la variable en sitios o periodos de tiempo no observado mediante el uso de técnicas geoestadísticas.

Figura 4

Fases generales del marco estadístico de la investigación



La figura 4 muestra en resumen el marco estadístico del trabajo desarrollado.

3.4. Fuente de Información y procesamiento de los datos

La muestra de estudio o de entrenamiento, comprende la producción total de aceite del campo desde enero 1996 hasta julio de 2008 de veintidós (22) pozos, ubicados en un campo petrolero. La población total del campo se desconoce; sin embargo, solo se dispone de la información de la muestra. La fuente de la información es propia y consiste de un conjunto de datos abiertos

obtenidos a partir de una capacitación académica que fueron exportados a una base de datos Excel para su uso en este trabajo. Para el procesamiento de los datos se utilizó el software de acceso libre Rstudio versión 1.1.463 - © 2009 – 2018 a lo largo de toda la extensión del desarrollo de esta investigación. El paquete gstat (Pebesma, 2004) fue utilizado para la estimación de los semivariogramas teóricos y empírico. Así mismo, los paquetes Forecast y fBasics para el análisis de las series de tiempo, ggplot2 para los gráficos, sp para desarrollar coordenadas, lmtest para el análisis de las pruebas de normalidad y modelamiento de mínimos cuadrados y paswr para el análisis exploratorio de la información. De igual manera, el uso de la versión abierta OpenBugs versión 3.2.3 rev1012 - © 2010 – 2011, para especificar los modelos teóricos variográficos, que en combinación con la aplicación de la técnica de las cadenas de Markov para el método de Monte Carlo (librerías de Rstudio: arm; BRugs; R2WinBUGS, MCMCglmm y mcmcplots), permitieron conocer el ajuste de los modelos de inferencia bayesiana a través del criterio de ajuste DIC (Hadfield, 2010).

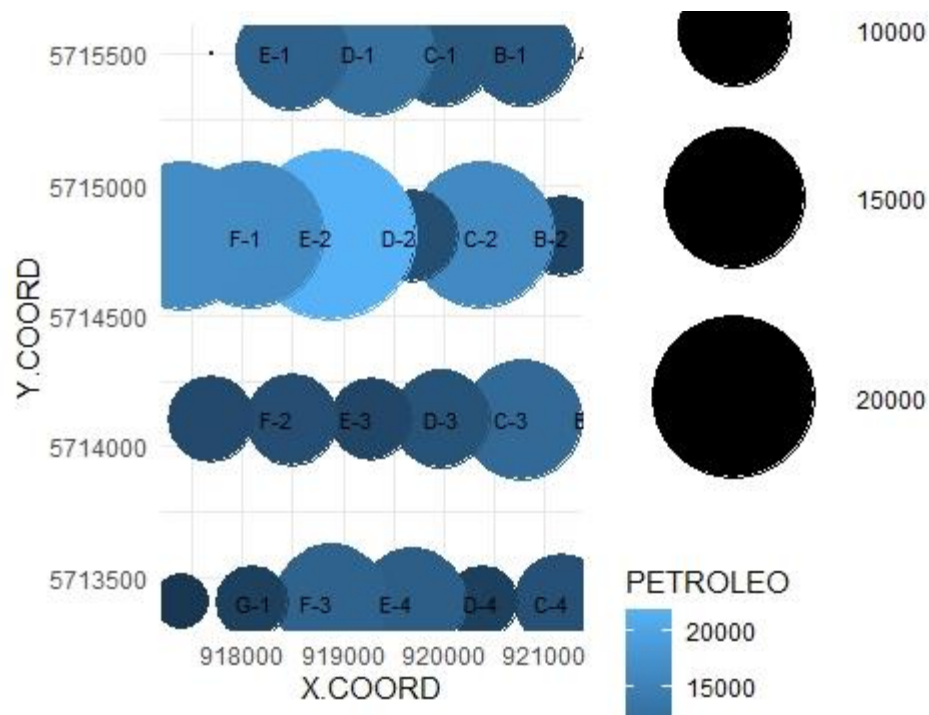
CAPÍTULO 4: ANÁLISIS DE LOS RESULTADOS

4.1. Fase Descriptiva

El conjunto de datos utilizado para la metodología propuesta consistió de variables numéricas correspondientes a las producciones de petróleo, gas y agua de veintidós (22) pozos ubicados en el yacimiento de un campo petrolero (ver figura 5)

Figura 5

Mapa de burbujas de la producción total de petróleo por pozo para el área del campo de estudio

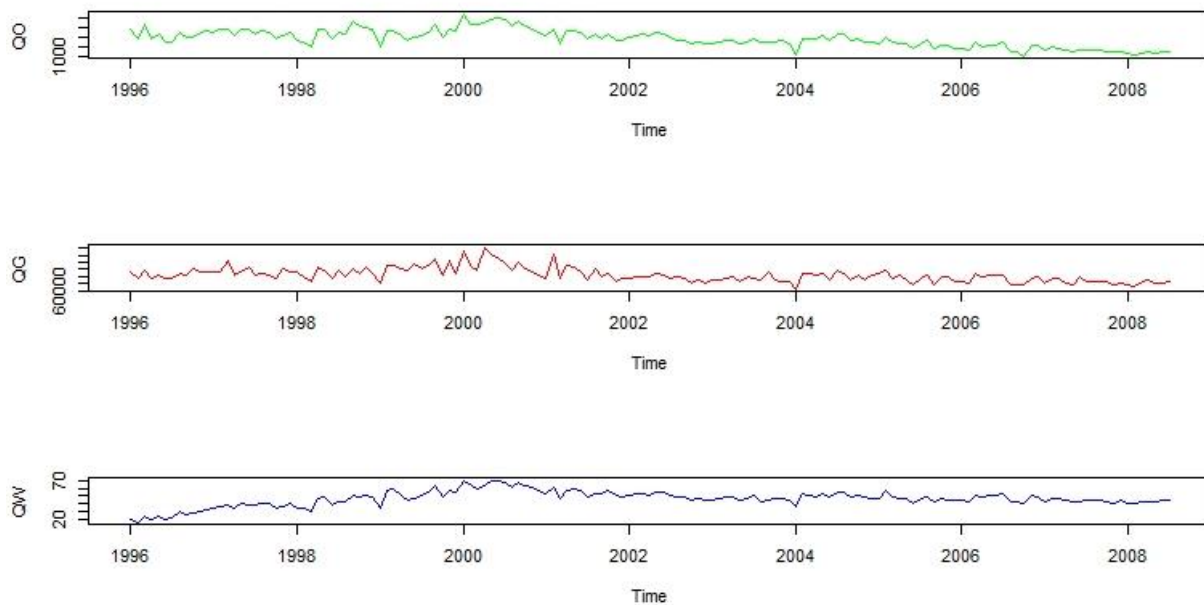


La figura 5 muestra un mapa de burbujas de la producción total de petróleo que cada uno de los pozos aporta al yacimiento del campo de estudio, reflejando que la contribución no es equitativa y existen pozos con mayores producciones en comparaciones entre estos. De esta manera, las variables correspondientes a las producciones de petróleo, gas y agua se analizaron como los volúmenes totales que el campo produce (ver figura 6).

Las producciones totales en el campo fueron graficadas para analizar el comportamiento que el yacimiento ha tenido desde el inicio de su explotación 01/01/1996 hasta la fecha hasta la cual se dispone de información (01/07/2008). Como se aprecia en la figura 6, el pico más alto de producción lo alcanzó en el año 2000, y después de ese año la declinación de su producción se hizo evidente, cerrando en Julio de 2008 con 1089 barriles de petróleo por día aproximadamente.

Figura 6

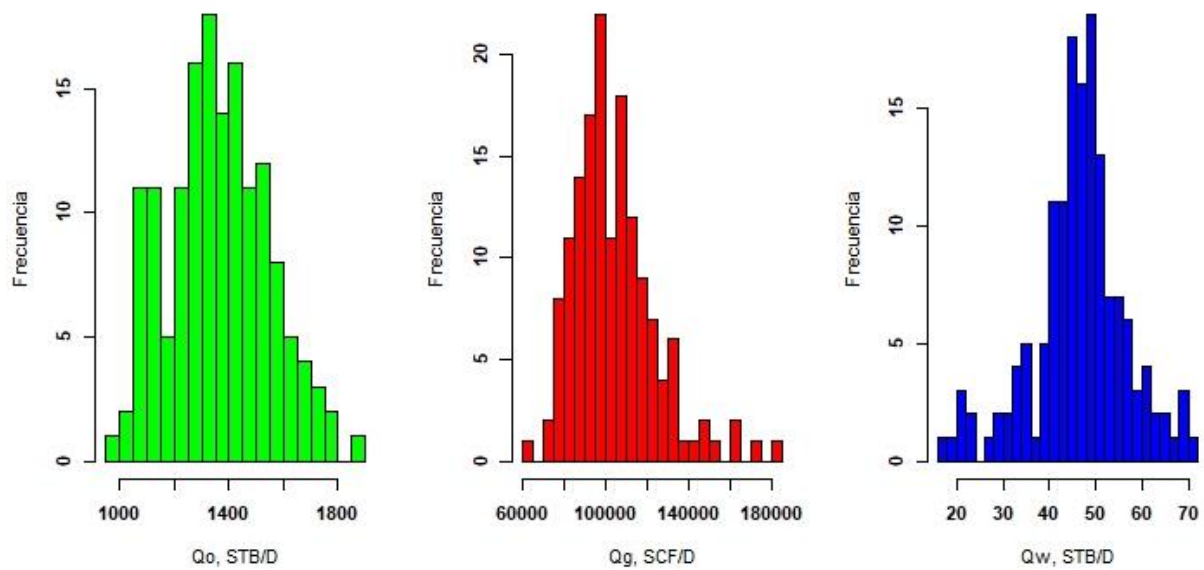
Histórico de producción de petróleo, gas y agua para el yacimiento del campo de estudio.



Para iniciar la estadística descriptiva de las variables, las figuras 7 y 8 muestran los histogramas y gráficos de densidades para las variables de escala lineal observadas en el conjunto de datos; donde visualmente la producción total de petróleo muestra un comportamiento de distribución normal y simétrico; sin embargo, las producciones totales de gas y agua reflejaron un sesgo hacia la derecha y hacia la izquierda, respectivamente.

Figura 7

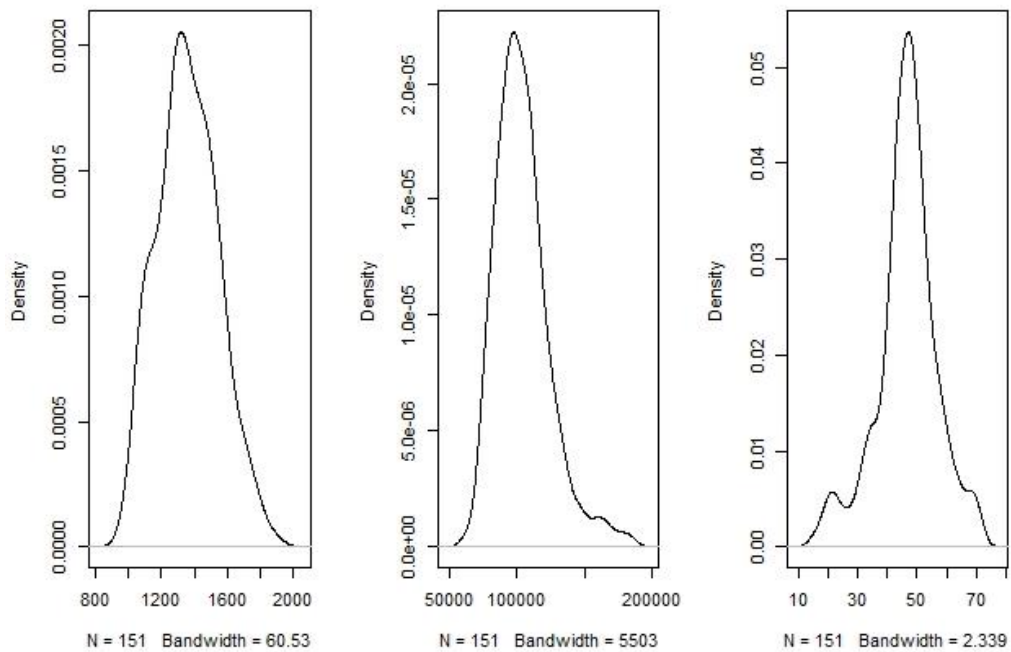
Histogramas de la producción histórica de hidrocarburos total para el yacimiento del campo de estudio.



