

Optimización de la ubicación de un nuevo pozo infill productor mediante el uso de simulación numérica de yacimientos y modelos proxy. Aplicado a un campo colombiano.

Rafael David Llanos Castellanos

Diego Enrique Berdugo González

Trabajo de Grado para Optar por el Título de Ingeniero de Petróleos

Director

Andrés Felipe Ortiz Meneses M.Sc.

Codirector

Carlos Alfonso Mantilla Duarte Ph.D.

Universidad Industrial de Santander

Facultad de Ingenierías Fisicoquímicas

Escuela de Ingeniería de Petróleos

Bucaramanga

2017

Dedicatoria

A mi madre, por ser mi más grande motivación. Por enseñarme con su ejemplo, a dar todo de mí. Por inspirarme con su apoyo incondicional a llenarme de fortaleza para afrontar y superar los obstáculos que la vida me brinda. Y a nunca darme por vencido.

A mi hermano, por ser aquella persona que me ha demostrado que el emprendimiento y la disciplina son claves para lograr las metas propuestas. Por demostrarme que la perseverancia sí genera resultados gratificantes y por creer que yo puedo generar un impacto positivo en este mundo.

A esa persona que me vio nacer y ya no se encuentra conmigo. Que, a pesar de ser una señorita difícil, me enseñó que la bondad y el amor son más grandes que cualquier defecto, siempre permanecerá en mis recuerdos, con su apoyo incondicional me ayudo a salir adelante.

A mi amigo Diego, una persona brillante con un gran futuro por delante, que a lo largo de la carrera forjamos una gran amistad.

A todos aquellos que me brindaron su amistad, confianza y afecto durante mi estadía en la Universidad.

Por último, a la vida, por permitirme estar rodeado de personas con grandes capacidades e inteligencia y de quienes aprendo todos los días.

Muchas gracias...

Rafael David Llanos Castellanos

Dedicatoria

Madre, incansable luchadora, mujer de firme carácter, pero llena de un inmenso amor, a ti dedico este y todos mis logros venideros. Por tu apoyo incondicional, tus sabios consejos y tu valioso ejemplo, por aquello que con el mayor de los esfuerzos y cariño me has brindado y me ha permitido estar aquí hoy, por enseñarme a seguir adelante sin importar los obstáculos, a luchar por mis sueños, a enfrentarme a la vida y a no rendirme jamás.

A mi hermano, hombre inigualable y bondadoso, por tu constante apoyo, por tu desinteresado cariño, por tus siempre acertadas palabras, por demostrarme con tu ejemplo que el esfuerzo, la disciplina y el trabajo duro son la base para un éxito constante, por creer y confiar en mí y recordarme que he llegado a este mundo para grandes cosas.

A Karina A. Bolaños, mujer inteligente y tenaz, mi compañera de vida, conocedora de mis batallas, mis tristezas y mis alegrías, por tu tierno cariño, por tu invaluable apoyo, por ser un pilar que se alza cada vez con más fuerza, por los momentos vividos, por las sonrisas compartidas, por inspirarme a ser siempre mejor, por lo que fue y por lo que vendrá.

A Rafael Llanos, mi gran amigo, hombre perseverante como ningún otro, con un futuro brillante ante sus ojos, por esa incondicional amistad.

Y a todos aquellos que han pasado por mi vida, me han brindado su afecto y han dejado una huella en mí.

Gracias...

Diego Enrique Berdugo González

"Sólo es posible avanzar cuando se mira lejos. Solo cabe progresar cuando se piensa en grande." José Ortega y Gasset

Agradecimientos

Al profesor Andrés, por darnos la oportunidad de pertenecer al grupo de modelamiento en procesos de hidrocarburos, y por orientarnos incondicionalmente en el proceso de Investigación.

Al profesor Carlos, por guiarnos durante la última etapa del desarrollo de nuestro trabajo de grado, incorporando ideas innovadoras y oportunas para la culminación del proyecto.

A ECOPETROL por brindarnos uso de su información, parte indispensable para el proyecto.

A la Universidad Industrial de Santander y a la Escuela de Ingeniería de Petróleos.

A todas las personas que de algún modo estuvieron involucradas con la realización de este proyecto.

Tabla de contenido

Introducción	14
1. Antecedentes de la investigación	16
1.1 Perforación Infill	18
1.2 Simulación numérica de yacimientos	18
1.3 Función Objetivo	19
1.4 Algoritmos de optimización	21
1.4.1 Programación lineal.	21
1.4.2 Método del gradiente.	22
1.4.3 Método probabilístico y optimización estocástica.	22
1.5 Generalidades de los modelos Proxy	22
1.5.1 Modelo regresión polinomial.	23
1.5.2 Modelos kriging.	24
1.5.3 Modelos Splines de placa delgada.	24
1.5.4 Redes neuronales artificiales.	24
2. Modelo del sector	27
2.1 Modelo del campo	28
2.2 Modelo del sector	30
3. Optimización de la ubicación de un nuevo pozo	33
3.1 Simulación del sector	33
3.2 Optimización con CMOST	33
3.3.1 Variables financieras.	34
3.3.2 Hard y Soft constraints.	35
3.3.3 Algoritmos de optimización.	36
4. Desarrollo del modelo proxy	44
4.1 Modelamiento proxy	44
4.1.1 Variables de entrada.	45
4.1.2 Análisis de Sensibilidad.	45

4.1.3 Muestreo del set de datos de entrada.	45
4.1.4 Estimación del modelo proxy.	47
4.1.5 Comparación de tiempos de simulación.	47
4.1.6 Comparación de resultados de modelos.	47
5. Validación del modelo Proxy y análisis complementario	51
5.1 Diferencia de Medias (Prueba de Welch)	51
5.2 Error Porcentual	53
5.3 Aplicación de la curva ROC	55
5.4 Análisis complementario de la data espacial	57
6. Conclusiones	60
7. Recomendaciones	62
Referencias bibliográficas	63

Lista de figuras

Figura 1. Infraestructura de optimización mediante simulación	17
Figura 2. Esquema de una red neuronal artificial	25
Figura 3. Procesamiento de nodos	25
Figura 4. Modelo del campo colombiano de interés	28
Figura 5. Ubicación del Bloque 5 en el modelo de campo	31
Figura 6. Modelo del sector de interés	31
Figura 7. Modelo del sector de interés en CMG	32
Figura 8. Ubicación óptimo global del nuevo pozo vertical	37
Figura 9. Resultados del algoritmo Random Brute Force para la zona 1	38
Figura 10. Resultados del algoritmo Random Brute Force para la zona 2	38
Figura 11. Resultados del algoritmo Random Brute Force para la zona 3	39
Figura 12. Resultados del algoritmo Random Brute Force para la zona 4	39
Figura 13. Tasa de aceite mensual a condiciones estándar caso valor óptimo global del sector	40
Figura 14. Tasa de agua mensual a condiciones estándar caso valor óptimo global del sector	40
Figura 15. Aceite acumulado a condiciones estándar caso valor óptimo global del sector	41
Figura 16. Agua acumulada a condiciones estándar caso valor óptimo global del sector	41
Figura 17. Resultados del algoritmo Designed Exploration Controlled Evolution	42
Figura 18. Resultados del algoritmo Particle Swarm Optimization	42
Figura 19. Resultados del algoritmo Differential Evolution	43
Figura 20. Diagrama de flujo para el modelamiento proxy	44
Figura 21. Gráfica de los Resultados del modelo base vs. Resultados de R. P.	48
Figura 22. Gráfica de los Resultados del modelo base vs. Resultados de la R.N.A.	49
Figura 23. Gráfica del V.P.N. vs. Número total de coordenadas para el modelo de R. P.	49
Figura 24. Gráfica del V.P.N. vs. Número total de coordenadas para el modelo de R.N.A.	50
Figura 25. Ejemplos de curvas ROC	56
Figura 26. Grafica de los Resultados de la curva ROC	57

Lista de tablas

Tabla 1. Resumen de datos geológicos del campo de interés	29
Tabla 2. Resumen de datos petrofísicos del campo de interés	30
Tabla 3. Variables financieras para optimización de la función objetivo	35
Tabla 4. Variables estadísticas para el muestreo del sector de interés	46
Tabla 5. Tiempos de simulación empleados por cada método de optimización	47
Tabla 6. Resultados del valor P obtenido en cada prueba para los modelos proxy	53
Tabla 7. Errores porcentuales de los valores arrojados por cada modelo proxy en comparación con el resultado óptimo de la función objetivo del modelo base	54

RESUMEN

TÍTULO: OPTIMIZACIÓN DE LA UBICACIÓN DE UN NUEVO POZO INFILL PRODUCTOR MEDIANTE EL USO DE SIMULACIÓN NUMÉRICA DE YACIMIENTOS Y MODELOS PROXY. APLICADO A UN CAMPO COLOMBIANO*

AUTORES: RAFAEL DAVID LLANOS CASTELLANOS & DIEGO ENRIQUE BERDUGO GONZALEZ**.

PALABRAS CLAVE: Pozo infill, Modelo proxy, Simulación numérica de yacimientos, Algoritmo de optimización, Curva ROC, I de Moran

DESCRIPCIÓN:

La optimización de la ubicación de un nuevo pozo es una tarea desafiante y compleja, y es un factor determinante para el exitoso desarrollo de un campo. Tradicionalmente cientos o inclusive miles de ubicaciones deben ser consideradas a fin encontrar la más óptima. El gran número de grados de libertad hace que utilizar un proceso de simulación numérica de yacimientos convencional sea una tarea de un alto consumo de tiempo y recursos económicos.

A pesar de que los recientes progresos en el desarrollo de hardware y software computacional han abierto nuevas fronteras en el área de simulación, la disponibilidad de estos recursos es aún vista como un factor limitante, así pues, la definición de un modelo proxy se propone como la solución a este problema y su aplicación conjunta a la labor convencional de simulación permite la realización de la tarea de optimización en tiempos más cortos generando menores costos y brindando la ubicación del nuevo pozo que permitirá el incremento de la producción.

Se optimiza la ubicación de un nuevo pozo para un sector de un modelo de simulación representativo de un campo colombiano, se realiza simulación y optimización mediante el uso de software especializado aplicando algoritmos de búsqueda y se propone un modelo proxy que replique la respuesta de la tarea de simulación. La aplicación del modelo proxy permite ahorrar tiempo en comparación a la optimización con la simulación completa. Este estudio se desarrolló en el marco del proyecto de investigación 531-2011 suscrito entre la UIS, Ecopetrol S.A. y Colciencias.

* Trabajo de Grado

** Facultad de Ingenierías Físico-Químicas. Escuela de Ingeniería de petróleos. M. Sc. Andrés Felipe Ortiz Meneses, Ph. D. Carlos Alfonso Mantilla Duarte.

ABSTRACT

TITLE: WELL PLACEMENT OPTIMIZATION OF A NEW INFILL PRODUCER WITH THE USE OF RESEVOIR NUMERIC SIMULATION AND PROXY MODELS. APPLIED TO A COLOMBIAN FIELD.*

AUTHORS: RAFAEL DAVID LLANOS CASTELLANOS & DIEGO ENRIQUE BERDUGO GONZALEZ**

KEYWORDS: Infill well, Proxy model, Reservoir numeric simulation, Optimization algorithm, ROC Curve, Moran´s I.

DESCRIPTION:

Well placement optimization of a new well is a complex and challenging task, and it is a decisive element for the successful development of a field. Traditionally, hundreds or even thousands of locations must be considered to find the most optimum. The great amount of degrees of freedom makes a conventional numeric reservoir simulation process become a task of high time consumption and of substantial amount of economic resources.

Even though recent advancements in hardware and computational software development have opened new frontiers in the area of simulation, the availability of these resources is still seen as a limiting factor, therefore, the definition of a proxy model is proposed as a solution to this problem which allows the execution of the optimization task along with its joint application of the conventional simulation task in shorter times resulting in lower costs and providing a new well location which will increase production rates.

A well placement optimization of a new well is done in a representative sector simulation model of a Colombian field. The simulation and optimization processes are done with specialized software applying search algorithms and a proposal of a proxy model that can reproduce the response of the simulation task is made. The application of a proxy model will allow to save time compared to a full simulation optimization. This study was made in the context of research project 531-2011 endorsed between UIS, Ecopetrol S.A. and Colciencias.

* Bachelor Thesis

** Facultad de Ingenierías Físico-Químicas. Escuela de Ingeniería de petróleos. M. Sc. Andrés Felipe Ortiz Meneses, Ph. D. Carlos Alfonso Mantilla Duarte.

Introducción

La optimización de la ubicación de un pozo es una tarea desafiante y compleja, y es un factor determinante para el exitoso desarrollo de un campo. El desempeño de un yacimiento es altamente dependiente de las ubicaciones de los pozos, las cuales, deben ser determinadas basadas en las propiedades del reservorio y sus fluidos, así como en las especificaciones de los sistemas de producción.

Tradicionalmente cientos o inclusive miles de ubicaciones deben ser consideradas a fin de encontrar la más óptima, acorde con los objetivos de producción. El gran número de grados de libertad hace que utilizar un proceso de simulación numérica de yacimientos convencional sea una tarea de un alto consumo de tiempo, esfuerzo computacional y recursos económicos.

Actualmente, a pesar de que los recientes progresos en desarrollo de hardware y software computacional han abierto nuevas fronteras en el área de simulación de yacimientos, la disponibilidad de estos recursos de computación es aún vista como un factor limitante en estos procesos, por lo cual, esta investigación afrontará este problema y estudiará, mediante el uso de simulación numérica junto a modelos proxy, la forma de sortear estas barreras y facilitar la realización de esta labor en tiempos más cortos y a menores costos.

Este trabajo presenta los siguientes capítulos:

En el capítulo I se identifican las distintas metodologías ya empleadas en otros estudios de optimización en la ubicación de pozos donde se consideran las variables determinantes y la función

objetivo para este procedimiento, dando una base referente para el desarrollo del proceso previamente mencionado.

El capítulo II está compuesto por el procedimiento para establecer un modelo apropiado del sector del campo colombiano de interés abarcando información relevante y pertinente para el proceso de optimización. Posteriormente en el capítulo III se realiza, con base en el modelo del sector de campo previamente establecido, la simulación numérica de todas las posibles ubicaciones para el nuevo pozo infill y mediante el uso de software especializado la optimización de este mismo.

Para abordar la parte más relevante de la investigación, en el capítulo IV se proponen dos modelos proxy partiendo de la información del modelo del sector de campo ya constituido, buscando replicar los resultados obtenidos por medio de la simulación.

Finalmente, en el capítulo V se ofrece una comparación de los resultados obtenidos por medio de la simulación con los obtenidos de los modelos proxy propuestos y a su vez se realiza una validación de su eficacia en el proceso de optimización.

1. Antecedentes de la investigación

El proceso de optimizar la ubicación de un nuevo pozo *infill* presenta una tarea desafiante ya que este consiste en realizar una búsqueda exhaustiva, mediante el uso de diferentes técnicas, para hallar el lugar donde el nuevo pozo genere el mejor rendimiento para la explotación del área de interés. Una práctica común en la industria es la optimización mediante el método manual, donde el ingeniero toma un juicio basado en su experiencia y conocimiento, y además se apoya mediante la simulación de yacimientos para determinar la ubicación más apropiada (Pallav & Wen, 2008).

La optimización combinatoria es un término aplicado hacia problemas donde las variables de decisión, aquellos aspectos del problema que pueden ser manipulados, son discretos en vez de ser continuos. Se debe comprender que, al trabajar con variables de decisión discretas, estas complican la optimización porque limitan el uso de las técnicas más eficientes de búsqueda del valor óptimo. Aunque exista este contratiempo, es posible de encontrar el valor óptimo global con los recursos computacionales modernos donde su tiempo depende del número de casos que necesiten ser evaluados para que la función objetivo sea optimizada (Nikraves & Zadeh, 2003).

La optimización por simulación estocástica surge de problemas cuya naturaleza es estocástica, y la meta del proceso es encontrar los valores de las variables de decisión óptimos que sean de interés y cuya simulación es de eventos discretos (Beers, 2007). Para problemas de optimización determinística, la aleatoriedad o incertidumbre no es considerada, por esta razón sus soluciones son relativamente simples. Esto se debe a que la incertidumbre no es crítica. Sin embargo, en el mundo real, las incertidumbres si son críticas en el diseño, análisis y operación de los sistemas modernos. A consecuencia de lo anterior, se tiene que tratar con modelos o sistemas de no

linealidad, dimensionamiento y complejidad altos. Cualquier intento de usar modelos analíticos en dicho sistema implica simplificar suposiciones (Chen & Lee, 2011)

La infraestructura de optimización mediante simulación consiste en tres niveles de operación, el modelo de simulación, el simulador estocástico y la optimización/búsqueda. El modelo de optimización es el núcleo de la infraestructura y su papel es de evaluador de desempeño. Para cada configuración de las variables de decisión, múltiples repeticiones de simulación o muestras deben ser desarrolladas para poder capturar la propiedad de aleatoriedad y obtener un estimado estadístico. Luego el simulador estocástico realiza un bucle de repeticiones múltiples del modelo de simulación. Y, por último, el nivel más alto es la optimización o buscador, que itera la búsqueda de un espacio solución para encontrar la mejor configuración de las variables de decisión, donde el simulador estocástico es aplicado para evaluar diferentes candidatos alternativos (Chen & Lee, 2011). La Figura 1 adaptado de (Chen & Lee, 2011) resume la infraestructura.

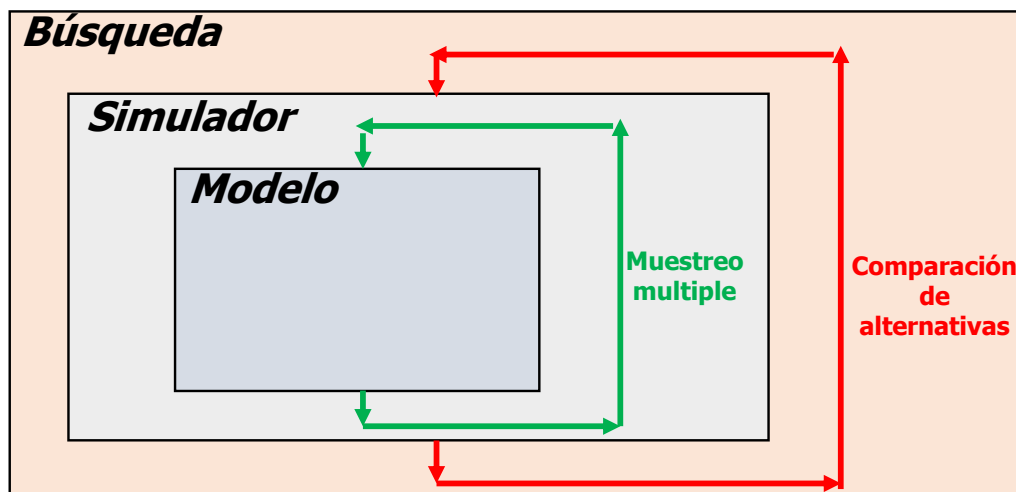


Figura 1. Infraestructura de optimización mediante simulación

El problema que conlleva el optimizar la ubicación de un pozo radica en su misma naturaleza, la no linealidad, ya que se encuentran presentes variables, restricciones y un número indefinido de escenarios que pueden impactar de manera directa o indirecta en la solución óptima final. Generalmente el proceso parte con la selección de una posición inicial basado en un juicio ingenieril con un software de simulación numérica, luego se determina una función objetivo para que junto con la ayuda de un algoritmo de optimización se realice una búsqueda basado en los resultados obtenidos del simulador en conjunto con la metodología que dicho algoritmo posee y luego este mismo determina la ubicación con las mejores características para la producción de un nuevo pozo.

1.1 Perforación Infill

Hace referencia la adición de pozo(s) en un campo, disminuyendo la separación promedio de los pozos. Esta práctica acelera la recuperación esperada e incrementa la recuperación estimada final en yacimientos heterogéneos mejorando la continuidad entre inyectores y productores. A medida que disminuye la separación entre pozos, los patrones de pozos de desplazamiento alteran las trayectorias de flujo de fluidos de la formación e incrementan el barrido en áreas donde existe mayor saturación de hidrocarburos (Schlumberger, 2017).

1.2 Simulación numérica de yacimientos

La tecnología de simulación numérica de yacimientos surge del deseo de tratar en forma adecuada un yacimiento con algún grado de exactitud. Su aplicación es diversa y amplia, permitiendo el estudio de fenómenos locales alrededor de la cara del pozo y ha demostrado ser una herramienta

fundamental para el ingeniero de petróleos. La tecnología permite evaluar de manera segura el comportamiento de un esquema de explotación de un campo o inclusive de un sector del mismo.

El simulador numérico está integrado por tres modelos: el diferencial, el numérico y el de computador. El modelo diferencial está conformado por el conjunto de ecuaciones diferenciales que describen los procesos fisicoquímicos que ocurren en el yacimiento en función del espacio y del tiempo. El modelo numérico constituye la forma en que se da solución al modelo diferencial. Para dar solución al número de ecuaciones se requiere la elaboración de un programa de computador conocido como el modelo de computador del simulador (Osorio Gallego, 2002). La información básica que se requiere para que el modelo de computador funcione y genere los resultados deseados son los siguientes:

- Geometría del yacimiento
- Ubicación de pozos productores e inyectores con sus respectivas tasas de flujo
- Propiedades de la roca y del fluido en cada bloque
- Distribución inicial de fluidos en el yacimiento
- Presión promedio en cada nodo en el tiempo cero

1.3 Función Objetivo

Para la ingeniería, la función objetivo se define como aquella función que incorpora variables de decisión que permiten establecer una rentabilidad óptima de un proceso. Su definición es fundamental en cualquier proceso de optimización y por esta razón su determinación al inicio del proceso es clave. Para la ubicación de un nuevo pozo *infill*, la función Valor Presente Neto (V.P.N.) es la más recomendada ya que relaciona la producción de todos los fluidos del yacimiento (Al-Mudhafar, Al-Jawad, & Al-Shamma, 2010). El V.P.N. es una función financiera de variables

económicas que permite representar ingresos y egresos. En términos de la industria de los hidrocarburos, los ingresos equivalen a la producción de aceite y gas acumulado que ha sido pronosticado por un modelo de yacimiento. Y los egresos equivalen a los costos asociados por fluidos producidos diferente al aceite y el gas, además de los gastos asociados por el CAPEX y OPEX, gastos generados por inversión de capital y de operación respectivamente (Al-Mudhafer, 2013).

La ecuación (1) describe la diferencia de los ingresos y egresos y se conoce como Flujo de Caja Neto (F.C.N.). Esta ecuación hace parte y permite definir el V.P.N. mediante la ecuación (2).

$$F.C.N. = [q_o * P_o + q_g * P_g] - [q_w * C_w + q_i * c_i + OPEX + CAPEX] \dots \dots \dots (1)$$

Donde:

q_o : Producción de aceite acumulada (STB)

P_o : Precio de venta del barril de aceite (\$ USD / STB)

q_g : Producción de gas acumulada (MSCF)

P_g : Precio de venta del volumen de gas (\$ USD / MSCF)

q_w : Producción de agua acumulada (bbl)

C_w : Costo asociado al tratamiento de agua acumulada producida (\$ USD / bbl)

q_i : Agua inyectada acumulada (bbl)

c_i : Costo asociado al agua acumulada inyectada (\$ USD / bbl)

$OPEX$: Gastos asociados a la operación (\$ USD)

CAPEX : Gastos asociados al capital (\$ USD)

$$V.P.N. = \sum_t \frac{F.C.N.}{(1+i)^{TO}} \dots\dots\dots (2)$$

Donde:

V.P.N. : Valor Presente Neto (\$ USD)

F.C.N. : Flujo de Caja Neto (\$ USD)

i : Tasa de oportunidad (%)

TO : Tiempo de la producción (años o meses)

1.4 Algoritmos de optimización

Resolver el problema que conlleva la optimización de la ubicación de un nuevo pozo *infill* ha llevado a diferentes autores a recurrir a distintos métodos. Cada método posee diferentes limitaciones y manejo de la función objetivo, pero todos buscan obtener la ubicación más óptima de un nuevo pozo.

1.4.1 Programación lineal. El método más antiguo de optimización para la ubicación de un pozo ha sido los algoritmos con base en programación lineal. Estos pasos automatizados parten de la programación de enteros mixtos que maximiza mediante relaciones matemáticamente lineales a la función objetivo. Sin embargo, este método carece de aplicaciones para las condiciones de flujo no lineales que se presenta en la mayoría de la producción de pozos, limitando la solución más óptima (Belegundu & Chandrupatla, 2011).

1.4.2 Método del gradiente. En este método, una solución máxima o mínima de una función objetivo de n-dimensiones es determinada a través de la aplicación de la expansión de las series de Taylor y el cálculo del gradiente de la función objetivo. Las dos aproximaciones que se pueden usar son el método de diferencias finitas o por el método adjoint. Desafortunadamente, la limitación de este método radica en su alta posibilidad de quede atrapado en un óptimo local (Beers, 2007).

1.4.3 Método probabilístico y optimización estocástica. El método probabilístico de Hibridación simulada o S.A. por sus siglas en inglés, permite ubicar el óptimo global imitando el proceso de recristalización de un objeto sólido que ha sido calentado y luego enfriado a una estructura congelada teniendo en cuenta una configuración de energía mínima. El método de optimización estocástico Optimización por Enjambre de Partículas o P.S.O. por sus siglas en inglés, se basa en poblaciones que imita los esfuerzos de búsqueda que conforman los grupos de animales en la naturaleza, dando muy buenos resultados en trabajos recientes de ubicación de pozo. Existen otros algoritmos que se basan en poblaciones y que al igual que el P.S.O. han presentado resultados óptimos tales como Evolución Dirigida mediante Evolución Controlada o D.E.C.E. por sus siglas en inglés, Algoritmo de Competencia Imperialista o I.C.A. por sus siglas en inglés y Evolución Diferencial o D.E. por sus siglas en inglés (Al Dossary & Nasrabadi, 2015).

1.5 Generalidades de los modelos Proxy

Se hace referencia a un modelo proxy como una función definida matemática o estadísticamente, que replica la salida de un proceso para parámetros de entrada seleccionados (Temizel , et al.,

2014). También se le conoce como un “meta-modelo” o “modelo *surrogate*”. Estos modelos típicamente incluyen en simulación numérica lo siguiente:

- Análisis de sensibilidad para variables inciertas
- *Forecasting* (ajuste histórico probabilístico) y análisis de riesgo
- Condicionamiento del modelo de simulación a data observada (Ajuste histórico)
- Planeación del desarrollo del campo y optimización en la producción

Los retos de crear modelos proxy de alta calidad están relacionados con la calidad de los datos de entrada. La simulación de yacimientos trata con la salida altamente no lineal. Por lo tanto, la data de entrada con experimentos uniformemente distribuidos sobre un rango de incertidumbre puede no ser suficiente para la construcción de un modelo proxy adecuado. Para sobreponer esta limitación, existen varias técnicas propuestas (Zubarev, 2009). Dentro de las que se han usado en la industria de los hidrocarburos, se encuentran modelos de regresión polinomial (P.R. por sus siglas en inglés), modelos Kriging multivariable (KG. por sus siglas en inglés), modelos Splines de placa delgada (T.P.S. por sus siglas en inglés) y Redes neuronales artificiales (A.N.N. por sus siglas en inglés).

1.5.1 Modelo regresión polinomial. Los modelos de regresión polinomial son comúnmente referidos como modelos de superficie de respuesta (Kuehl, 2001). Estos fueron propuestos por primera vez para el análisis de experimentos físicos y luego fueron adoptados también para experimentos computacionales. Aun si este tipo de modelo no se aproxima exactamente a la data experimental, los modelos de regresión polinomial han sido ampliamente adoptados en la industria de los hidrocarburos debido a su facilidad de comprenderlos, su flexibilidad y eficacia

computacional. Es importante considerar que este tipo de modelo proxy tiene un pobre desempeño para espacios multidimensionales altamente no lineales (Temizel , et al., 2014).

1.5.2 Modelos kriging. Los modelos *kriging* están basados en técnicas geoestadísticas para correlación espacial aleatoria *Kriging*. Este tipo de modelo proxy replica exactamente la muestra de datos inicial lo que lo hace un candidato muy atractivo para experimentos computacionales. El tiempo requerido para construirlo depende del número de parámetros de entrada, el número de experimentos de la muestra, la función de aproximación usada y la CPU. Culinado la construcción del modelo *kriging*, este corre igual de rápido como cualquier otro tipo de modelo proxy (Sacks, Welch, Mitchell, & Wynn, 1989).