Nota. Los gráficos muestran de izquierda a derecha en color verde el petróleo, en color rojo el gas y en color azul el agua.

Figura 8

Gráficos de densidad de las variables de producción histórica total



Nota: De izquierda a derecha el comportamiento de la producción total del petróleo, gas y agua, respectivamente.

Para corroborar lo visualizado en las figuras 7 y 8, la tabla 1 reporta las estadísticas descriptivas de las variables aleatorias del estudio: Producción histórica total de petróleo (Qo) en unidades barriles estándar por día (STB/D, por sus siglas en inglés), producción histórica total de gas (Qg) en unidades pies cúbicos estándar por día (SCF/D, por sus siglas en inglés) y finalmente, producción histórica total de agua (Qw) en unidades barriles estándar por día. El número de las observaciones para cada variable es de ciento cincuenta y una (151).

El análisis de estas estadísticas nos conduce a ratificar en los valores de sesgo la asimetría de las variables de producción total de gas y agua.

En estadística se llama sesgo de un estimador a la diferencia entre su esperanza matemática y el valor numérico del parámetro que estima. En este caso el coeficiente de sesgo positivo (1.12) para la producción total de gas y negativo (-0.41) para la producción total de agua, confirmando lo observado en las figuras 7 y 8; es decir una asimetría hacia la derecha y hacia la izquierda, respectivamente.

Tabla 1

Estadísticas descriptivas de las variables del estudio: De izquierda a derecha petróleo, gas y agua

Variable	Qo, barriles/día	Qg, pies cúbicos/día	Qw, barriles/día
Media	1,360.50	104,142	46.58
Desviación Estándar	183.50	20,018	10.11
Varianza	33,657.50	400,728,196	102.23
Coeficiente de Variación	13.48	19	21.70
Mínimo	992.70	64,449	16.03
Q1	1,241.30	90,867	41.90
Mediana	1,354.80	100,000	47.01
Q3	1,490.40	113,786	51.63
Máximo	1,875.90	182,395	70.67
IQR	249.20	22,919	9.72
Sesgo	0.24	1.12	-0.41
Curtosis	-0.37	2.02	0.94

Adicionalmente de la tabla (1) se puede deducir que siendo los coeficientes de variación para el petróleo, gas y agua menores al 80%, los datos son homogéneos y por tanto las medias son representativas de sus respectivos datos, sobresaliendo en este aspecto la del petróleo.

Alternativamente, las pruebas de normalidad de Kolmogorov-Smirnov y Anderson-Darling fueron aplicadas en la muestra para las variables de estudio. Los resultados se muestran en la tabla 2.

Tabla 2

Pruebas de normalidad aplicadas a las variables de la producción histórica total

Prueba de Normalidad	Kolmogorov - Smirnov		Anderson - Darling		Análisis
	KS	p-value*	AD	p-value*	
Qo	0.045	>0.150	0.380	0.390	KS < p-value AD < p-value
Qg	0.085	<0.010	1.967	<0.005	KS > p-value AD > p-value
Qw	0.095	<0.010	1.666	<0.005	KS > p-value AD > p-value

Nota: *Estadístico de la prueba (nivel de significancia)

Las pruebas de normalidad aplicadas buscan satisfacer a las siguientes hipótesis {Hipótesis Nula (Ho): Las variables aleatorias del estudio siguen una distribución normal; Hipótesis Alternativa (H1): Las variables aleatorias del estudio no siguen una distribución normal}

En general se observa de la tabla 2 que, para la producción total de aceite del campo, ambos estadísticos KS y AD resultan en valores menores al valor p. De esta manera, no se rechaza la hipótesis nula y se concluye que la variable sigue una distribución normal. No obstante, para las restantes dos variables aleatorias (producción total de gas y agua), los estadísticos KS y AD son mayores a sus correspondientes p-valores; rechazándose en consecuencia la hipótesis nula, y estableciendo que estas variables aleatorias **no** siguen distribuciones normales.

4.2. Fase Inferencial

Considerando los resultados obtenidos en la fase 1, se decide continuar el análisis exclusivamente para la producción total de aceite del campo.

El análisis de la serie de tiempo inicia mediante el análisis de gráficos que pueden revelar patrones aleatorios, tendencias, cambios de nivel, períodos o ciclos, observaciones inusuales o una combinación de patrones. La muestra de estudio comprende la producción total de aceite de un campo petrolero en el periodo de enero de 1996 a julio de 2008.

La figura 9 muestra la descomposición de la serie, en sus componentes: tendencia, estacionalidad y residual.

La tendencia se presenta estable, sin mayores fluctuaciones desde el inicio de la producción de aceite en el campo (enero, 1996) hasta el año 2000 donde se manifiesta un pico creciente que desde ese mismo año y hasta el final de la información (julio, 2008) se presenta esta tendencia como decreciente.

En relación al componente residual, presenta un comportamiento estable alrededor de un valor específico; indicando la presencia de ruido blanco.

Cuando la serie además de tendencia y residuales (componente aleatorio) refleja estacionalidad, los métodos de descomposición suponen que los datos se generan como una contribución de estos tres efectos (Peña, 2010). Expandiendo la ecuación (9), la serie temporal de la producción de petróleo en el estudio queda expresada por la ecuación (27):

$$x_t = \mu_t + S_t + a_t \quad (28)$$

Donde, x_t son las observaciones; μ_t representa la tendencia, S_t es el componente estacional y a_t , corresponde al componente puramente aleatorio.

Figura 9

Descomposición de la serie de tiempo original de la producción total de aceite.

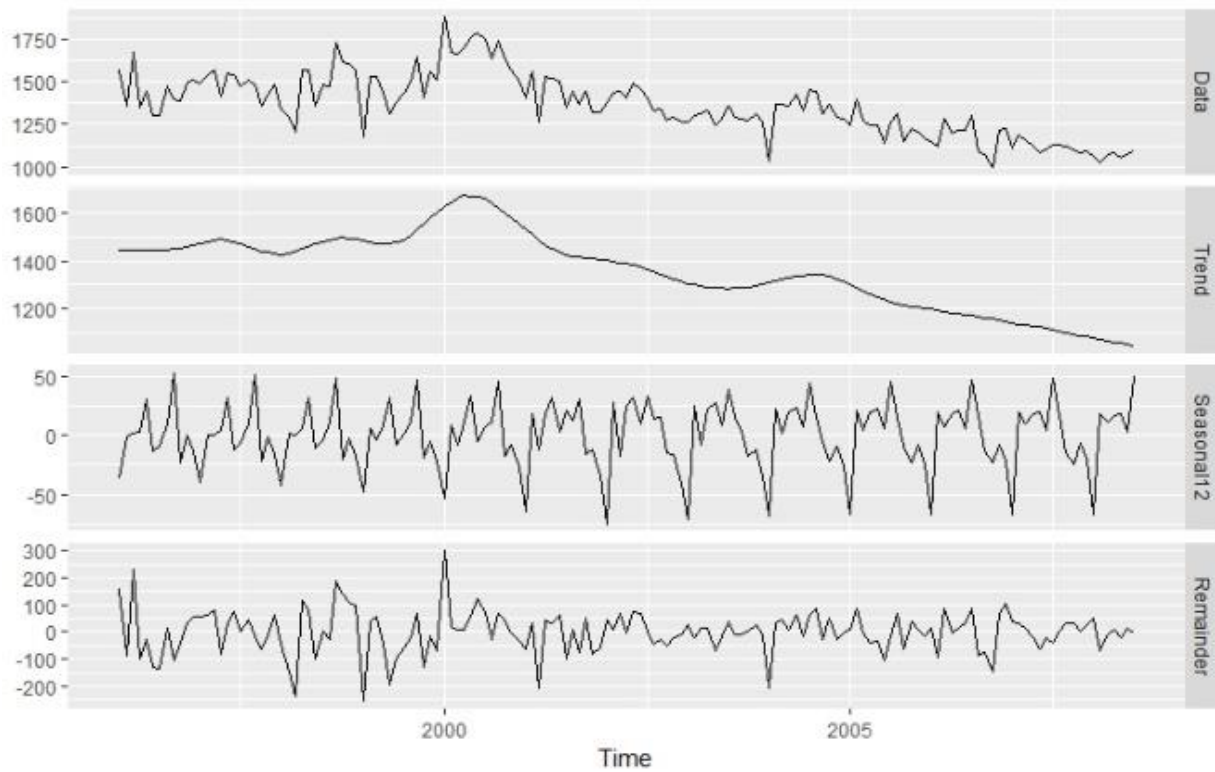
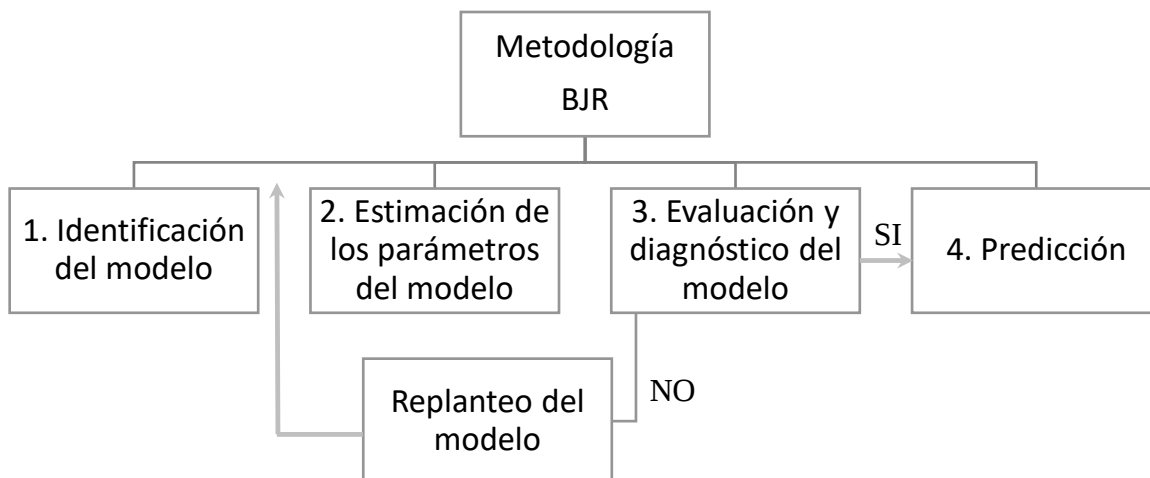


Figura 10

Metodología Box Jenkins y Reinsel.



Por otra parte, la estacionalidad se modela como una función periódica, que verifica la condición: $S_t = S_{t-s}$, donde s es el periodo de la función. En la serie temporal de este estudio, se evidencia una estacionalidad anual, debido a que la serie es mensual ($s=12$ meses) ; es decir, los coeficientes estacionales, S_t , se repetirán cada doce (12) observaciones.

Considerando el comportamiento de la serie, se emplea la metodología Box Jenkins y Reinsel (2008) para su modelado y predicción (Figura 10).