1.5.3 Modelos Splines de placa delgada. Los *Splines* de placa delgada son un esquema de interpolación multidimensional para data arbitrariamente espaciada. Modelos de este tipo reproducen exactamente la data de entrada, pero requieren que el número de experimentos de la muestra sea mayor al número de los datos de entrada. Este modelo está conformado por dos partes, una función de regresión de aproximación global y funciones de base radial que define el mapeo espacial entre puntos en el espacio solución (Li & Friedmann, 2005).

1.5.4 Redes neuronales artificiales. La red neuronal artificial es una emulación del sistema neuronal biológico. Una red neuronal artificial consiste en elementos básicos; nodos, análogos a neuronas en sistemas biológicos. Cualquier nodo recibe señales de nodos vecinos y los procesa para proveer una sola salida. Un proceso esquemático de una red neuronal artificial con múltiples niveles ocultos se observa en la Figura 2 adaptado de (Zubarev, 2009).

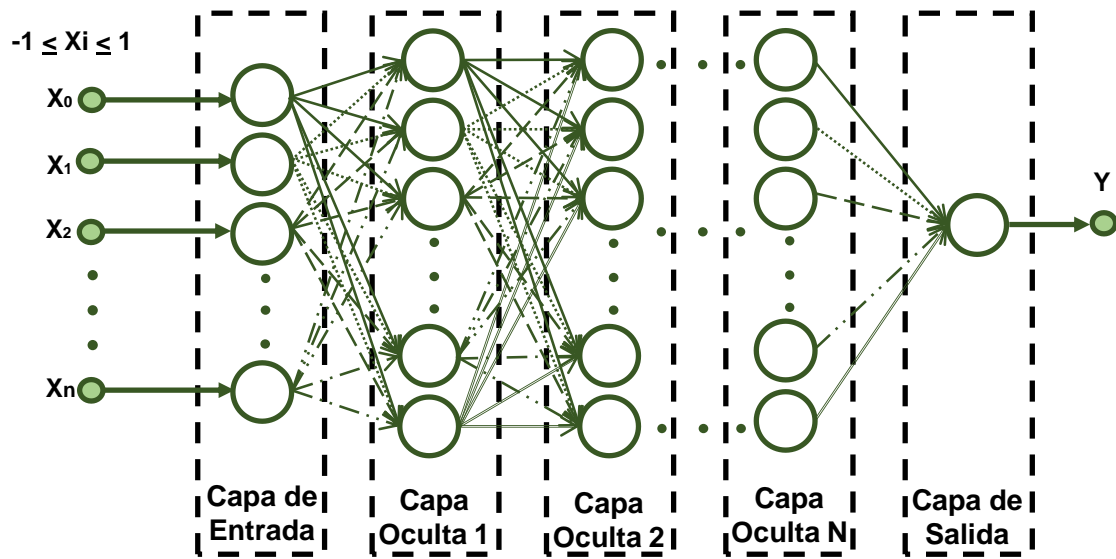


Figura 2. Esquema de una red neuronal artificial

Las señales, valores para las variables de entrada, pasan de los nodos al nivel de entrada hacia los nodos del nivel de salida mediante múltiples nodos de niveles ocultos. Cada nodo de un nivel oculto procesa las señales como lo muestra la Figura 3 adaptado de (Zubarev, 2009).

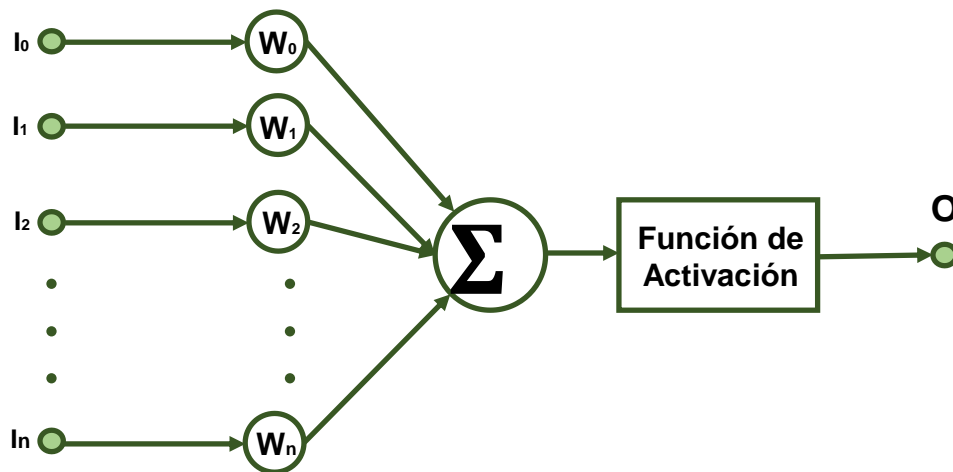


Figura 3. Procesamiento de nodos

Cada nodo de los niveles ocultos recibe múltiples señales de entrada de los nodos del nivel anterior, que puede ser el nivel de entrada u otro nivel oculto. Las señales recibidas son multiplicadas por sus pesos correspondientes y luego es sumado. Posteriormente, la función de activación es usado para definir su topología; número de niveles ocultos, numero de nodos por nivel oculto y la función de activación. Después los pesos de entrada deben ser estimados para cada nodo. El número de niveles ocultos y nodos influyen en la habilidad de la red neuronal para reproducir diferentes grados de no linealidad. Sin embargo, el número de nodos ocultos está restringido por el número de experimentos usados en una construcción de una red neuronal artificial. Una vez la topología de la red neuronal artificial es definida, los pesos desconocidos del nodo pueden ser estimados a través de un algoritmo de aprendizaje supervisado. El algoritmo de aprendizaje se basa en el error de propagación de retorno, este permite actualizar los pesos del nodo mediante un proceso iterativo que se basa en la disparidad de la salida de la red neuronal artificial con los valores de salida del set de datos de entrenamiento para luego ajustar los pesos y minimizar la disparidad (Cullick, Johnson, & Shi, 2006).

2. Modelo del sector

Un modelo se considera como una entidad que permite el estudio de un fenómeno, bajo condiciones de prueba apropiadas, que tiene la probabilidad de que ocurra en la práctica. En general los modelos pueden clasificarse como físicos y matemáticos. Los modelos físicos son representaciones de laboratorio con tendencias a reproducir lo que ocurre en el yacimiento, y las ecuaciones que describen de manera teórica este comportamiento constituyen el modelo matemático (Osorio Gallego, 2002).

Ambos modelos juegan un papel importante. Por ejemplo, las leyes que gobiernan el flujo de fluidos en un medio poroso fueron descubiertos y delineados empleando modelos físicos. La ley de Darcy, los conceptos de permeabilidad relativa, presión capilar, densidad y correlaciones de viscosidad, entre otras, tienen sus orígenes en experimentos con modelos físicos. Sin embargo, estos tienen sus limitaciones: es poco práctico el modelamiento riguroso de sistemas de gran escala como lo es un yacimiento de petróleo, por lo que podría decirse que los modelos físicos son muy útiles en el estudio de fenómenos a pequeña escala (Osorio Gallego, 2002). Dentro de los modelos físicos, se aplica una sub-clasificación según el tamaño que, de mayor a menor escala, podemos encontrar los siguientes: yacimiento, campo, sector y laboratorio. Este trabajo se enfoca en el modelo de sector.

2.1 Modelo del campo

Para constituir el modelo del sector del campo colombiano de interés, primero hay que entender que es un modelo estático. Un modelo estático se puede definir como aquel modelo que contiene información básica como: fallas, propiedades petrofísicas, pozos etc.; poblado en una malla de simulación de un tamaño predeterminado y donde no se ha evaluado el comportamiento de la información a través del tiempo. El modelo estático bien definido y poblado del campo con el que se contó para desarrollar este trabajo, fue facilitado por la empresa operadora del campo. La Figura 4 muestra una visualización del modelo estático del campo colombiano.

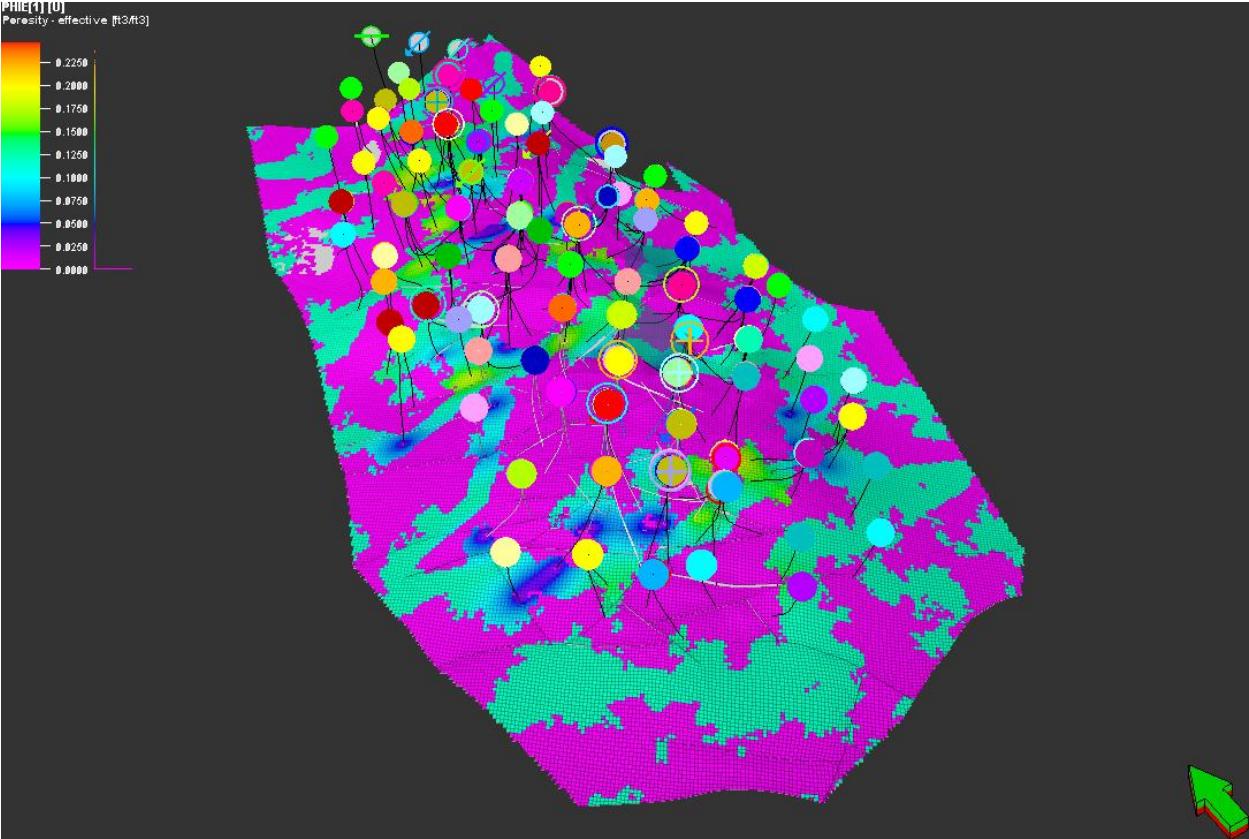


Figura 4. Modelo del campo colombiano de interés

El modelo de simulación del campo tiene una geometría *Corner point* y posee un área superficial aproximada de 11253 acres, ubicado en un yacimiento heterogéneo. Este modelo contiene 8 bloques de pozos, 81 fallas y aproximadamente un total de 230 pozos distribuidos en los sectores de los cuales 227 pozos son productores y 3 pozos inyectores. Por cuestiones de confidencialidad, algunos de los datos del modelo de campo se renombraron. A continuación, en la Tabla 1 se presenta un resumen de los datos geológicos y en la Tabla 2 un resumen de los datos petrofísicos del campo de interés.

Tabla 1. *Resumen de datos geológicos del campo de interés*

Ubicación Geológica	Valle Superior del Magdalena
Ubicación Geográfica	24 Km. NW de la ciudad más cercana. 13 Km. W del campo más cercano.
Roca Generadora	Shales ricos en materia orgánica.
	Formaciones AS y AI.
Yacimiento Principal	Rocas cretáceas de Albiano-Aptiano. Ambiente deposicional deltaico.
	Areniscas cuarzosas con intercalaciones de limolita y shales grises oscuros. Formación B.
Yacimiento Secundario	Rocas cretáceas de Albiano-Aptiano. Ambiente deposicional deltaico.
	Calizas gris claro y marrón, generalmente cristalinas y dolometizadas con margas e intercalaciones de shales calcáreos.

Tabla 2. *Resumen de datos petrofísicos del campo de interés*

Porosidad	AS 14 – 18 %; AI 9 – 13 %
Permeabilidad	AS 200 – 2000 mD ; AI 50 – 200 mD
Espesor Neto	AS 65ft ; AI 46 ft
Profundidad del yacimiento	1700' - 3800'

2.2 Modelo del sector

Luego de analizar la distribución de pozos y las propiedades petrofísicas en cada Bloque del campo colombiano, se seleccionó trabajar con el Bloque 5. Este sector tiene un área superficial aproximada de 1136 acres y está conformado por un total de 36 pozos, donde 35 pozos son productores y 1 pozo es inyector. Además, cuenta con un total de 27 fallas, donde 22 fallas atraviesan el sector completamente, 1 falla es dividida en 3 partes y otra en 2 partes por que pasan por el Bloque parcialmente. La Figura 5 muestra la ubicación del Bloque 5 en el modelo de campo y la Figura 6 muestra la visualización del modelo del sector de interés.

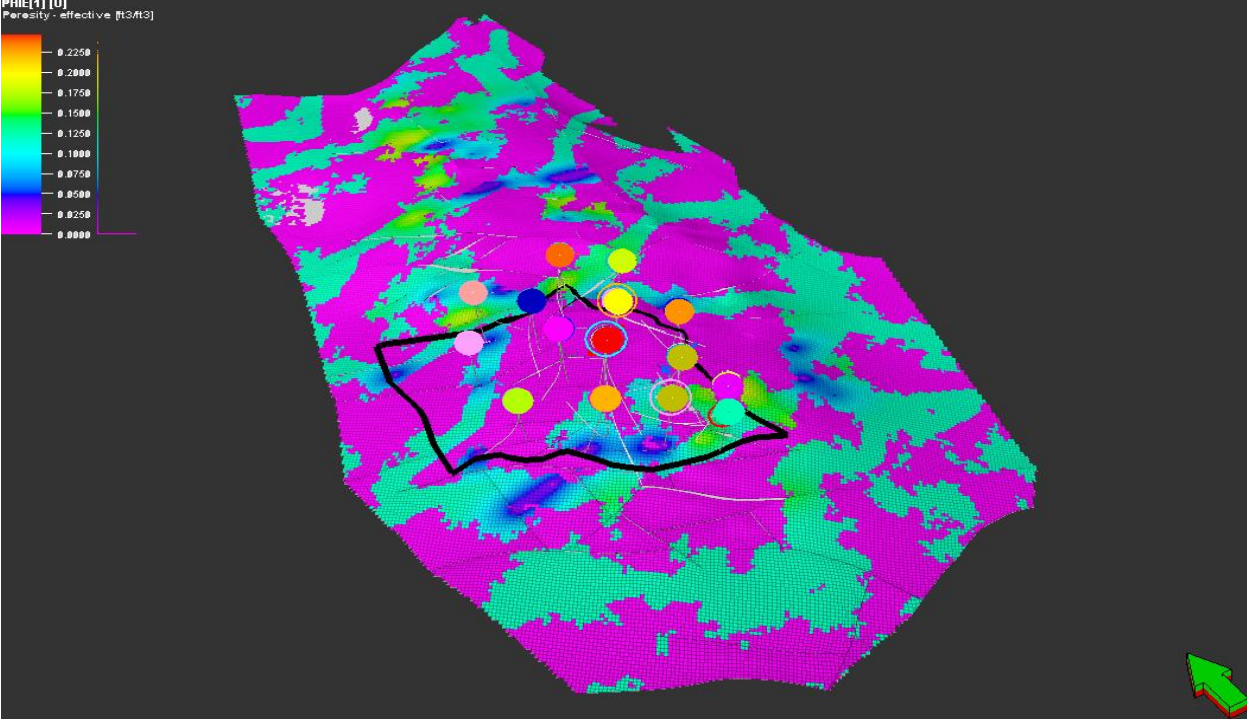


Figura 5. Ubicación del Bloque 5 en el modelo de campo

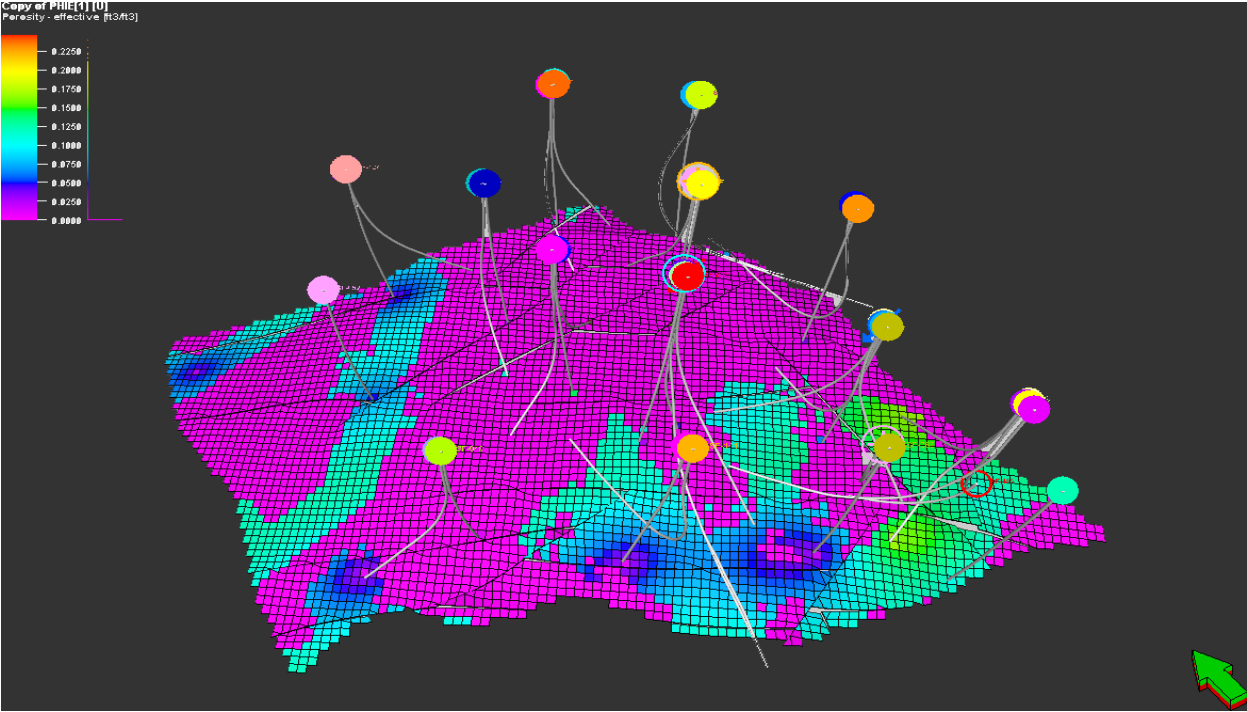


Figura 6. Modelo del sector de interés

El software de simulación seleccionado para este trabajo es IMEX-CMG versión 2015. IMEX es uno de los tres simuladores que posee CMG. Este simulador es apto para yacimientos de aceite negro a condiciones de temperatura constante o casi constante, permitiendo realizar de manera adecuada cálculos pertinentes mediante correlaciones a las propiedades de los fluidos y su respectivo comportamiento a través del tiempo para yacimientos convencionales y no convencionales. Builder-CMG versión 2015 es una herramienta software que corre bajo el Sistema Operativo Windows de Microsoft que le permite al usuario crear de manera intuitiva *datasets* o archivos de entrada para simulación. Adicionalmente, Builder posee una interfase interactiva que permite realizar procesos de importación de otros softwares de simulación. Al modelo del sector en la Figura 6 se le realizó un proceso de exportación mediante el formato RESCUE, para que este fuera compatible con el software de simulación finalmente usado. Al importar, cinco pozos presentaron problemas ya que no interceptaban realmente en la malla del sector y se decidió removerlos. La Figura 7 muestra una visualización del modelo del sector ya importado en CMG.

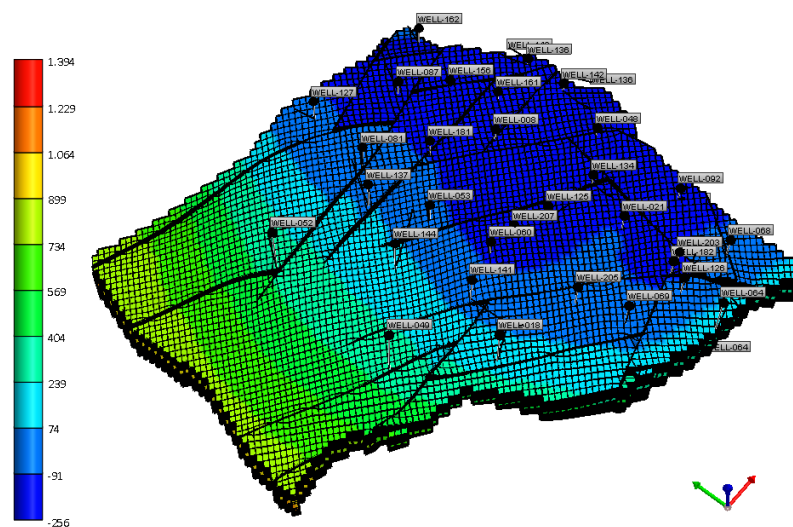


Figura 7. Modelo del sector de interés en CMG

3. Optimización de la ubicación de un nuevo pozo

3.1 Simulación del sector

Para poder generar la simulación base del modelo estático del sector de interés en *Builder* (Computer Modelling Group Ltd., 2015), que se denotara de ahora en adelante como el caso base de simulación, se construyó el modelo de fluidos y las curvas de permeabilidad relativa con base en la información que se dispone del campo. Luego, se procedió a ingresar la información pertinente a las condiciones iniciales; la profundidad del contacto agua-aceite (WOC por sus siglas en inglés), la presión y profundidad de referencia, y la información del punto de burbuja, información que también se dispuso del campo. Por último, se ingresó la fecha de inicio y la fecha final para realizar su respectiva corrida con *IMEX* (Computer Modelling Group Ltd., 2015). En la fecha final se ubicó el nuevo pozo vertical para que después de la corrida del caso base se generara el archivo de formato RESTART, necesario para su posterior optimización. El caso base de 16 años, tarda aproximadamente 13 minutos en un equipo computacional de 4 procesadores, con una malla de simulación final de 177840 celdas activas.

3.2 Optimización con CMOST

CMOST (Computer Modelling Group Ltd., 2015) es el programa de CMG que se encarga de realizar diversas tareas que involucren la simulación de yacimientos. Las tareas que puede realizar son: análisis de sensibilidad, ajuste histórico o *History Matching*, optimización y análisis de incertidumbre. Para este trabajo se seleccionó la tarea de optimización. La interfase de

optimización de CMOST le permite al usuario ingresar datos de manera intuitiva con la función objetivo del V.P.N. ya constituido dentro de los datos a ingresar por parte del usuario. Además, se requiere definir las variables financieras, *Hard* y *Soft constraints* y el algoritmo de búsqueda a usar.

3.3.1 Variables financieras. Las variables financieras por ingresar son el precio de venta del crudo, precio de venta del gas y los costos asociados a procesos de producción e inyección de agua. El software permite ingresar constantes a la ecuación de la función objetivo e inclusive exportar la ecuación para un posterior procesamiento (Computer Modelling Group Ltd., 2015). Por falta de información financiera pertinente al campo colombiano de interés en la literatura, se buscó otro campo con características semejantes que si poseía dichas variables y se modificó el precio de aceite para incorporar regalías y el *OPEX*. No se presentó una producción representativa de gas en el campo que generara un impacto en la función objetivo, además no se realizó un proceso de inyección de agua, por esta razón esas variables no se usaron. Además, se tomó la decisión de no incorporar el *CAPEX* dentro la función objetivo ya que, después de evaluarlo en el proceso de optimización, no generó un impacto sustancial al removerlo de la ecuación. La ecuación (3) representa la ecuación (1) con las variables descritas modificadas para el F.C.N. que a su vez modifican la ecuación (2). Adicionalmente, la

Tabla 3 resume las cifras de las variables financieras que se usaron para el proceso de optimización.

$$F.C.N. = [q_o * P_o] - [q_w * C_w] \dots\dots\dots (3)$$

Donde:

q_o : Producción de aceite acumulada (STB)

P_o : Precio de venta del barril de aceite (\$ USD / STB)

q_w : Producción de agua acumulada (bbl)

C_w : Costo asociado al tratamiento de agua acumulada producida (\$ USD / bbl)

Tabla 3. *Variables financieras para optimización de la función objetivo*

q_o	Producción de aceite acumulada.
q_w	Producción de agua acumulada.
P_o	\$22.80 \$ USD / STB
C_w	\$ 2.00 \$ USD / bbl
t	24 meses
i	13% o 0.13

Nota: Datos preliminares para el análisis financiero caso base. Adaptado de Muñoz, L. y Santamaria, S. (2016). EVALUACIÓN DE LOS MÉTODOS DE CONTROL DE ARENA IMPLEMENTADOS EN EL CAMPO LA CIRA INFANTAS (Tabla 21 y 23). Tesis de pregrado, Universidad Industrial de Santander, Bucaramanga.

3.3.2 Hard y Soft constraints. *Constraint* o restricción es aquel valor que viola el normal funcionamiento de un proceso. Existen dos tipos. El *soft constraint* es aquel valor que, aunque viole un normal funcionamiento, permite la sustitución del valor violado por un valor constante permitido que genera la continuación normal funcionamiento de este. El *hard constraint* es aquel valor que, al violar un normal funcionamiento, se omite y se continúa el normal funcionamiento sin procesar el valor que genera la violación (Computer Modelling Group Ltd., 2015). Para efectos

de este trabajo, no se presentó ningún caso de *soft constraint* pero si se presentaron 284 *hard constraints*, porque al realizar la optimización de la ubicación del nuevo pozo, se consideran la ubicación de pozos ya preexistentes y sus respectivas trayectorias ubicaciones prohibidas para una nueva ubicación. La ventaja que otorgan los *hard constraints* es que estos permiten ahorrar tiempos de computo porque la simulación y el cálculo de la función objetivo de la posible ubicación son omitidos.

3.3.3 Algoritmos de optimización. El software CMOST posee varios algoritmos de búsqueda para ubicar el valor óptimo y estos son: *Random Brute Force* (Fuerza Bruta Aleatoria), *Designed Exploration Controlled Evolution* (Exploración dirigida mediante Evolución Controlada), *Particle Swarm Optimization* (Optimización por Enjambre de Partículas) y *Differential Evolution* (Evolución Diferencial). Se inició la búsqueda con el algoritmo *Random Brute Force* debido a que este algoritmo encuentra el valor óptimo global del espacio solución probando todas las combinaciones posibles de manera aleatoria sin repetir combinaciones realizadas anteriormente por el mismo. De todos los algoritmos de búsqueda, este es el menos eficiente en términos de tiempo y el de mayor exigencia computacional, sin embargo, este garantiza el valor óptimo global y para poder implementarlo se dividió el sector de interés en 4 zonas y se corrió el algoritmo para cada zona. Luego de obtener los valores del óptimo global de cada zona, se seleccionó el de mayor valor que se encontró en la zona 2 en la coordenada $\{i, j\}$ (42,25), la Figura 8 muestra la ubicación del nuevo pozo vertical en la celda. El tiempo consumido en esta tarea fue de aproximadamente 1 semana continua con 2 equipos de cómputo de 4 procesadores cada uno.