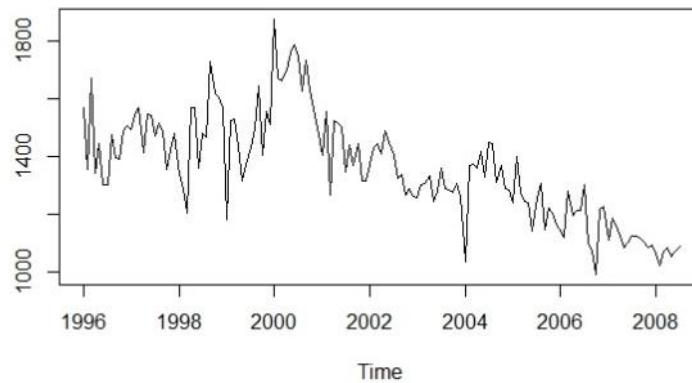
4.2.1. Identificación del modelo.

El objetivo de la primera etapa de la metodología BJR (figura 10), es determinar la estacionaridad de la serie de tiempo. Para esto, el uso de correlogramas simples y parciales es incluido dentro del análisis. De no encontrarse la estacionaridad en la serie original, esta debe ser inducida mediante transformaciones apropiadas de los datos (i.e., Diferenciación, transformación Box-Cox)

La figura 11 detalla el comportamiento original de la producción de petróleo en un campo petrolero en el tiempo, donde se observa en general una tendencia creciente desde los inicios de su explotación (enero 1996) y hasta enero 2000 donde la declinación de la producción se hace muy marcada debido a la merma de los pozos. Según Peña (2010), la tendencia de una serie puede eliminarse diferenciando los datos.

Figura 11

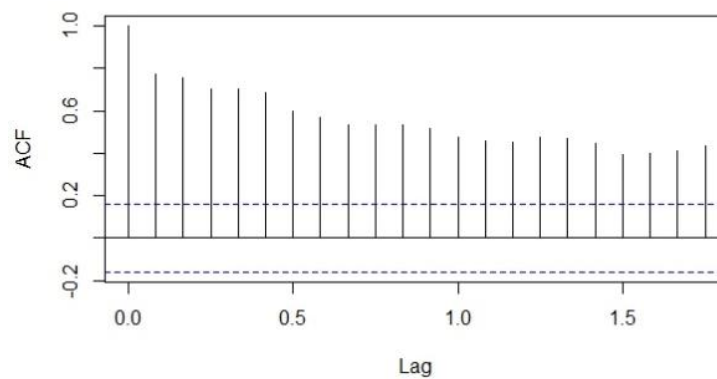
Serie de tiempo original de la producción total de petróleo de un campo en el periodo 01/1996 a 07/2008



La figura 12 presenta la función de autocorrelación simple de la serie original donde se observa un decrecimiento lento, indicando la no estacionaridad en el nivel de la serie. En la medida que el retardo aumenta, la correlación debe tender a cero, al no evidenciarse esto, la serie temporal muestra la no ergodicidad, ya que esta, requiere que las observaciones muy separadas en el tiempo estén poco correlacionadas; sin embargo, toda la información que queda por fuera del intervalo de confianza del 95% (líneas azules), muestra significancia.

Figura 12

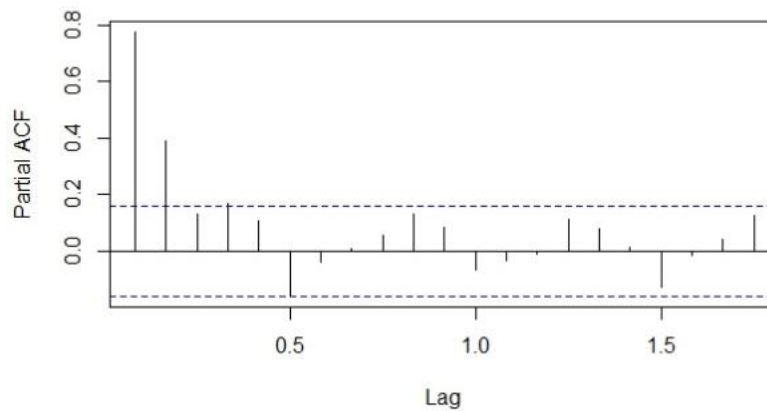
Correlograma simple de la serie de tiempo original.



En el correlograma de la función de autocorrelación simple, se puede apreciar la estacionalidad y al no caer rápidamente a cero, también refleja que la serie original no es homogénea y una vez más, para resolver esta situación se debe obtener la primera diferencia regular.

Figura 13

Correlograma parcial de la serie de tiempo original.



Finalmente, la figura 13 muestra la función de autocorrelación parcial de la serie original. Al contrario de la función de autocorrelación simple que identifica la naturaleza del proceso generador de datos; la PACF identifica con claridad el orden del modelo mediante el truncamiento del retardo.

Sin embargo, al no evidenciarse el cumplimiento de la condición de estacionaridad, no aplica ningún modelo para la serie original, el truncamiento en el segundo retardo no nos aporta información del orden del modelo. Adicionalmente, la estacionalidad está muy marcada.

Para confirmar lo evidenciado en las figuras 12 y 13, se corrió la prueba Dickey-Fuller a la ST original, que busca determinar la existencia o no de raíces unitarias en una serie de tiempo. La hipótesis nula de esta prueba es que existe una raíz unitaria en la serie, y cuando la raíz es unitaria, la serie resulta no estacionaria. El resultado demostró un $DF = -2.4409$, con un p-value de 0.3927. Al ser el p-valor mayor a 0.05, luego no se rechaza la hipótesis nula y la serie resultó ser no estacionaria.

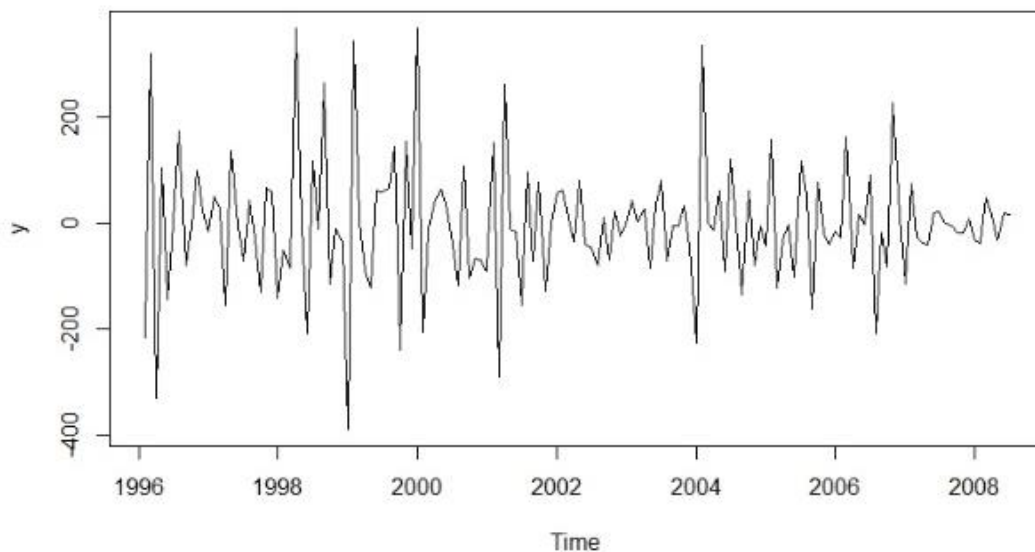
El paso siguiente fue el de diferenciar la serie original de la producción total de petróleo en el campo. Con la primera diferencia (Figura 14) se observa que los valores de la serie fluctúan alrededor de un valor central (cero); sin embargo, los fluctuantes altos y bajos valores de la serie diferenciada evidencian que la varianza es variable y evidenciando heterocedasticidad.

Para afirmar que la serie diferenciada en primer orden ahora es estacionaria, nuevamente se utilizó la prueba Dickey-Fuller, arrojando en esta oportunidad un estadístico $DF = -4.1433$, y un p-valor = 0.01.

Al ser el p-valor menor a 0.05, luego se rechaza la hipótesis nula y la serie diferenciada resultó ser estacionaria.

Figura 14

Serie de tiempo transformada de la producción total de aceite.

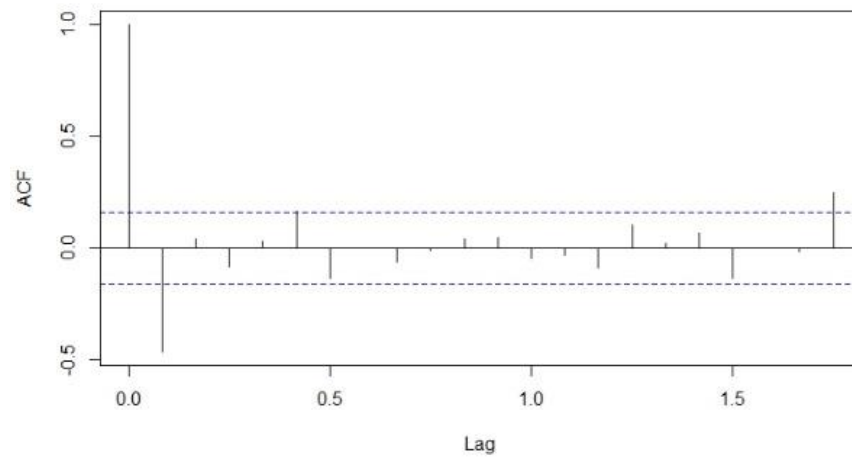


En el correlograma simple (Figura 15) se observa que en el retardo 1 la correlación es altamente significativa; sin embargo, todos los demás retardos se ubican dentro de las bandas de confianza.

El correlograma se trunca; es decir no tiene una caída distribuida pero rápida a cero como en el caso de los modelos autorregresivos (AR); indicando que el modelo es de promedio móvil (MA).

Figura 15

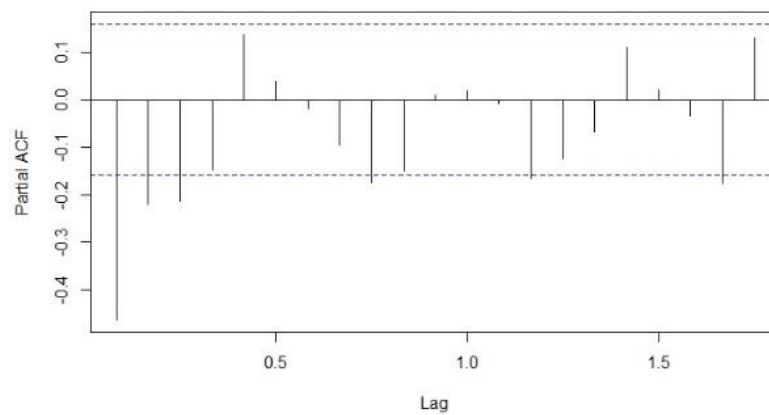
Correlograma simple de la primera diferencia de la serie de tiempo original.



Al tomarle la primera diferencia de la serie de tiempo, se observa que la parte regular ahora es estacionaria (figura 16) al igual que la parte estacional.

Figura 16

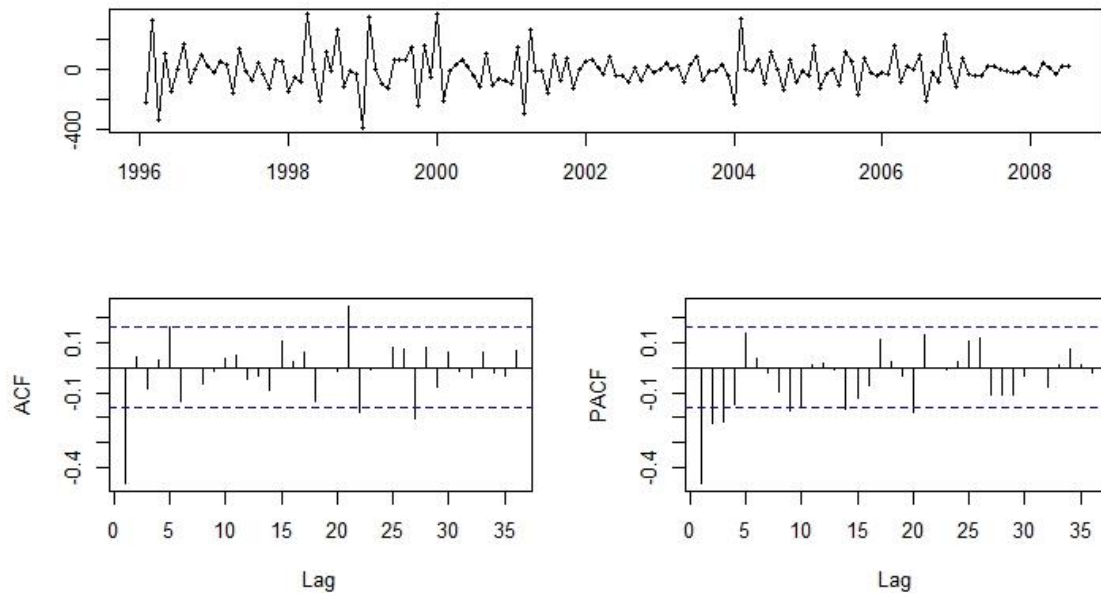
Correlograma parcial de la primera diferencia de la serie de tiempo



En resumen, la figura 17 muestra en su conjunto la serie diferenciada de la producción total de petróleo, simultáneamente con los comportamientos de las funciones ACF y PACF. El aumento en el retardo 21 no significa que la serie no es estacionaria, sino que la autocorrelación es espúrea.

Figura 17

Tsdisplay de la serie de tiempo de la producción total de aceite con la primera diferencia de la parte regular.



De esta manera, la serie de tiempo de la producción total de petróleo bajo la luz de este estudio presenta una diferencia en la parte regular ($d=1$) y una estacionalidad de periodo $s=12$ estacionaria ($D=0$). En consecuencia, el modelo que se presume es SARIMA $(p,1, q) (P,0, Q)_{12}$.

4.2.2. Estimación del modelo

En esta etapa, son usados los criterios de información en la selección del modelo. Para usar estos criterios, se estima por máxima verosimilitud un conjunto de potenciales modelos ARIMA. De acuerdo a la tabla 3, todos los modelos podrían ajustar a la ST; sin embargo, se identifica al modelo m3: SARIMA $(0,1,1) (1,0,0)_{12}$ como el de mejor ajuste de acuerdo a los criterios Akaike y Bayesiano; ya que estos criterios dependen de la suma de los errores cuadrados medios entre las observaciones de la serie temporal y los predictores estimados a partir de los modelos. Así, el modelo que ofrezca los menores valores AIC y BIC será el óptimo. Las características del modelo seleccionado (m3) se resumen en la tabla 4.

Tabla 3

Comparación de la bondad de ajuste de distintos modelos predictivos.

Modelo	Criterio AIC	Criterio BIC
m1:SARIMA(0,1,1)(0,0,1) ₁₂	1815.38	1824.41
m2:SARIMA(0,1,2)(0,0,1) ₁₂	1817.3	1829.34
m3:SARIMA(0,1,1)(1,0,0) ₁₂	1815.32	1824.35
m4:SARIMA(0,1,2)(1,0,0) ₁₂	1817.23	1829.27

Tabla 4

Características del modelo seleccionado.

	ma1	sar1
Coeficientes =	-0.6304	-0.0941
s.e. =	0.0663	0.0836
sigma ² estimado =	10235	
log likelihood=	-904.66	

4.2.3. Evaluación y Diagnóstico del modelo

En esta tercera etapa de la metodología BJR, se analizan los errores residuales, verificando la ausencia de patrones y la presencia de ruido blanco mediante la prueba Q.

Un error residual es la diferencia entre el valor observado y el valor estimado a partir de un modelo en un tiempo t (Cowpertwait y Metcalfe, 2009). Por ejemplo, si el modelo es definido por la variable y_t y el estimador es $\hat{y}_t$, el error ε_t viene dado por la ecuación (29)

$$\varepsilon_t = y_t - \hat{y}_t \quad (29)$$

La prueba Q, también llamada prueba Ljung-Box aplicado a los residuales establece como hipótesis nula que todos los coeficientes de autocorrelación son nulos.

Esto se expresa por la ecuación (30), como:

$$H_0: \rho_1 = \rho_2 = \rho_3 = \dots = \rho_k = 0 \quad (30)$$

Así, si se acepta la hipótesis nula la información de la ST es netamente ruido blanco. De otra manera, si se rechaza la hipótesis nula, la ST tiene información que puede modelarse.

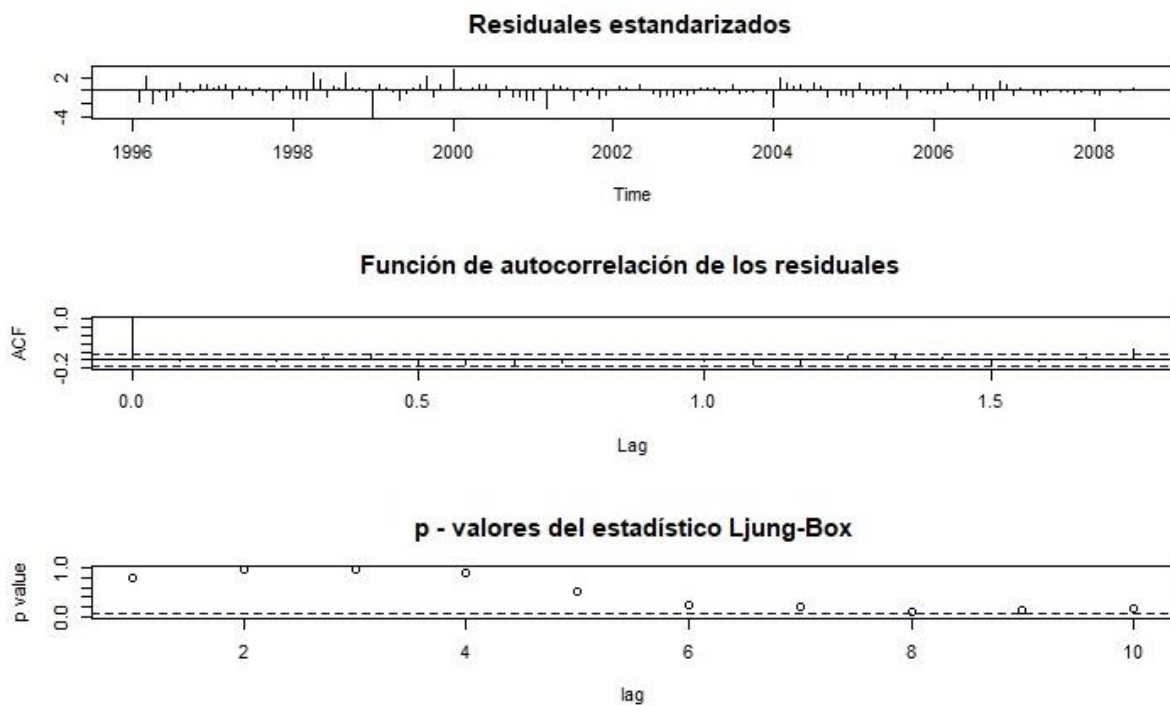
Las condiciones para que el proceso ruido blanco ocurra son: a) La esperanza matemática de los valores de la serie de tiempo, es siempre constante e igual a cero, b) la varianza es constante y c) las variables del proceso están incorrelacionadas para todos los retardos (Peña, 2010).

De esta manera, el proceso estocástico llamado ruido blanco no contiene información útil para predecir valores futuros, así $x_t = \varepsilon_t$.

La figura 18 muestra que los p -values del modelo seleccionado son mayores al 5%. En consecuencia, no se rechaza la hipótesis nula; y se verifica que los residuos son ruido blanco; es decir, el modelo capturó toda la información.

Figura 18

Análisis de los residuales del modelo SARIMA (0,1,1) (1,0,0)₁₂ empleado para la serie de tiempo de la producción total de petróleo.

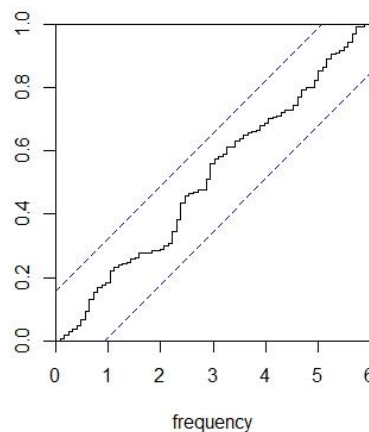


Adicionalmente a la prueba Q, el periodograma que es análogo al Correlograma, es una estrategia gráfica alternativa para verificar la presencia de ruido blanco (Brockwell y Davis, 1991).

Este es definido en términos de la transformada discreta de Fourier y puede ser expresado en términos de la función de autocovarianza muestral (ver figura 19).

Figura 19

Análisis del periodograma de los residuales del modelo SARIMA (0,1,2) (0,0,1)₁₂.



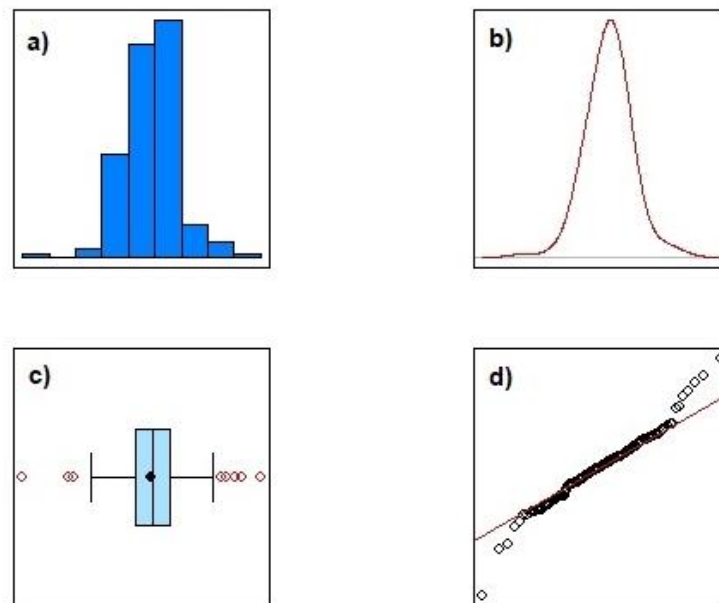
La figura 19 muestra como el modelo de estimación (m3) recogió toda la información de la serie de tiempo, porque en el residuo ya no hay más información debido a que al pasar dentro de la banda definida por el intervalo de confianza del 95%, confirma que estos errores residuales son aleatorios; confirmándose la independencia de los residuos.

Para probar si los residuales son normales, la prueba de Shapiro-Wilk es usada. Esta establece como hipótesis nula, que los residuos son normales. Los resultados mostraron un valor $W = 0.97021$, y un $p\text{-valor} = 0.002314$. Como el valor p es pequeño, rechazamos la hipótesis nula y en consecuencia, los errores residuales del modelo aplicado a la primera diferencia de la serie de tiempo de la producción total de petróleo, no se distribuyen normalmente. Esto es importante de resaltar, a pesar de que la variable en la serie original se distribuye normalmente, la serie diferenciada no lo hace. Cuando los errores no son normales, las estimaciones de mínimos cuadrados pueden no ser óptimas. Seguirán siendo las mejores estimaciones lineales insesgadas

(como se evidencia en la tabla 5, donde el sesgo es igual al valor de -0.053), pero otros estimadores robustos pueden ser más efectivos. En estos casos, se recomienda verificar la normalidad de los errores mediante la aplicación del gráfico Q – Q (ver figura 20). El Q-Q plot de los residuales del modelo (figura 20) confirmó la no normalidad de los mismos; de esta manera la resolución depende del tipo de problema encontrado. De acuerdo con Faraway (2015), para las distribuciones de cola corta (como la de la serie temporal de estudio), las consecuencias de la no normalidad no son graves y pueden ignorarse razonablemente. Adicionalmente, de la figura 20 se observó, que en el Boxplot de los residuales se presentaron valores extremos (*Outliers*) tanto por la izquierda como por la derecha de los límites mínimos ($Q1-1.5IQR$) y máximos ($Q3+1.5IQR$), producto de la diferenciación de la serie temporal.