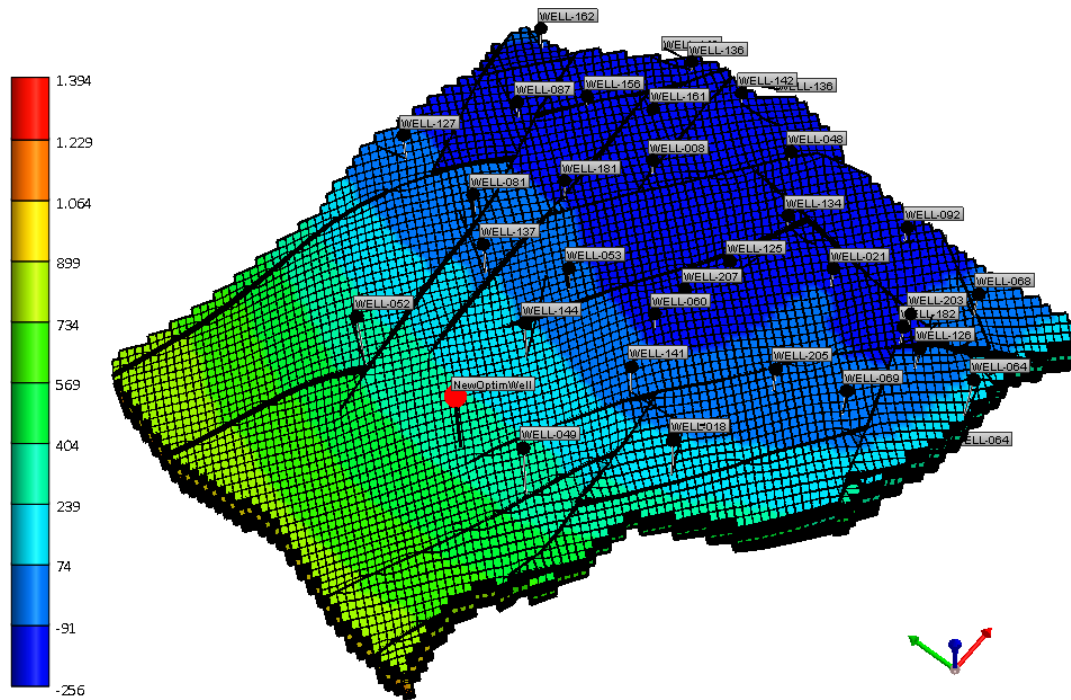


Figura 8. Ubicación óptimo global del nuevo pozo vertical

Finalmente, teniendo el valor óptimo global del sector de interés, se procedió a ejecutar los algoritmos de búsqueda que tiene disponible CMOST. Estos algoritmos aplican métodos de optimización mediante simulación y fueron seleccionados para evaluar su rendimiento en tiempo para encontrar el valor óptimo global de la nueva ubicación de un pozo infill. Los algoritmos son Designed Exploration Controlled Evolution o D.E.C.E., Particle Swarm Optimization o P.S.O., y Differential Evolution o D.E. Se encontró que cada algoritmo de búsqueda seleccionado halló el valor óptimo global en aproximadamente 3 días con un equipo computacional de 4 procesadores. La Figura 9, Figura 10, Figura 11 y Figura 12 visualizan los resultados del óptimo global del algoritmo Random Brute Force para cada zona. La Figura 13 y Figura 14 detallan las tasas de aceite y agua mensual respectivamente para el caso del valor óptimo global, la Figura 15 y Figura

16 detallan las gráficas de aceite y agua acumulada respectivamente. Las Figura 17, Figura 18 y Figura 19 visualizan los resultados del óptimo global para cada algoritmo aplicado.

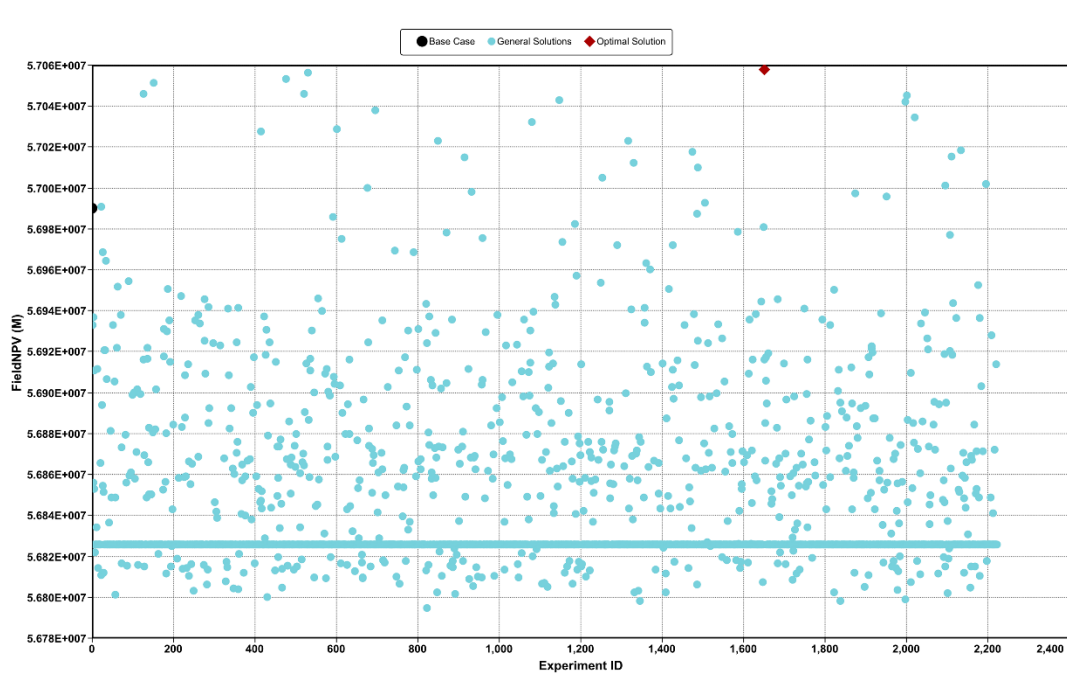


Figura 9. Resultados del algoritmo Random Brute Force para la zona 1

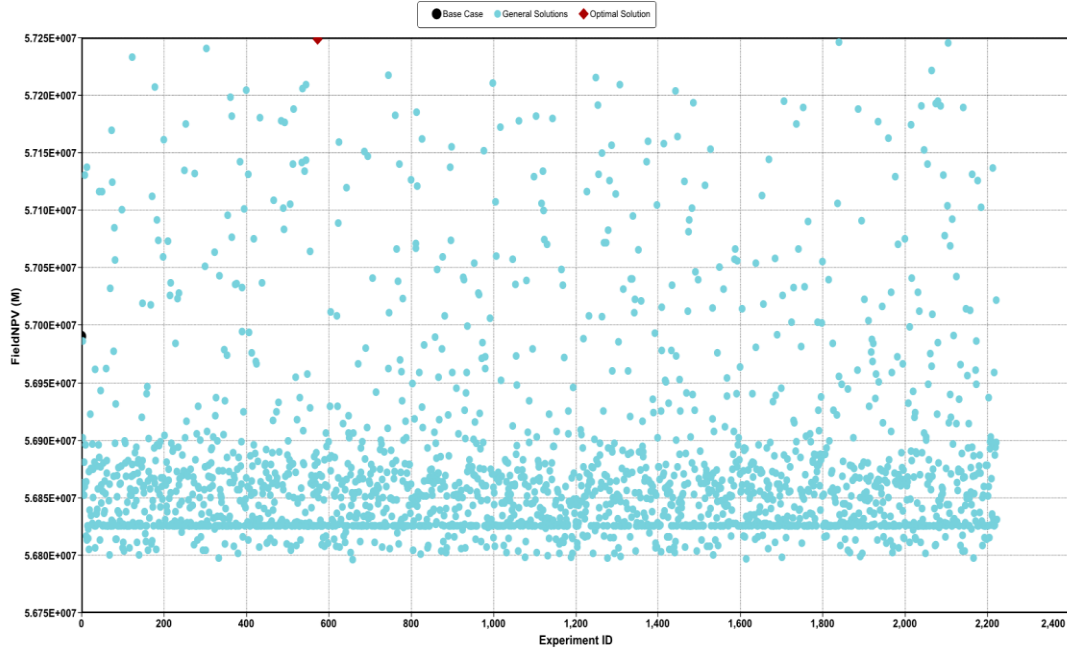


Figura 10. Resultados del algoritmo Random Brute Force para la zona 2

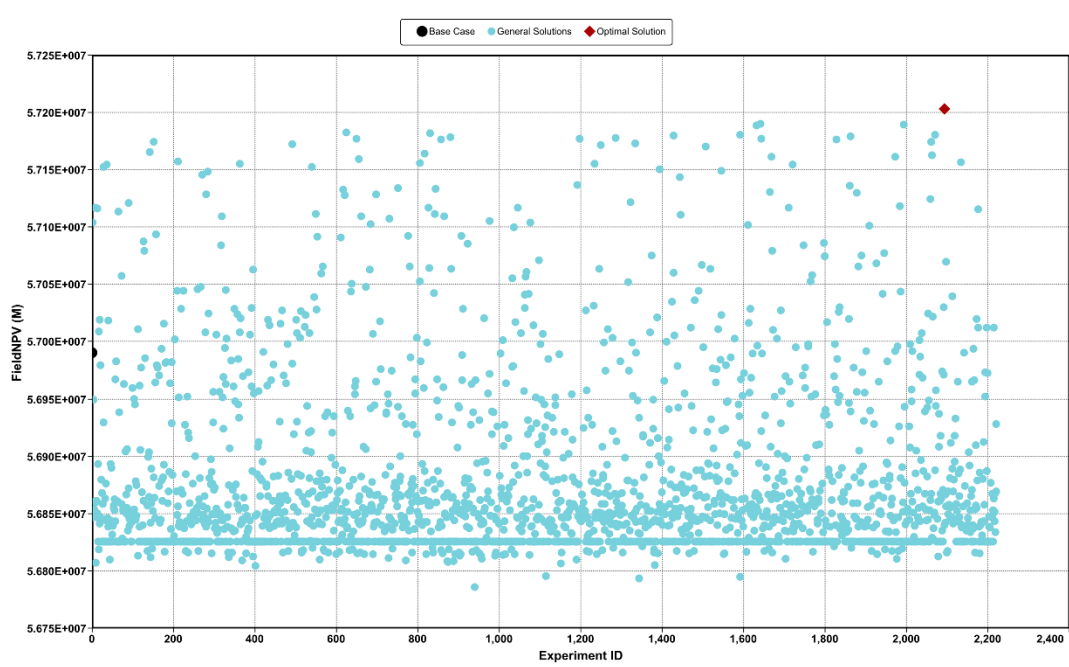


Figura 11. Resultados del algoritmo Random Brute Force para la zona 3

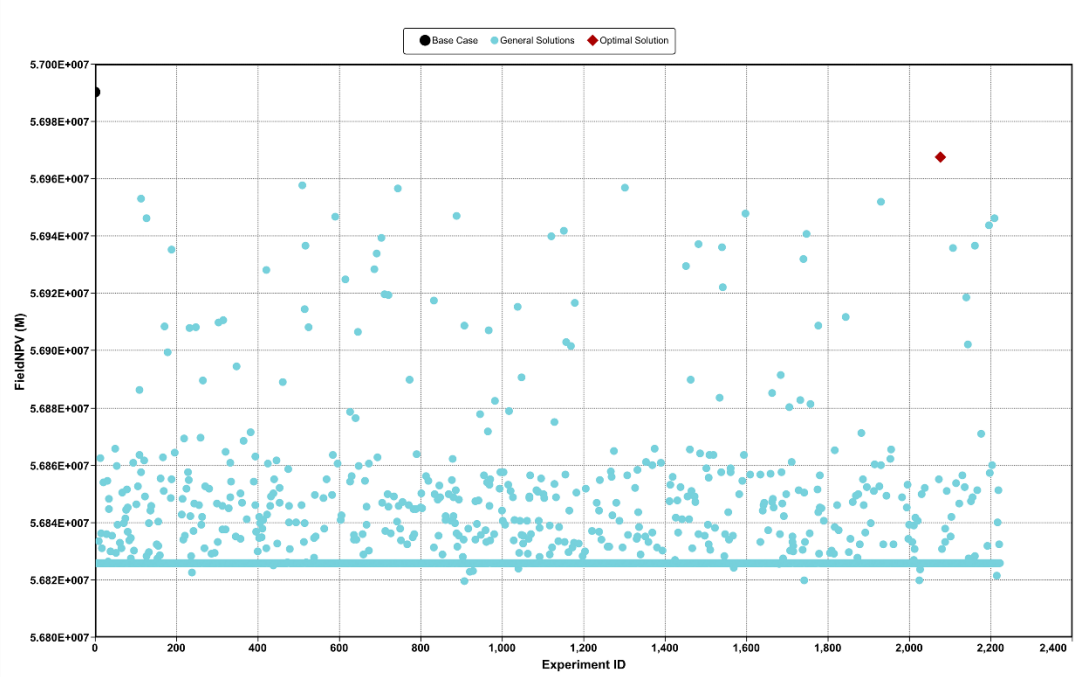


Figura 12. Resultados del algoritmo Random Brute Force para la zona 4

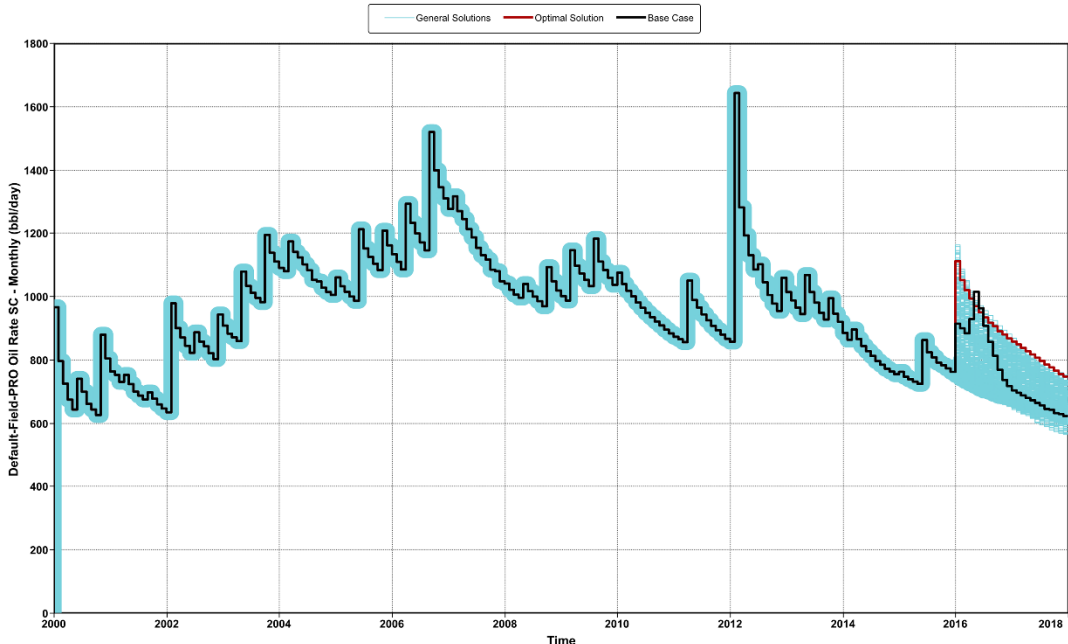


Figura 13. Tasa de aceite mensual a condiciones estándar caso valor óptimo global del sector