Figura 20

Análisis exploratorio de los residuales para el modelo SARIMA (0,1,2) (0,0,1)₁₂.



Nota: De izquierda a derecha, de arriba abajo tenemos a) El histograma de los residuales, b) El gráfico de densidad, c) El boxplot y d) el Q-Q plot.

Tabla 5

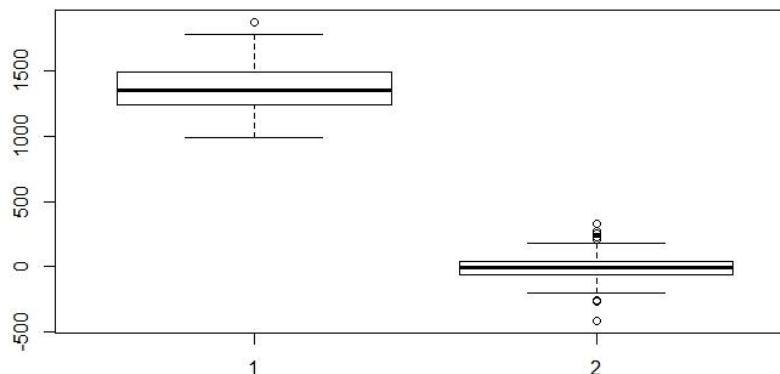
Estadísticas descriptivas de los errores residuales del modelo seleccionado para la ST diferenciada del estudio

Variable	Residuales del modelo seleccionado
Tamaño muestral	151
Media	-8.73
Desviación Estándar	100.11
Varianza	10,021.91
Mínimo	-410.56
Q1	-61.48
Mediana	-6.23
Q3	46.17
Máximo	330.48
IQR	107.65
Sesgo	-0.05
Curtosis	2.05

Pese a que, en la etapa de estimación de los parámetros del modelo, el seleccionado por los criterios de información AIC y BIC para la serie diferenciada de la producción total de petróleo fue el m3: SARIMA (0,1,1) (1,0,0)₁₂, se decidió correr el coeficiente de determinación R^2 , el cual indicó un valor de $R^2 = 0.7022387$, concluyéndose que el 70.22% de la información de la serie temporal es capturada por el modelo.

Figura 21

Boxplots comparativos entre (1) la serie original y (2) la variabilidad de los residuos del modelo SARIMA (0,1,1) (1,0,0)₁₂.



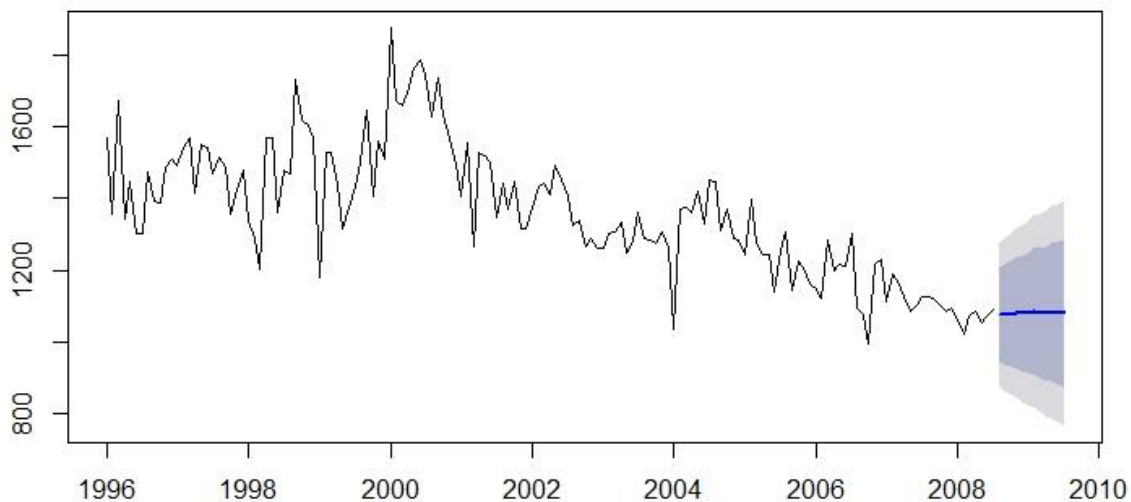
La figura 21 compara como está la variabilidad de los residuos con respecto a la serie original. Los fluctuantes altos y bajos valores en los residuos del modelo seleccionado introducen incertidumbre acerca de si realmente es el modelo predictivo de la serie temporal.

4.2.4. Predicción de la Serie Temporal

Esta etapa final de la metodología BJR busca hacer uso del modelo para obtener predicciones futuras de la serie. Las estimaciones del modelo consideraron la muestra histórica completa de la producción total de petróleo del campo y aun así, el modelo sólo recogió aproximadamente el 70% de la información. Esta serie dentro de su componente regular no pudo ser modelada autorregresivamente; es decir, no pudo ser expresada como una función lineal de sus observaciones previas o pasadas, pero si como una función lineal de sus errores actuales y previos ($p=0$ y $q=1$). La figura 22 muestra las predicciones del modelo para un ciclo de 12 observaciones futuras (1 año), donde se observaron predicciones lineales, in-sesgadas y de varianza mínima para la serie de tiempo que no se ajustan con la naturaleza de la misma.

Figura 22

Predicciones de la serie de tiempo producción total de petróleo.



A pesar de que en su fase predictiva el modelo no presentó satisfactorios resultados, este ejercicio se realizó para completar el ciclo metodológico BJR y obtener predicciones comparativas considerando el modelo m3: SARIMA (0,1,1) (1,0,0)₁₂.

4.3. Fase Predictiva

La aplicación de la metodología BJR para el análisis y el modelamiento de procesos estocásticos no estacionarios, como el del objeto de este estudio, se relaciona con la función aleatoria intrínseca de orden k , introducida en Geoestadística por Matheron (Bruno y De Fouquet, 2005), donde el componente de tendencia es interpretado usualmente como un componente determinístico, modelado como un polinomio local. Si el valor esperado del proceso no es constante, pero puede ser expresado como una función polinómica, el predictor Kriging puede ser usado. Este predictor, cumple las condiciones de ser insesgado y disponer del mínimo error de varianza. De igual manera, el componente periódico de la serie (estacional) pueden ser articulado por el variograma.

4.3.1. Estudio de la estructura de autocorrelación a través del variograma

Considerando los resultados obtenidos en el modelamiento de la información de la serie de tiempo (producción total de petróleo), un análisis variográfico para esta serie fue desarrollado. El punto de partida fue la construcción del semivariograma empírico definido por la ecuación (24) y mostrado en forma de nube variográfica (puntos negros) en la figura 23.

Para el desarrollo de este variograma empírico, el retardo τ_t representa la separación entre cada dos (2) mediciones de la serie de tiempo, separados una distancia h , permitiendo que el análisis del variograma o semivariograma como es comúnmente conocido, se realizara en un dominio temporal y no espacial.

El siguiente paso consistió en el ajuste de un modelo teórico. Existen muchos y diferentes semivariogramas teóricos propuestos en la literatura (Journel y Huijbregts, 1978); sin embargo, para el análisis de la producción total de petróleo se probaron los tres tipos de modelos que representan a los más comúnmente empleados en geoestadística debido a su flexibilidad, que se definen según los parámetros de variabilidad: 1) Nugget, también conocido como efecto pepita, el cual se ubica en el intercepto de la ordenada en la gráfica del semivariograma y representa una variación de escala o errores en la medición de los datos; 2) la meseta (psill) que representa el valor al cual, el semivariograma alcanza la máxima variabilidad de los datos; y 3) el rango (a), que representa la distancia o en este caso el tiempo en el cual la meseta es alcanzada. Para tiempos posteriores al rango, se presume que la correlación de los datos es cero.

Modelo esférico, definido por la ecuación (31), donde los coeficientes θ y β , se definen en la tabla 6 y h es la distancia entre las mediciones en tiempo del conjunto de observaciones.

$$\gamma_h = \text{nugget} + [1.5\theta\beta h - 0.5\theta\beta^3 h^3], \quad h \leq a \quad (31)$$

$$\gamma_h = \theta, \quad h > a$$

Modelo Gaussiano, definido por la ecuación (32) y los mismos parámetros de variabilidad.

$$\gamma_h = \text{nugget} + [\alpha \exp(\beta h^2) + \theta], \quad |h| \geq 0 \quad (32)$$

Modelo Exponencial, definido por la ecuación (33)

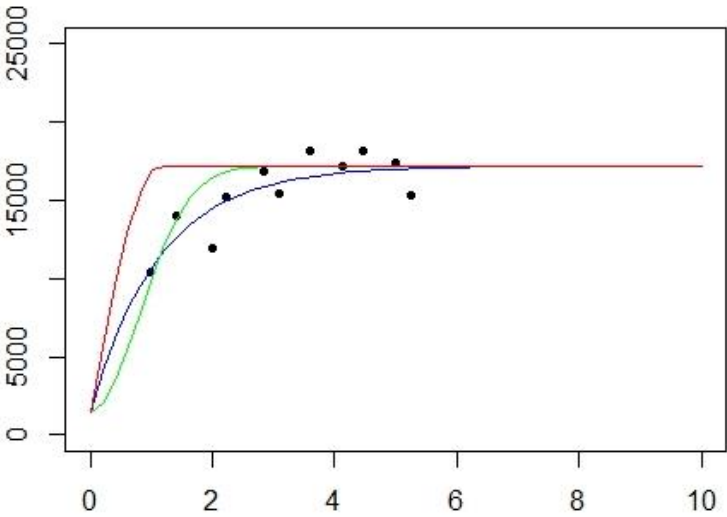
$$\gamma_h = \text{nugget} + [\alpha \exp(\beta h) + \theta], \quad |h| \geq 0 \quad (33)$$

Tabla 6.
Coefficientes de modelos teóricos relacionados con los parámetros de variabilidad del análisis estructural de los datos

Coeficiente	Modelo Esférico	Modelo Gaussiano	Modelo Exponencial
alpha (α)	-	-psill	-psill
beta (β)	$1/a$	$-1/a^2$	$-1/a$
theta (θ)	psill	psill	psill

Después de un ajuste automático a partir de la función `fit.variogram` de la librería `gstats`, se visualizó que el modelo exponencial fue el que mejor se ajustó al semivariograma empírico, ya que representó a una mayor cantidad de información (Figura 23). Al mismo tiempo, los valores para los parámetros de variabilidad resultantes para este modelo exponencial, están reportados en la tabla 6.

Figura 23
Modelos de variograma para el histórico de la producción total de petróleo.



Nota: Los puntos negros representan el semivariograma empírico y las líneas representan a los modelos teóricos generados: Esférico (línea roja), Gaussiano (línea verde) y Exponencial (línea azul).

La tabla 7 resume los valores resultantes los parámetros de variabilidad, de acuerdo al ajuste del modelo exponencial con el semivariograma empírico, considerando todos los datos observados del histórico de producción total de petróleo para el campo de estudio (periodo 01/01/1996 – 01/07/2008).

Tabla 7

Parámetros de variabilidad de acuerdo al análisis estructural de los datos

Parámetros de Variabilidad	Valor
Nugget	1487.926 barriles/día
Meseta	15706.968 barriles/día
Rango	1.140872 años*mes

Sin embargo, para la verificación del ajuste entre cada modelo teórico variográfico y el variograma empírico, fue implementada una metodología propuesta por Spiegelhalter, Best, Carlin, y Van der Linde (2002).

Esta metodología propone la aplicación del criterio de información de desviación (DIC, por sus siglas en inglés), para conocer el ajuste del mejor modelo de inferencia bayesiana. Mediante el paquete BUGS, software para realizar inferencias bayesianas utilizando muestreo de Gibbs, en su versión abierta OpenBugs, se especificaron los modelos estadísticos, de complejidad (casi) arbitraria, simplemente indicando las relaciones entre las variables relacionadas. Posteriormente, desde Rstudio, un esquema apropiado de las cadenas de Markov en el método de Monte Carlo (Hadfield, 2010) fue analizado a partir de la función bugs, de la librería R2WinBUGS.

Para cada modelo se corrieron 4 cadenas independientes con 10,000 iteraciones para un periodo de descarte (burn –in) de 5000 iteraciones. Las Figuras 24 a la 26 muestran las distribuciones a posteriori de las cadenas de Markov, donde se observan que cada cadena reprodujo la distribución de cada uno de los coeficientes α , β y θ ; respectivamente.

Adicionalmente, se muestran las gráficas de autocorrelación, las gráficas del comportamiento numérico de las corridas y de las variaciones por iteración por cada cadena para cada coeficiente del modelo teórico exponencial.

Figura 24

Gráficos de Diagnóstico para el coeficiente alpha (α) del modelo teórico exponencial.

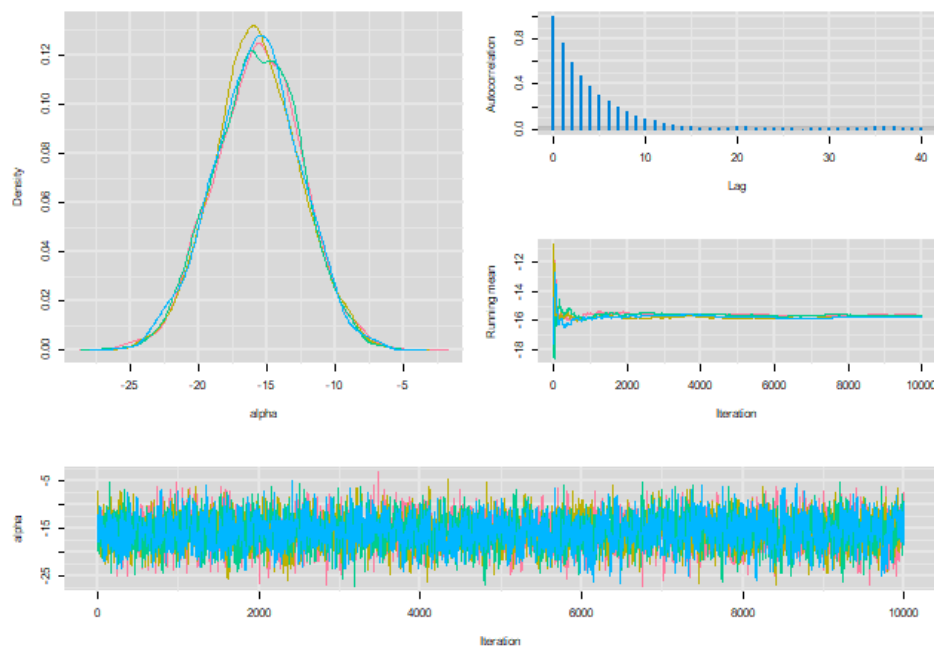


Figura 25

Gráficos de Diagnóstico para el coeficiente beta (β) del modelo teórico exponencial.

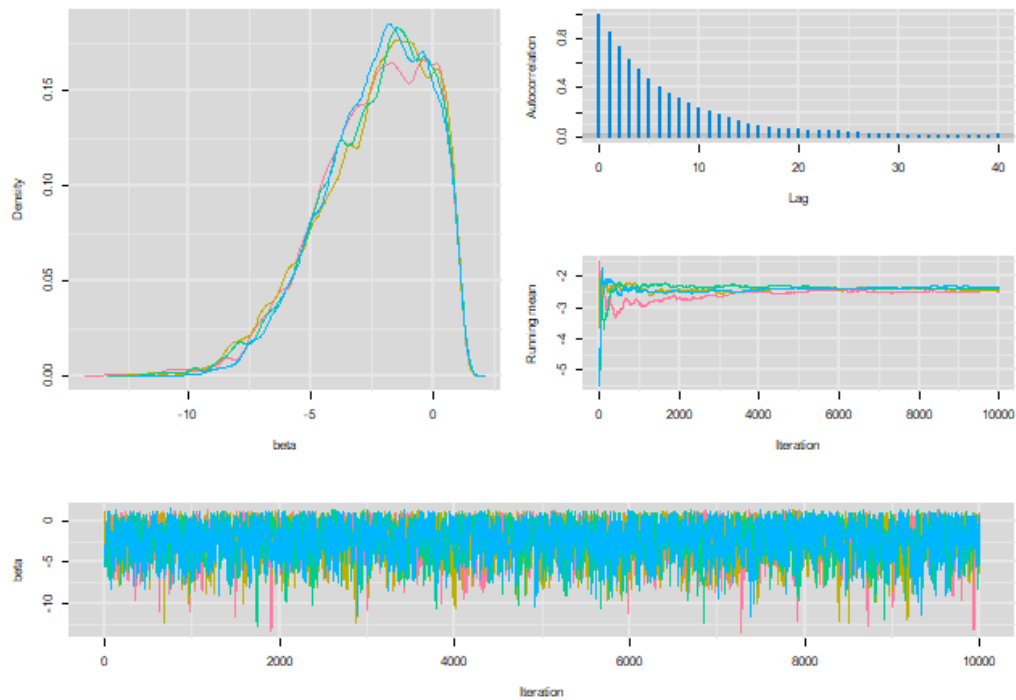
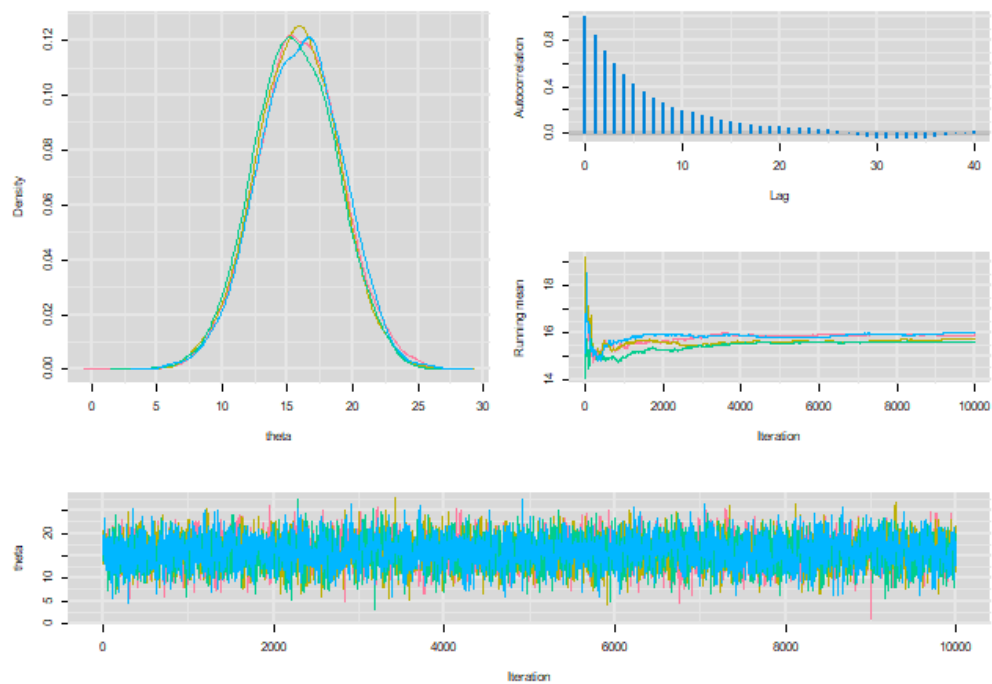


Figura 26

Gráficos de Diagnóstico para el coeficiente theta (θ) del modelo teórico exponencial.



Por cada modelo teórico se corrieron gráficos diagnósticos; donde se observó respuestas similares a las mostradas por las figuras 24 a la 26 en lo que respecta al modelo gaussiano y esférico. La tabla 8 resume los valores del promedio de las desviaciones de las observaciones a-posteriori y el valor DIC por cada modelo teórico, donde se confirma que el modelo exponencial es el que mejor se ajustó al semivariograma empírico.

Tabla 8

Resumen de las desviaciones para los modelos de inferencia Bayesiana ajustados a los datos de la producción total de petróleo del campo de estudio, en el periodo 01/01/1996 – 01/07/2008.

Modelo	Exponencial	Gaussiano	Esférico
$\bar{D}$	191.35	200.48	195.99
DIC	193.81	202.40	198.50

4.3.2. Comparación predictiva entre la metodología BJR de la serie temporal y la técnica geoestadística Kriging

La técnica Kriging es un método de interpolación comúnmente usado en geoestadística. Esta depende de la autocorrelación de los datos observados para realizar inferencias en los valores no observados en un proceso aleatorio (Journel y Huijbregts, 1978).

Antes de usar esta técnica con fines predictivos para datos faltantes, objetivo final de este trabajo, en esta sección se propuso comparar la técnica con las predicciones realizadas en la sección 4.2.4, generadas por el análisis de la serie temporal y la aplicación de la metodología Box, Jenkins y Reinsel (2008).

Se resalta el hecho de que, para este ejercicio, el conjunto de datos conocidos fue reducido, eliminando únicamente 6 valores. Así, considerando únicamente el periodo desde el 01/01/1996 hasta el 01/12/2007; se predijo la producción de petróleo a lo largo de seis (6) meses del año 2008, y de las estimaciones se pudo calcular errores que permitieron cuantificar las diferencias entre ambas técnicas.

Al reducir el tamaño de la muestra, el semivariograma empírico se ve afectado y en consecuencia un nuevo modelo debe ser ajustado, con diferentes parámetros de variabilidad. A pesar de que el modelo exponencial fue el que mejor se ajustó a la serie temporal completa, la predicción orientada hacia la estructura externa de la serie, generó los nuevos parámetros de variabilidad como: nugget = 0, para darle continuidad en todos los retardos; meseta igual a la varianza de los datos del periodo seleccionado (30804 barriles/día) y la distancia de correlación que se ajustó a esa variabilidad (rango = 5 años*mes). La figura 27 muestra el resultado obtenido del nuevo semivariograma teórico. No obstante, nuevamente se determinó a partir del criterio de información de la desviación, DIC, si efectivamente, este resultaba el mejor modelo de ajuste. En esta oportunidad se utilizó directamente de Rstudio la librería MCMCglmm (Hadfield, 2010).

La tabla 9 resume los valores del promedio de las desviaciones de las observaciones a-posteriori y el valor DIC por cada modelo teórico, donde se confirma que el modelo exponencial es el que mejor se ajustó al nuevo semivariograma empírico (nube variográfica en la figura 27).

Figura 27

Semivariograma exponencial definido para la predicción a futuro de la producción total de petróleo del campo de estudio (año 2008).