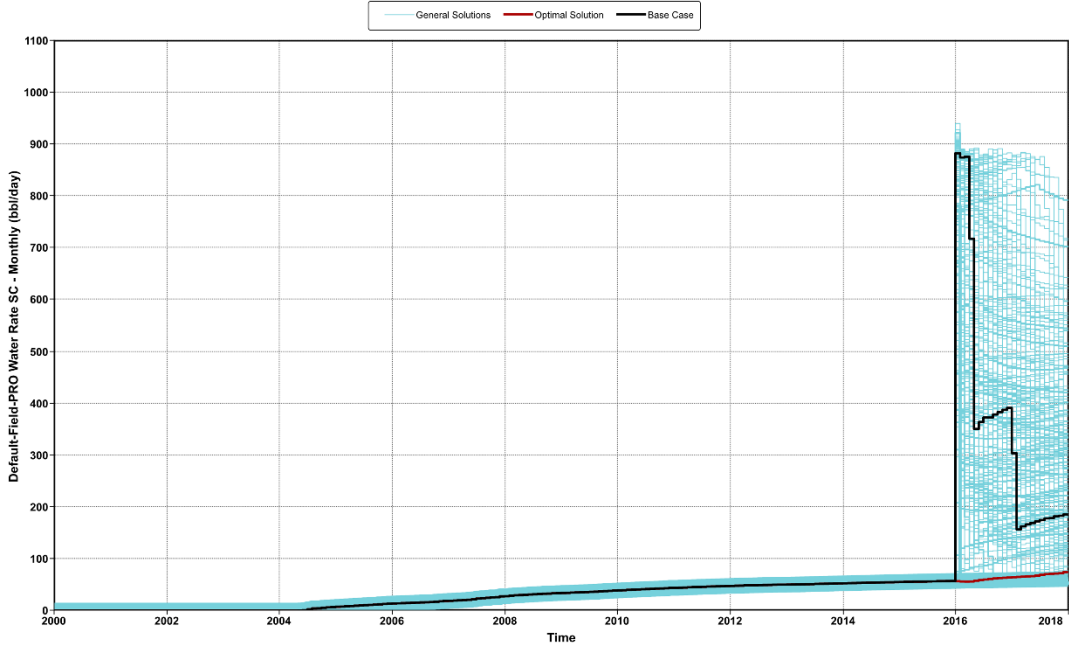


Figura 14. Tasa de agua mensual a condiciones estándar caso valor óptimo global del sector

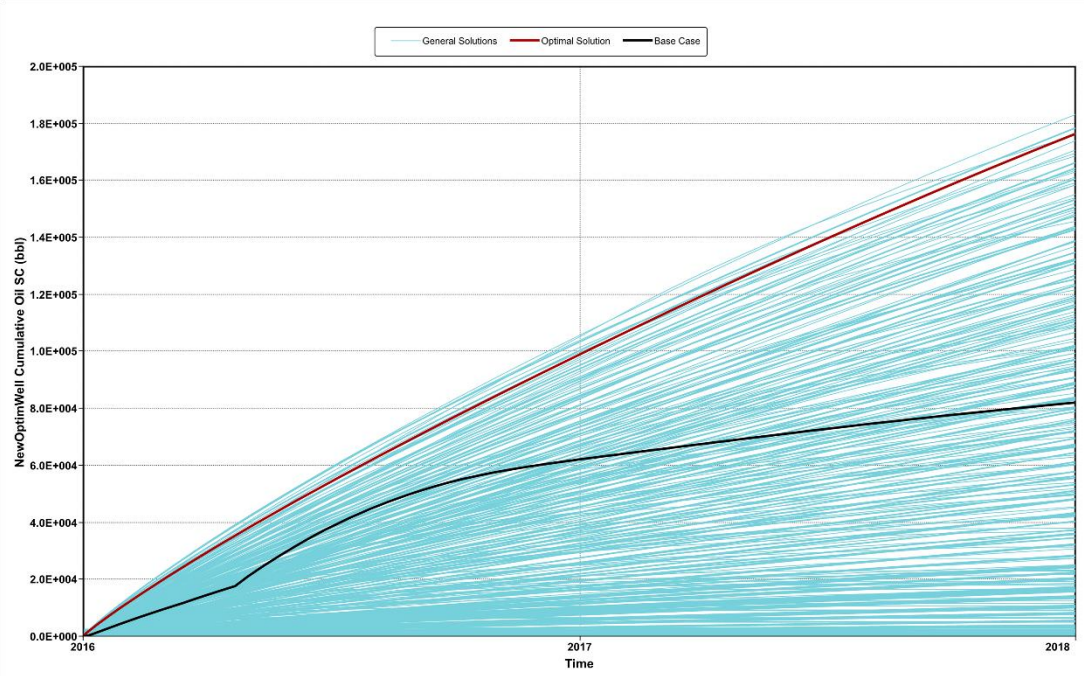


Figura 15. Aceite acumulado a condiciones estándar caso valor óptimo global del sector

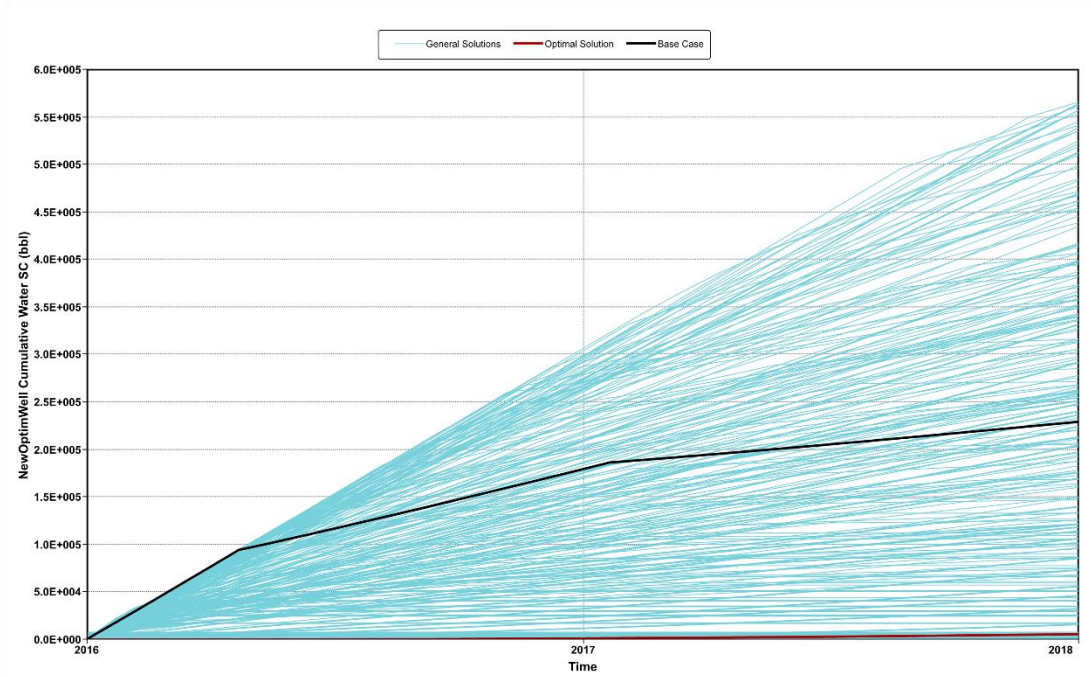


Figura 16. Agua acumulada a condiciones estándar caso valor óptimo global del sector

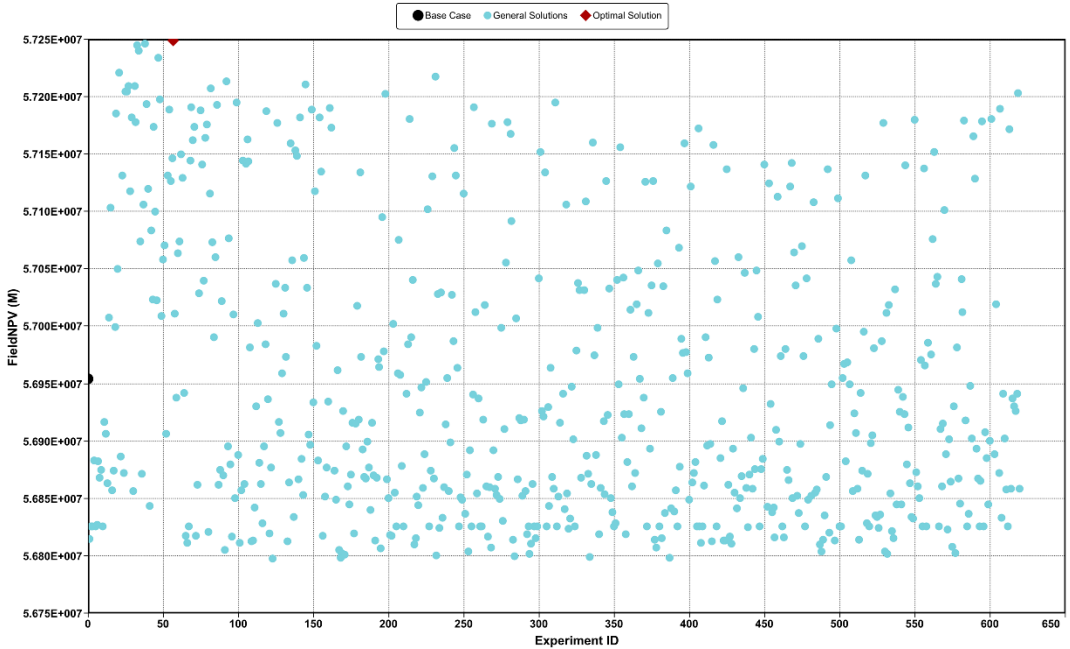


Figura 17. Resultados del algoritmo *Designed Exploration Controlled Evolution*

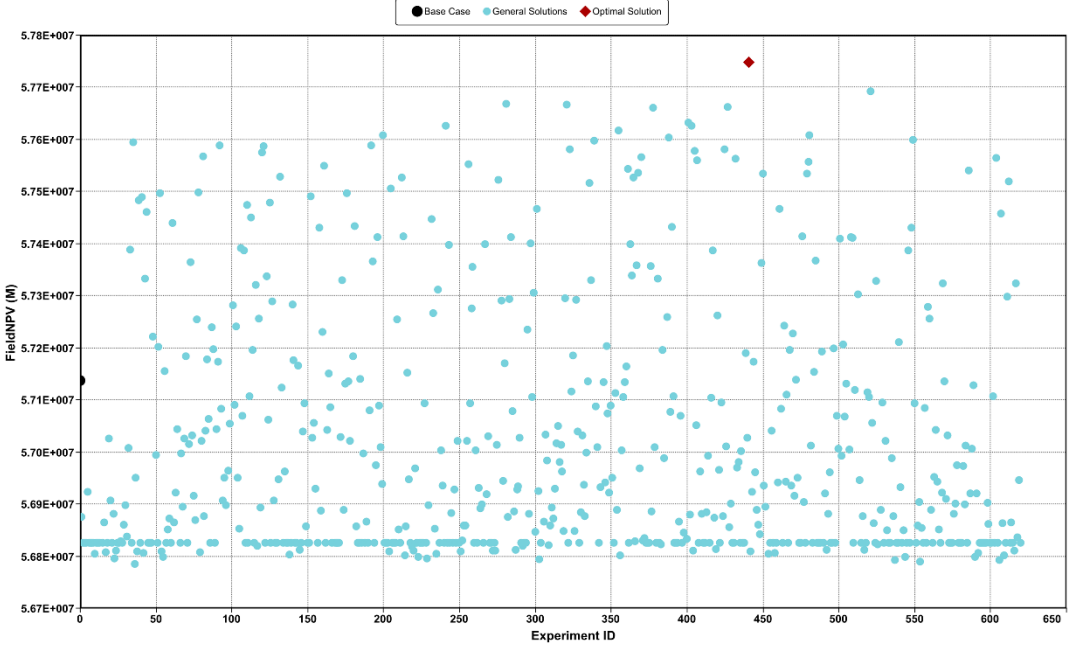


Figura 18. Resultados del algoritmo *Particle Swarm Optimization*

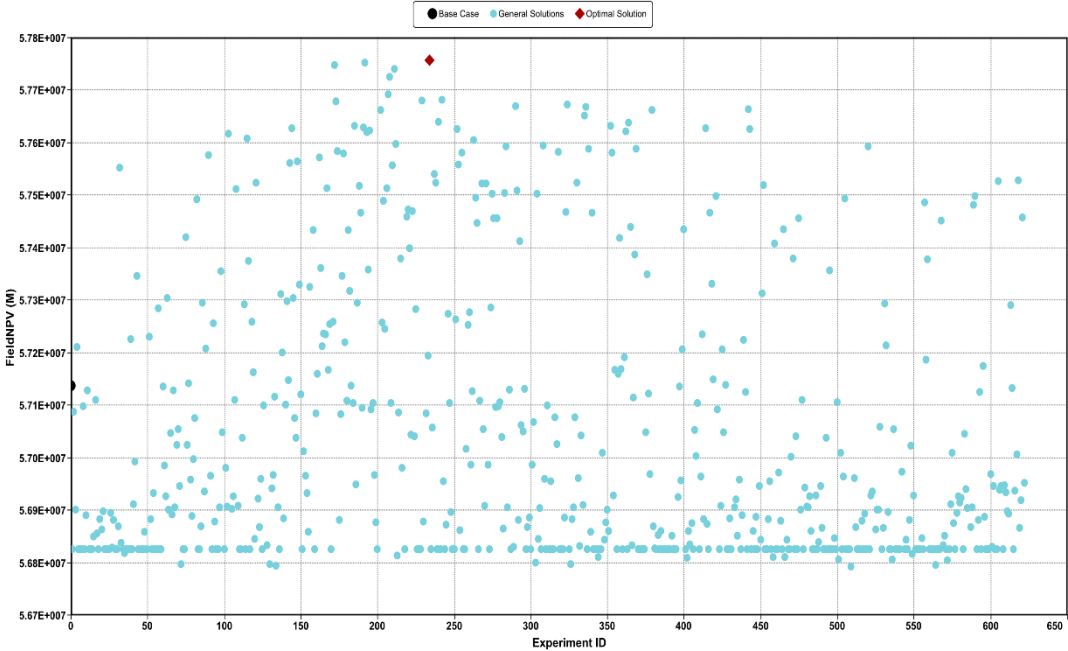


Figura 19. Resultados del algoritmo *Differential Evolution*

4. Desarrollo del modelo proxy

4.1 Modelamiento proxy

Para el desarrollo de un modelo Proxy, primero se debe seleccionar el tipo de modelo que se va a emplear y el que mejor se adecúa a los datos de los resultados generados por la simulación. Se decidió trabajar los modelos proxy de tipo regresión polinomial y red neuronal artificial, el primero por su simplicidad en su construcción y ejecución y el segundo por sus buenos resultados al tratar con información no lineal. Luego se procedió a sus respectivas construcciones a través de un set de pasos a seguir para sus respectivos modelamientos. La Figura 20 adaptado de (Zubarev, 2009) resume el procedimiento que se empleó para la construcción de los modelos proxy.