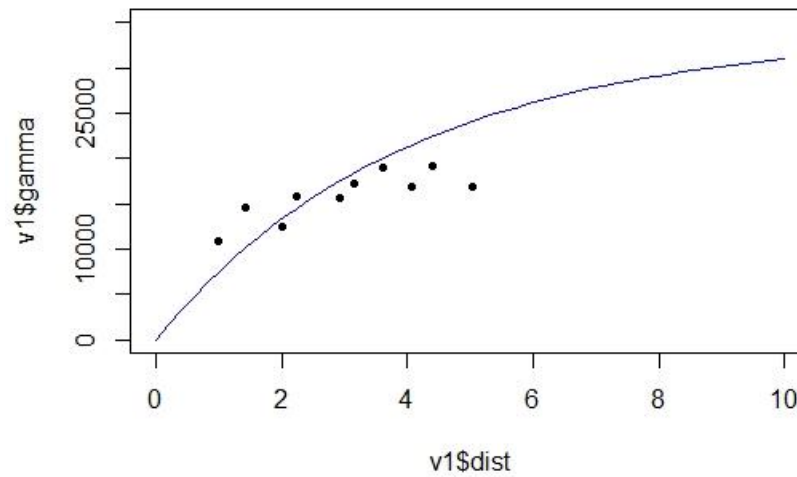


Tabla 9

Resumen de las desviaciones para los modelos de inferencia Bayesiana ajustados a los datos de la producción total de petróleo del campo de estudio, en el periodo 01/01/1996 – 01/12/2007.

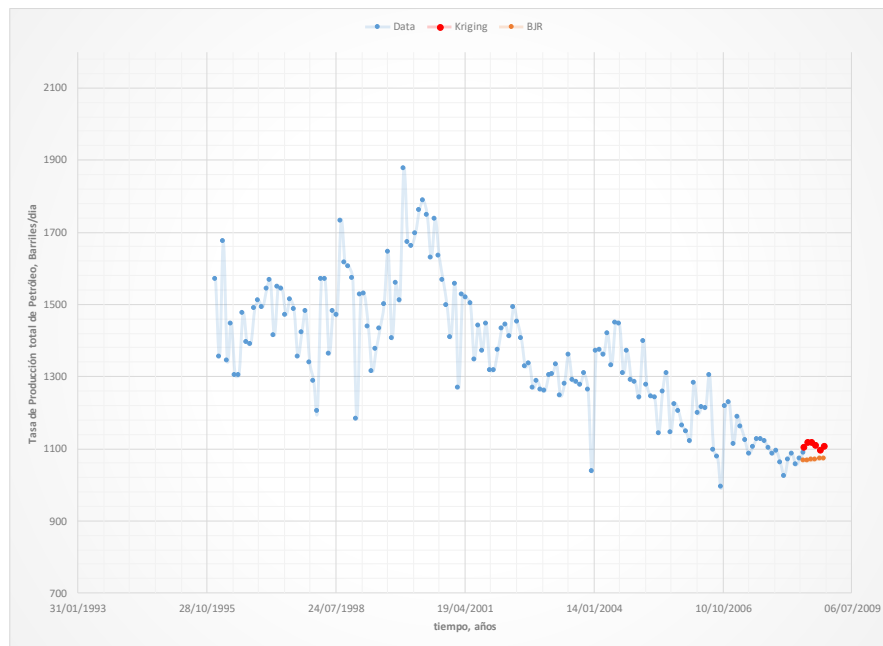
Modelo	Exponencial	Gaussiano	Esférico
$\bar{D}$	174.85	178.75	176.41
DIC	177.32	181.18	178.89

Para la aplicación de la técnica geoestadística Kriging, se forma el sistema matricial de la ecuación (27) y posteriormente con la ecuación (26) se determina las estimaciones en las locaciones no observadas. El análisis conjunto, del componente estacional y de tendencia de la serie histórica, siguió la metodología propuesta por Baca (2016), definiendo la ubicación de las mediciones mensuales en el eje coordenado Y, y los años en el eje coordenado X. La etapa predictiva de la producción total de petróleo para el periodo 01/01/2008 - 01/12/2008, resultó satisfactoria. La figura 28 muestra la comparación de los resultados estimados a partir de las dos

técnicas estadísticas empleadas en el desarrollo de esta investigación: En color naranja los datos de predicción de la serie de tiempo, usando la metodología BJR y en color rojo los datos de predicción de la técnica geoestadística Kriging.

Figura 28

Predicción a futuro de la producción total de petróleo del campo de estudio para el año 2008.



Nota: Los datos históricos en color azul, los datos estimados a partir de la metodología BJR en color naranja, mientras los datos estimados a partir de la técnica kriging en color rojo.

Adicionalmente, la tabla 10 muestra el cálculo de los errores de la predicción geoestadística, mostrando un error porcentual promedio absoluto (MAPE, por sus siglas en inglés) del 5%, además de la comparación tabulada de los resultados de las dos técnicas empleadas como predictivas.

Tabla 10

Errores estimados de la predicción geoestadística vía Kriging de la producción total de petróleo en el periodo 01/01/2008 – 01/07/2008

Geoestadística							Metodología BJR			
n	Tiempo	Data Medida (Qo, B/D)	Predicción Metodología BJR	Predicción (Kriging)	Error (Et)	Error Absoluto (Et)	Et/At	Error (Et)	Error Absoluto (Et)	Et/At
1	01/01/2008	1061.13		1140.93	-79.81	79.81	0.075			
2	01/02/2008	1023.04		1151.98	-128.93	128.93	0.126			
3	01/03/2008	1070.07		1138.41	-68.35	68.35	0.064			
4	01/04/2008	1085.60		1109.71	-24.11	24.11	0.022			
5	01/05/2008	1054.44		1088.07	-33.63	33.63	0.032			
6	01/06/2008	1072.36		1090.55	-18.19	18.19	0.017			
7	01/07/2008	1089.03	1067.481	1104.69	-15.66	15.66	0.014	21.545	21.545	0.020
8	01/08/2008		1067.514	1116.28						
9	01/09/2008		1068.08	1118.23						
10	01/10/2008		1069.761	1108.46						
11	01/11/2008		1071.415	1097.19						
12	01/12/2008		1070.568	1106.48						
Totales					-368.6	368.6				
Promedio del Error Absoluto Porcentual (MAPE)					5.008					

De acuerdo a los resultados obtenidos de la predicción de la serie a futuro se concluye que la estimación geoestadística reproduce la tendencia y el componente estacional de los datos que no se evidenció en la aplicación de la metodología BJR.

4.3.3. Predicción de los valores ausentes mediante el uso de la técnica geoestadística Kriging

El segundo procedimiento empleado para la técnica geoestadística, fue predecir valores faltantes, orientando las estimaciones hacia la estructura interna de la serie. Del histórico de la

producción total de petróleo del campo en el periodo 01/01/1996 – 01/12/2007, se procedió a eliminar datos observados adicionales; en esta oportunidad dentro de un periodo de dos ciclos estacionales, 24 meses (01/01/2001 – 01/12/2002); de tal manera que la técnica geoestadística pudiera predecir estos valores.

Nuevamente, al disminuir el tamaño de la muestra, el semivariograma empírico se ve afectado y en consecuencia un nuevo modelo debe ser ajustado, con diferentes parámetros de variabilidad.

Después de un ajuste automático a partir de la función `fit.variogram` de la librería `gstats`, se visualizó que el modelo esférico fue el que mejor se ajustó al semivariograma empírico, ya que representó a una mayor cantidad de información (Figura 30). Al mismo tiempo, los valores para los parámetros de variabilidad resultantes para este modelo fueron $\text{nugget} = 6878.941$ barriles/día, $\text{meseta} = 15785.753$ barriles/día, y $\text{rango} = 5.282695$ años*mes.

No obstante, nuevamente se determinó a partir del criterio de información de la desviación, DIC, si efectivamente, este resultaba el mejor modelo de ajuste. La tabla 11 resume los valores del promedio de las desviaciones de las observaciones a-posteriori y el valor DIC por cada modelo teórico, donde se confirma que el modelo esférico es el que mejor se ajustó al nuevo semivariograma empírico (nube variográfica en la figura 29), por presentar el menor valor DIC.

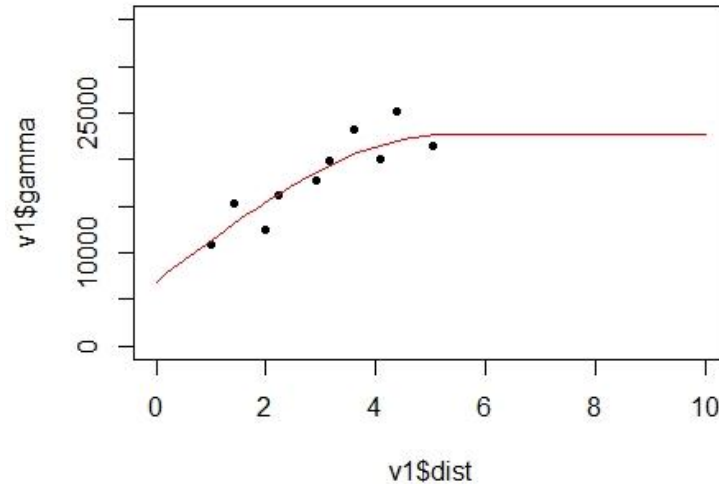
Tabla 11

Resumen de las desviaciones para los modelos de inferencia Bayesiana ajustados a los datos de la producción total de petróleo del campo de estudio, con los datos faltantes del periodo 01/01/2001 – 01/12/2002.

Modelo	Exponencial	Gaussiano	Esférico
$\bar{D}$	183.46	185.25	182.65
DIC	185.92	187.71	185.06

Figura 29

Semivariograma esférico definido para la predicción de valores faltantes de la producción total de petróleo del campo de estudio (periodo 01/01/2001 – 01/12/2002)



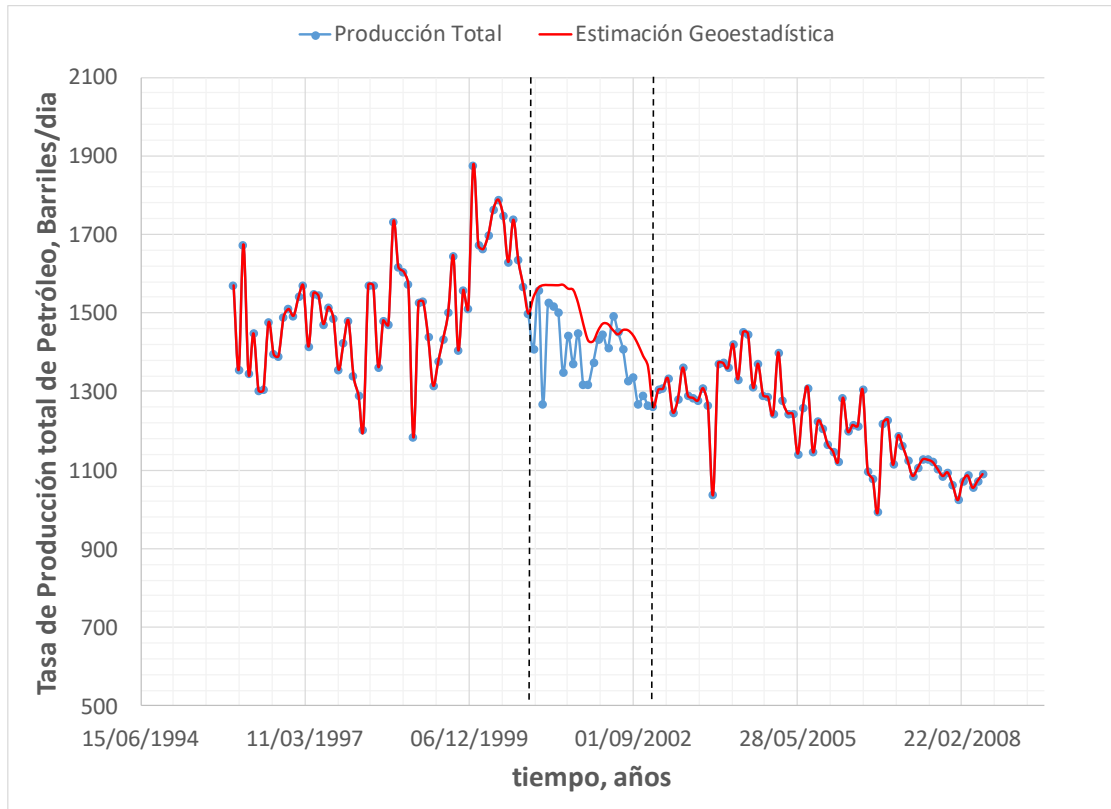
El modelo teórico esférico utilizado en la aplicación de la técnica geoestadística de Kriging ordinario, para la obtención de los datos faltantes arrojó estimaciones que comparadas con los datos observados generó un error porcentual promedio absoluto (MAPE, por sus siglas en inglés) del 7.15%, considerando aún el resultado satisfactorio. La figura 30 muestra como la técnica geoestadística honra no sólo la tendencia sino también la estacionalidad de la serie.

De la figura 31 se puede apreciar que las mejores producciones estuvieron concentradas en los primeros años de producción y existe una alta variabilidad mensual de acuerdo a cada año de producción.

En el análisis de los errores, la figura 32 muestra que son normales tanto el histograma, como el plot de densidad, la media muy cercana a la mediana en el boxplot y la tendencia lineal del Q-Q plot. Adicionalmente, la prueba de Shapiro Wilk arrojó un $W = 0.96769$ y un p-valor de 0.6104; resultando no significativo para un $\alpha = 0.05$, y en consecuencia no se rechaza la hipótesis nula; es decir los residuos son normales.

Figura 30

Predicción de los valores faltantes de la producción total de petróleo del campo de estudio para el periodo 01/01/2001 – 01/12/2002.



Nota: Los datos históricos en color azul, los datos estimados en color rojo.

Figura 31

Distribución de los volúmenes de producción total de petróleo anual (X) y mensual (Y) en el campo de estudio.

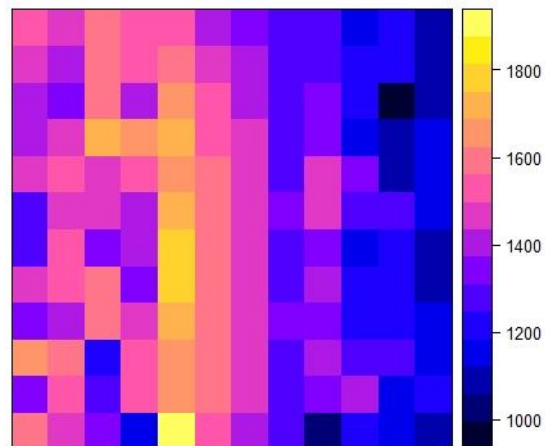
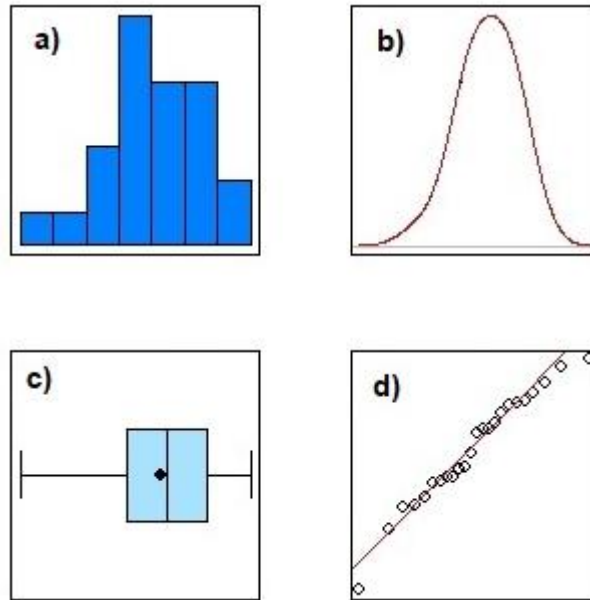


Figura 32

Análisis exploratorio de los residuales para las predicciones realizadas mediante estimación geoestadística de la producción total de petróleo en el periodo 01/01/2001 – 01/12/2002.



Nota: De izquierda a derecha, de arriba abajo tenemos a) El histograma de los residuales, b) El gráfico de densidad, c) El boxplot y d) el Q-Q plot.

Adicionalmente, la prueba de Durbin-Watson (Faraway, 2015) fue aplicada, obteniendo un $DW = 2.0735$ y un p-valor de 0.4852, que al ser un valor mayor al 5%, genera no rechazo a la hipótesis nula que establece que los errores están incorrelacionados; en consecuencia, son independientes.

CONCLUSIONES

Entre las características más resaltantes de la serie temporal producción total de petróleo del campo de estudio están la distribución normal de sus observaciones, el comportamiento no estacionario, no es homogénea, existe tendencia en sus observaciones, y el componente estacional está presente, así como la heterocedasticidad y la no ergodicidad en sus observaciones.

La metodología Box Jenkins y Reinsel (BJR) fue aplicada a la serie de tiempo de la producción total de petróleo diferenciada en su parte regular ($d=1$) y una estacionalidad de periodo $s=12$ estacionaria ($D=0$). Posterior a la evaluación de potenciales modelos, se presume la serie temporal es SARIMA $(0,1,1) (1,0,0)_{12}$, considerando el menor valor AIC y BIC (1815.32 y 1824.35; respectivamente). Este modelo recogió el 70% de la información de la serie; sus residuos resultaron incorrelacionados, pero no normales.

Cuando los errores no son normales, las estimaciones de mínimos cuadrados pueden no ser óptimas. Seguirán siendo las mejores estimaciones lineales insesgadas, pero otros estimadores robustos pueden ser más efectivos.

Esta serie dentro de su componente regular no pudo ser modelada autorregresivamente; es decir, no pudo ser expresada como una función lineal de sus observaciones previas o pasadas, pero si como una función lineal de sus errores actuales y previos ($p=0$ y $q=1$). Así, las predicciones del modelo para un ciclo de 12 observaciones futuras (1 año), donde se observaron predicciones lineales, in-sesgadas y de varianza mínima para la serie de tiempo no se ajustaron con la naturaleza de la misma.

Considerando los resultados obtenidos en el modelamiento de la información de la serie de tiempo producción total de petróleo por la metodología BJR, un análisis variográfico para esta serie fue desarrollado, como punto de partida para el análisis geoestadístico.

De acuerdo a los resultados obtenidos de la predicción de la serie a futuro se concluye que la estimación geoestadística reproduce la tendencia y el componente estacional de los datos que no se evidenció en la aplicación de la metodología BJR.

El análisis de los errores de las estimaciones de los datos o valores faltantes demostró que estos se distribuyeron normalmente y que son incorrelacionados; en consecuencia, son independientes.

RECOMENDACIONES

Para el análisis de la producción total de petróleo, objeto de este estudio, se probaron tres tipos de modelos que representan a los más comúnmente empleados en geoestadística debido a su flexibilidad: Esférico, Gaussiano y Exponencial; sin embargo, de acuerdo a la literatura existen muchos y diferentes semivariogramas teóricos que podrían ser evaluados buscando mejorar los resultados obtenidos.

Para la disciplina de la ingeniería de petróleos, se debe profundizar en la aplicación de las técnicas de estimación geoestadísticas en el dominio temporal. La revisión de los antecedentes para este trabajo permitió identificar a la autora, que es una metodología ampliamente usada en el dominio espacial pero muy poco explorada en el dominio temporal.

REFERENCIAS BIBLIOGRÁFICAS

- Arps, J. (1944). Analysis of decline curves. Transactions of the AIME, 160(1), 228-247.
- Baca, J. (2016). A geostatistical method for the analysis and prediction of air quality time series: Application to the Aburrá Valley Region (Master's Thesis). The Technische Universität München, Germany.
- Blaskovich, F. (April of 2000). Historical Problems with Old Field Rejuvenation, Society of Petroleum Engineering Asia Pacific Conference on Integrated Modelling for Asset Management , Yokohama, Japan.
- Box G., Jenkins G. y Reinsel G. (2008). Time Series Analysis: Forecasting and Control, John Wiley & Sons.
- Brockwell, P., y Davis, R. (Ed.). (1991). Time Series: Theory and Methods, Springer Verlag.
- Bruno, R., y De Fouquet Ch. (2005). The expansión of environmental geostatistics. En D. Krige, W. Kleingeld, M. Bilodeau, F. Meyer, y M. Schmitt (Eds.), Lecture Notes in Statistics 183. (pp.9-22). Springer-Verlag.
- Cressie N. (1993). Statistical for Spatial Data. John Wiley & Sons.
- Cowpertwait, P., y Metcalfe, A. (2009). Introductory Time Series with R. Springer.
- Cryer, J., y Chan, K. (Ed.). (2008). Time Series Analysis with applications in R. Springer.
- De Iaco S., Palma M. y Posa D. (2013). Geostatistics and the Role of Variogram in Time Series Analysis: A Critical Review. En S. Montrone y P. Perchinunno P (Eds.), Statistical Methods for Spatial Planning and Monitoring. (pp.47-73). Springer-Verlag.
- Faraway, J. (Ed.). (2015). Linear Models with R, CRC Press, Taylor and Francis Group.

- Giraldo, R., Pacheco, O., y Orozco, A. (2017). Geoestadística Aplicada A Series De Tiempo Autorregresivas: Un estudio De Simulación. *Revista Integración, Escuela de Matemáticas, Universidad Industrial de Santander*. 35(1), pp. 83–102.
- Hadfield, J.D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* 33 2 1-22
- Handayani, H., y Simamora, J. (2012). Challenge in Mature Field's Waterflood Project: Facing Limited GGRP Data & Budget Issue, *Society of Petroleum Engineering North Africa Technical Conference and Exhibition, Cairo, Egypt*.
- Haslett, J. (1997). On the sample variogram and sample autocovariance for non-stationary time series. *Statistician* 46(4), pp. 475-485.
- Höök, M., Davidsson, S., Johansson, S., y Tang, X. (2014). Decline and depletion rates of oil production: A comprehensive investigation, *Phil. Trans. R. Soc. A* 372(20120448). <http://dx.doi.org/10.1098/rsta.2012.0448>
- Hurtado de Barrera, J. (2000). El proyecto de investigación: Metodología de la investigación holística. *Sypal, Ediciones Quirón*.
- Journel, A., y Huijbregts, C. (1978). *Mining geostatistics: Academic press*.
- Khachatryan, D., y Bisgaard, S. (2009). Some results on the variogram in the time series analysis. *Qual. Reliab. Eng. Int.* <http://doi.org/10.1002/qre.1013>
- Ma, C. (2004). The use of the variogram in construction of stationary time series models. *J. Appl. Probab.* 41(4), pp.1093-1103.
- Medina, F., y Galván, M. (2007). Imputación de Datos: Teoría y Práctica, *Serie: Estudios Estadísticos y Prospectivos CEPAL* (54), p.7.

- Montes, E. (2016). Modelado de los pronósticos de producción a partir del análisis integrado de datos y métodos estadísticos no paramétricos (Tesis de maestría), Universidad Industrial de Santander, Bucaramanga, Colombia.
- Montes, E., Calvete, F., y Mantilla, C. (2016). Aplicación de series de tiempo en la realización de pronósticos de producción. *Revista Fuentes: El Reventón Energético*, 14(1), 79 – 88.
- Olominu, O., y Sulaimon, A. (2014). Application of Time Series Analysis to Predict Reservoir Production Performance. Society of Petroleum Engineering Nigeria Annual International Conference and Exhibition. Lagos, Nigeria.
- Pebesma, E. (2004). Multivariable geostatistics in S: the gsstat package. *Computers & Geosciences*, 30, 683 – 691.
- Peña, D. (2010). *Análisis de Series Temporales*, Alianza Editorial.
- Schabenberger O., y Gotway C. (2005). *Statistical Methods for Spatial Data Analysis*, Chapman & Hall.
- Spiegelhalter, D., Best, N., Carlin, B., y Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B* (64), 583-639.
- Yaglom, A. (1987). *Correlation Theory of Stationary and Related Random Functions*. Springer, New York.