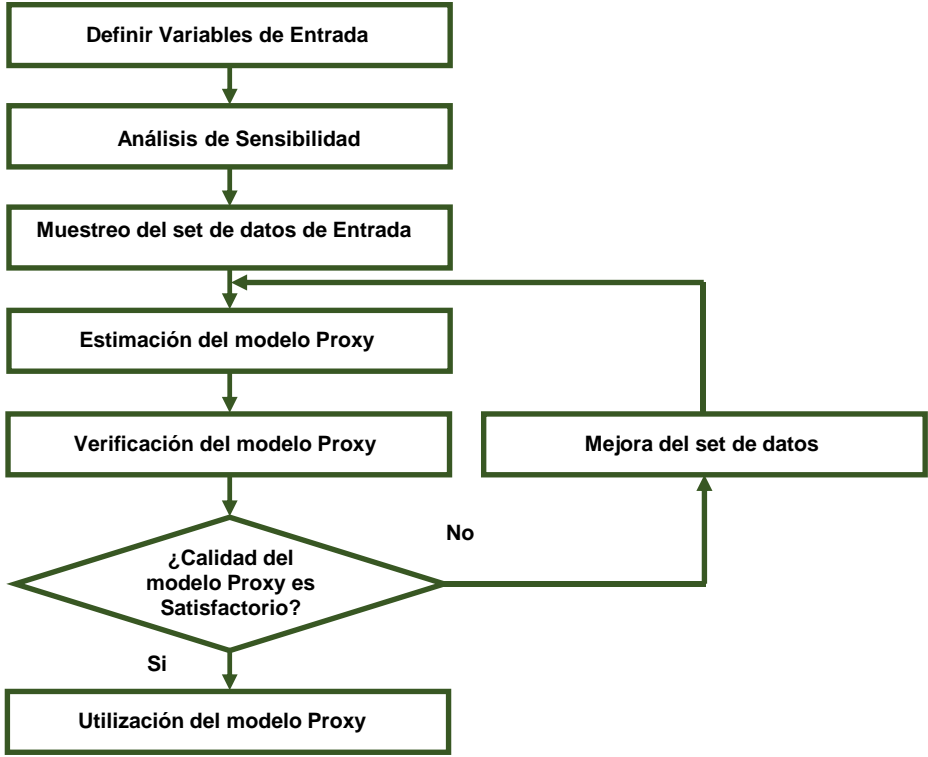


Figura 20. Diagrama de flujo para el modelamiento proxy

4.1.1 Variables de entrada. Las variables de entrada que se usaron para la construcción de los modelos Proxy fueron las coordenadas $\{i, j\}$ de las celdas con una ubicación disponible para un nuevo pozo, y el resultado obtenido de la función objetivo para dichas coordenadas. La función objetivo representa las variables de producción, información representativa de cada celda cuando se ubica un pozo en esa ubicación. Esta data fue obtenida por *CMOST* junto con *IMEX*, con el algoritmo de búsqueda *Random Brute Force*.

4.1.2 Análisis de Sensibilidad. El análisis de sensibilidad es la cuantificación del impacto de los parámetros de entrada en el modelo respuesta de la simulación. Esto es logrado a través del análisis de los cambios en el modelo de salida de simulación para diferentes combinaciones de las variables de entrada. Basado en los resultados del análisis de sensibilidad se puede eliminar las variables que no generan un impacto suficiente en los resultados del modelo de simulación (Zubarev, 2009). Para generar los modelos Proxy, este paso no es necesario porque las variables determinantes para la ubicación de un nuevo pozo son coordenadas $\{i, j\}$, producción de agua acumulada y producción de aceite acumulada; estas dos últimas son agrupadas y representadas por el Valor Presente Neto.

4.1.3 Muestreo del set de datos de entrada. El muestreo de la data de entrada es la etapa más importante en crear y adecuar los modelos proxy. Existen diferentes métodos de muestreo que se pueden aplicar. Para este trabajo se tomó la decisión de realizar muestreos bi-etapa, un muestreo que se caracteriza por combinar las características de un muestreo por conglomerados y un muestreo estratificado. En el muestreo por conglomerados, en lugar de seleccionar todos los sujetos de la población inmediata, el investigador realiza una selección de grupos o conglomerados y algunos conglomerados. En el muestreo estratificado, se aplica muestreo probabilístico en donde

el investigador divide a toda la población en diferentes subgrupos o estratos, luego selecciona aleatoriamente a los sujetos finales de los diferentes estratos en forma proporcional. Para poder determinar el tamaño de la muestra “n”, se usa la ecuación (4) que según la literatura permite calcular el tamaño de muestra ideal (Kuehl, 2001). La Tabla 4 resume los valores de las variables estadísticas que garantizan la media de la población y generan el tamaño de muestra óptimo.

$$n = \frac{N * p * q * Z_{\alpha}^2}{[E^2(N-1)] + [p * q * Z_{\alpha}^2]} \dots\dots\dots (4)$$

Donde:

n: Tamaño de la muestra

N: Tamaño de la población

p: Probabilidad de éxito

q: Probabilidad de fracaso

Z_α: Nivel de confianza

E: Error muestral admisible

Tabla 4. *Variables estadísticas para el muestreo del sector de interés*

<i>N</i>	8892 datos.
<i>p</i>	50 % = 0.5
<i>q</i>	50 % = 0.5
<i>Z_α</i>	99 % = 2.58
<i>E</i>	3 % = 0.03
<i>n</i>	1531 datos.

Nota: Control local de errores experimentales. Adaptado de Kuehl, R. (2001). *Diseño de Experimentos, Segunda edición (Capítulo 1.5)*. Editorial THOMSON LEARNING. México.

4.1.4 Estimación del modelo proxy. El modelo proxy es una aproximación de un modelo numérico de yacimiento existente. Consecuentemente este debe ser capaz de replicar la respuesta altamente no lineal del modelo real, ser fácil de construir y ser simple en su aplicación. Cada modelo es estimado individualmente. Para la estimación de los modelos Proxy, se procedió con la construcción de un polinomio y el entrenamiento de una red neuronal de base radial, un tipo de red neuronal artificial que calcula la salida de la función en función de la distancia a un punto denominado centro (Temizel , et al., 2014).

4.1.5 Comparación de tiempos de simulación. La comparación de los tiempos de simulación empleados tanto para la optimización por *Random Brute Force*, los algoritmos de búsqueda y los modelos proxy son presentados en la Tabla 5. Los resultados muestran que los tiempos empleados para la construcción de los modelos proxy son menores y también logran obtener el valor óptimo global como los demás procesos, exceptuando el modelo de regresión polinomial.

Tabla 5. *Tiempos de simulación empleados por cada método de optimización*

Método de Optimización	Tiempo empleado (horas)
<i>Ramdon Brute Force</i>	336
<i>Designed Exploration Controlled Evolution</i>	72
<i>Particle Swarm Optimization</i>	72
<i>Diferential Evolution</i>	72
<i>Modelo Proxy Regresión Polinomial</i>	62
<i>Modelo Proxy Red Neuronal Artificial</i>	62

4.1.6 Comparación de resultados de modelos. Para poder comparar los resultados obtenidos de la regresión polinomial y la red neuronal con los del modelo base, se procedió a graficar los

resultados de la función objetivo del caso base contra los generados por cada modelo proxy. En la Figura 21 puede observarse que, en el modelo de regresión polinomial, los datos están aglomerados en un intervalo definido y en ningún momento es posible visualizar la generación de una línea de pendiente igual a 1, lo cual es un indicativo de la poca funcionalidad del modelo. En el modelo de red neuronal artificial, se observó que la mayor parte de los datos generan una línea de pendiente de 45°, por lo cual, es presumible un comportamiento adecuado del modelo proxy, esta información se puede observar en la Figura 22. Adicionalmente, se graficaron los resultados obtenidos del modelo base y el modelo de regresión polinomial contra el número total de coordenadas para evaluar si ambos modelos presentan tendencias similares, evidenciándose una completa diferencia en el comportamiento de los modelos, esta información se puede observar en la Figura 23. Así mismo, se graficaron los resultados obtenidos del caso base y la red neuronal contra el número total de coordenadas, constatándose que ambos modelos presentan tendencias similares, esta información se puede visualizar en la Figura 24.

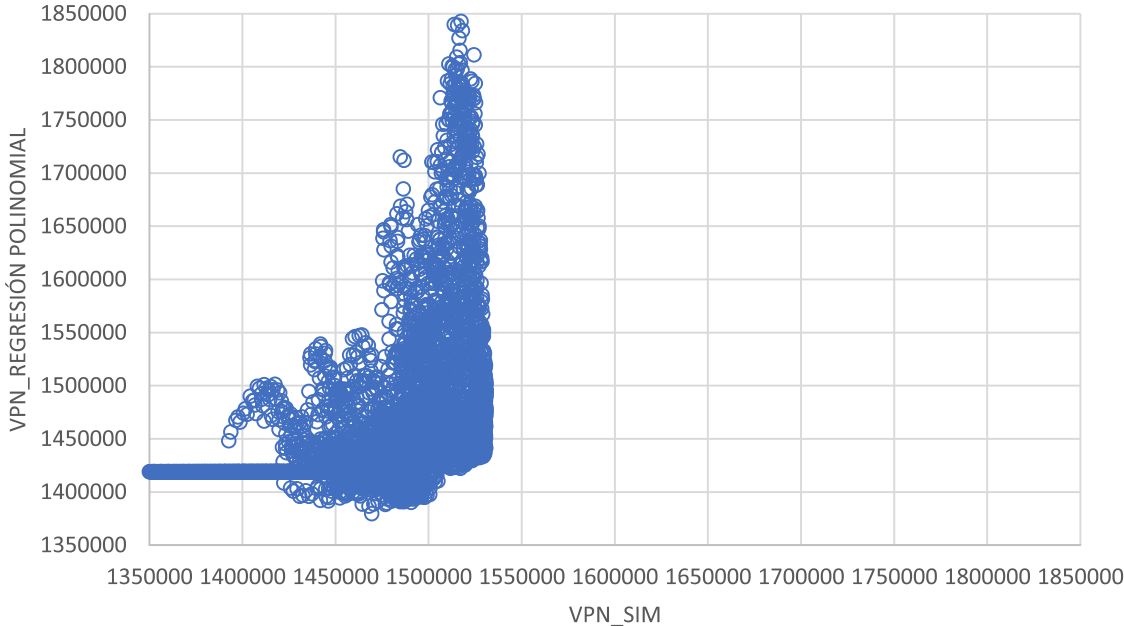


Figura 21. Gráfica de los Resultados del modelo base vs. Resultados de R. P.

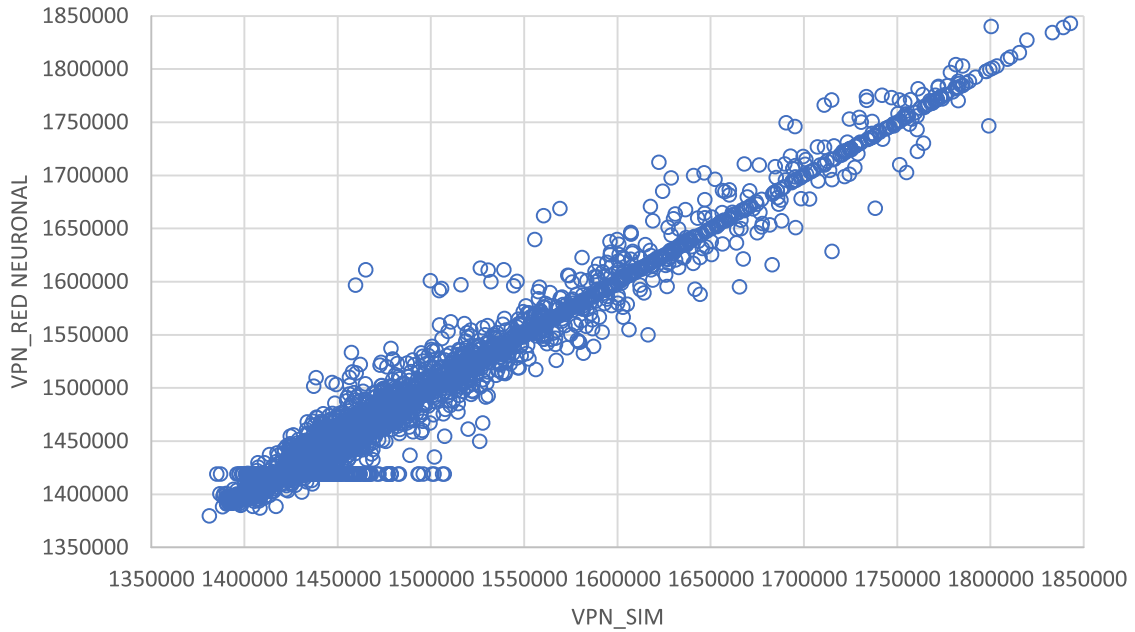


Figura 22. Gráfica de los Resultados del modelo base vs. Resultados de la R.N.A.

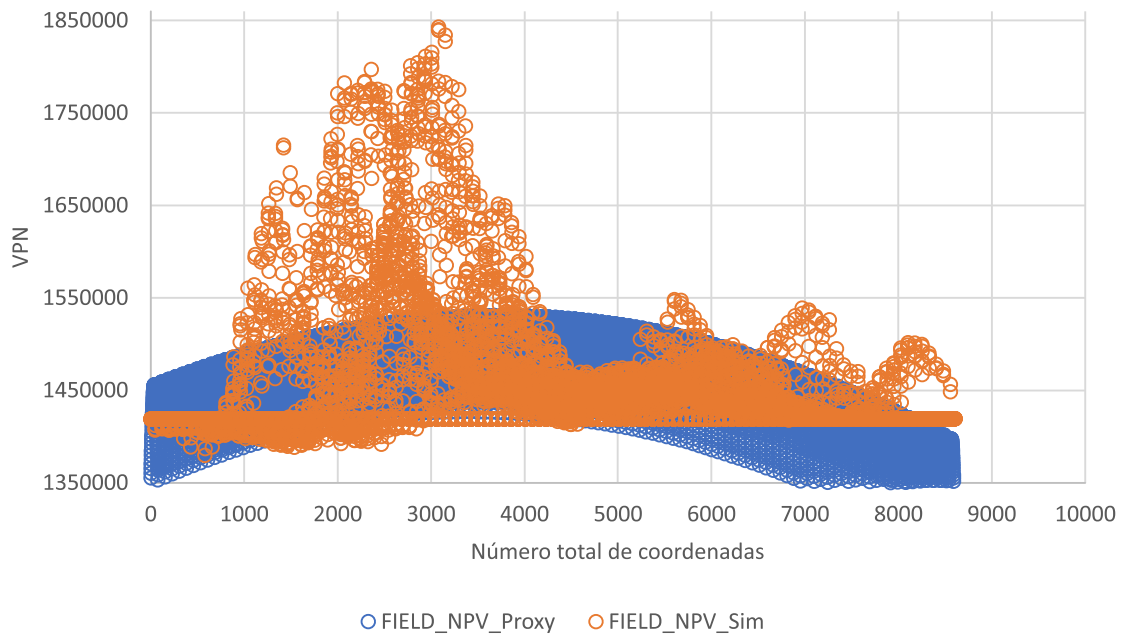


Figura 23. Gráfica del V.P.N. vs. Número total de coordenadas para el modelo de R. P.

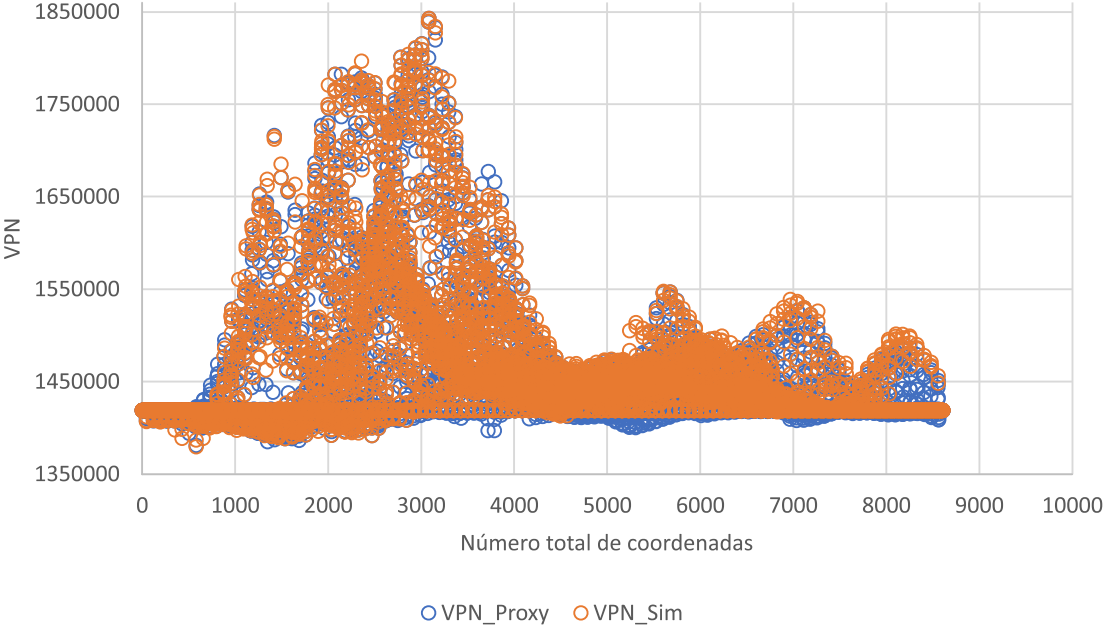


Figura 24. Gráfica del V.P.N. vs. Número total de coordenadas para el modelo de R.N.A.

5. Validación del modelo Proxy y análisis complementario

Un paso crítico del modelamiento Proxy es la validación de los modelos generados, para esto, se recurre a la Estadística Aplicada, específicamente al cálculo de errores porcentuales y la aplicación de pruebas de hipótesis como métodos de validación de datos. La prueba de hipótesis seleccionada fue la prueba de Welch, con la cual, se busca determinar si existe diferencia entre las medias de los modelos proxy con respecto a la media del modelo base.

Por otro lado, en aras de profundizar en el análisis de los modelos, se recurrió a las curvas ROC para realizar un proceso de comparación entre resultados. Estas curvas son aplicadas en la bioestadística, rama de la estadística aplicada a la medicina, para, por ejemplo, en los estudios de laboratorio, validar nuevos protocolos tales como un nuevo tratamiento, un nuevo medicamento, etc.; adicionalmente, se realizó un análisis complementario a la data espacial del modelo base apoyado en la Geoestadística, rama de la estadística aplicada al manejo de data espacial, para observar si existe alguna autocorrelación global de los datos.

5.1 Diferencia de Medias (Prueba de Welch)

Una prueba de Welch es una “prueba t de 2 muestras” aproximada. La prueba t de 2 muestras se utiliza para comparar si las medias de dos grupos independientes o para este trabajo, la data del modelo base con respecto a cada modelo proxy, son diferentes. Si las medias de ambos grupos son iguales, esto significa que ambos grupos independientes son iguales estadísticamente, de lo contrario cada grupo independiente es distinto. Esta prueba se deriva bajo el supuesto de que los

grupos independientes están normalmente distribuidos y poseen varianzas iguales. Sin embargo, las pruebas que involucran las varianzas están sujetas a limitaciones importantes porque son sensibles a desviaciones con respecto a la normalidad. Por esta razón, B.L. Welch desarrolló un método de aproximación para comparar las medias de dos poblaciones normales independientes cuando las varianzas no son necesariamente iguales (Minitab Inc., 2017), permitiendo de esta manera la comparación de medias sin depender de una igualdad entre las varianzas de los grupos independientes. Adicionalmente, la prueba de Welch trabaja con la incertidumbre de una distribución normal de la data de los grupos independientes porque aplicando el Teorema del Límite Central (Kuehl, 2001), solo se requiere que las poblaciones y muestras de los grupos independientes sean suficientemente grandes; los modelos base, regresión polinomial y red neuronal artificial cumplen con el Teorema.

Toda prueba estadística parte de los planteamientos de la hipótesis nula y la hipótesis alternativa. En este punto, es fundamental definir el elemento clave de la salida en el que hay que concentrarse cuando se realiza una prueba de hipótesis, el valor p , este, es el nivel de significancia más bajo en el que el valor observado de la estadística de prueba es significativo. Un valor p oscila entre 0 y 1 y mide la evidencia en contra de la hipótesis nula. El valor p se compara con el nivel de significancia (α) para decidir si debe rechazar la hipótesis nula (H_0), de esta manera, si el valor p es menor que o igual a α , se rechaza H_0 , por el contrario, si el valor p es mayor que el nivel de significancia (α), no puede rechazarse H_0 (Minitab, Inc., 2017).

La hipótesis nula para la prueba de Welch estipula que las medias de cada grupo independiente son iguales o en otras palabras que no existe diferencia alguna entre estas, por el contrario, la hipótesis alternativa establece que sí existe una diferencia entre las medias de las dos poblaciones. La Tabla 6 resume los valores resultantes de P con un nivel de significancia del 1% al aplicar la

prueba de Welch a cada uno de los modelos proxy, en esta prueba se compararon, para cada caso, la media de los resultados obtenidos por regresión polinomial y red neuronal artificial contra la media de los resultados arrojados por el modelo de simulación.

Para el modelo de regresión polinomial, se obtuvo que $P < \alpha$, por tanto, se rechaza la hipótesis nula y es posible establecer que sí existe una diferencia entre las medias de las dos poblaciones y cada grupo de datos es distinto. Así mismo, para el modelo de red neuronal artificial, se obtuvo que $P > \alpha$, por tanto, no se rechaza la hipótesis nula y se establece que ambos grupos de datos son iguales estadísticamente, de esta manera se valida la red neuronal artificial como modelo predictivo para la ubicación del nuevo pozo infill y se rechaza la regresión polinomial como modelo válido de predicción.

Tabla 6. Resultados del valor P obtenido en cada prueba para los modelos proxy

Prueba de Welch			
Regresión Polinomial		Red Neuronal Artificial	
Nivel de significancia (α)	0.01	Nivel de significancia (α)	0.01
Valor P	3.62E-15	Valor P	0.7468

5.2 Error Porcentual

El error es una medida de suma importancia para cualquier análisis de datos. El error hace referencia a que tanto una función, una formula e inclusive la misma estadística falla en explicar o modelar el verdadero valor teórico de la data (Minitab Inc., 2017). Identificar y cuantificar el error puede ayudar explicar la presencia de este mismo e inclusive permite medir la incertidumbre en los análisis estadísticos. Para cuantificarlo, se establece una relación que a su vez se puede representar como un porcentaje cuando este es multiplicado por 100. La relación del error

establece una división entre la diferencia del valor absoluto de la data teórica y experimental con respecto a la data teórica. La ecuación (5) denota la ecuación del error porcentual.

$$\%E = \frac{|V.T.-V.E.}|}{V.T.} * 100 \dots\dots\dots (5)$$

Donde:

V.T. : Valor teórico de la data

V.E. : Valor experimental de la data

Para evaluar el error porcentual de los modelos de regresión polinomial y red neuronal artificial, se escogió el valor óptimo global de la función objetivo del caso base como el valor teórico. Para los valores experimentales se tomaron los resultados obtenidos de las funciones objetivos pronosticados para cada modelo proxy. La *Tabla 7* muestra los resultados de la función objetivo obtenidos en la ubicación óptima del pozo infill para los modelos proxy y el modelo base y los respectivos valores del error porcentual. Es notable como la regresión polinomial falla estrepitosamente en el pronóstico del valor real del VPN con un error del 17.7% en su cálculo, mientras la red neuronal artificial, gracias a su proceso de aprendizaje, logra un adecuado pronóstico, con solo un 0.002% de error en esa ubicación, reafirmando su funcionalidad.

Tabla 7. Errores porcentuales de los valores arrojados por cada modelo proxy en comparación con el resultado óptimo de la función objetivo del modelo base

VPN Óptimo Caso Base		1842882.8 USD	
VPN R. Polinomial	1517451.8 USD	Error Porcentual	17.7%
VPN RNA	1842923.7 USD	Error Porcentual	0.002%

5.3 Aplicación de la curva ROC

ROC representa la sigla “Receiver Operating Characteristic” o en español Característica de la Operatividad de un Receptor. El nombre proviene de un sistema usado por los británicos durante la Segunda Guerra Mundial para distinguir las verdaderas señales de radio del ruido de fondo. Usaron estos gráficos para evaluar su capacidad de leer precisamente las señales que los alertaban de una invasión aérea alemana. Si acertaban, era un verdadero positivo; de ser incorrecto, era un falso positivo y fue aplicado por primera vez para la Bioestadística en la psicofísica, rama de la psicología que trata las relaciones entre el estímulo físico y las sensaciones y estados mentales resultantes de un receptor, así mismo, las curvas ROC también se han utilizado para evaluar la precisión de los motores de búsqueda en Internet.

Esta curva es una representación gráfica que ilustra la habilidad diagnóstica de un modelo clasificador binario (Aghajan, Delgado, & Augusto, 2010). En la evaluación de pruebas diagnósticas, la {sensibilidad} se coloca en la ordenada y $\{1 - \text{especificidad}\}$ en la abscisa. La sensibilidad más alta es del 100%, representada por un 1, y la especificidad más alta de 1 resulta en una $\{1 - \text{especificidad}\} = 0$. Las pruebas más precisas tendrán un punto que se acerque a la parte superior izquierda del gráfico donde la $\{sensibilidad\} = 1$ y $\{1 - \text{especificidad}\} = 0$. Las pruebas menos precisas no funcionarán mejor que los resultados al azar, representados por una línea recta con una pendiente de 45 grados (Dawson, 2009).

La Figura 25 muestra algunas curvas ROC. La curva ROC A corresponde a una prueba con sensibilidad y especificidad próximas al 100%. Ésta es una situación ideal poco habitual y que se acerca a la exactitud del estándar de referencia. La prueba B, con sensibilidad y especificidad más altas en el punto donde la curva está más próxima a la esquina superior izquierda. La prueba C no tiene mejores resultados que el propio azar.

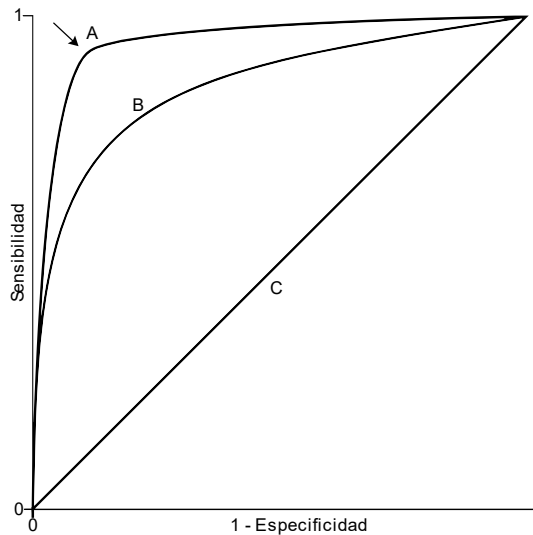


Figura 25. Ejemplos de curvas ROC

La curva ROC se usa en este trabajo para comparar la efectividad de los modelos Proxy generados, el de regresión polinomial y la red neuronal artificial. La Figura 26 muestra la gráfica de las curvas ROC obtenidas para ambos modelos proxy, con una condición binaria fijada como “resultados obtenidos por regresión polinomial y red neuronal con un porcentaje de error menor al 3% con respecto a los obtenidos por simulación, en las mismas coordenadas i y j ”; en esta gráfica puede observarse que la curva ROC para la red neuronal artificial se ubica considerablemente por encima de la curva ROC para la regresión polinomial, lo cual es una indicación del rendimiento superior de este modelo y una confirmación de la validez de la red neuronal como un modelo funcional comparado con la regresión polinomial.

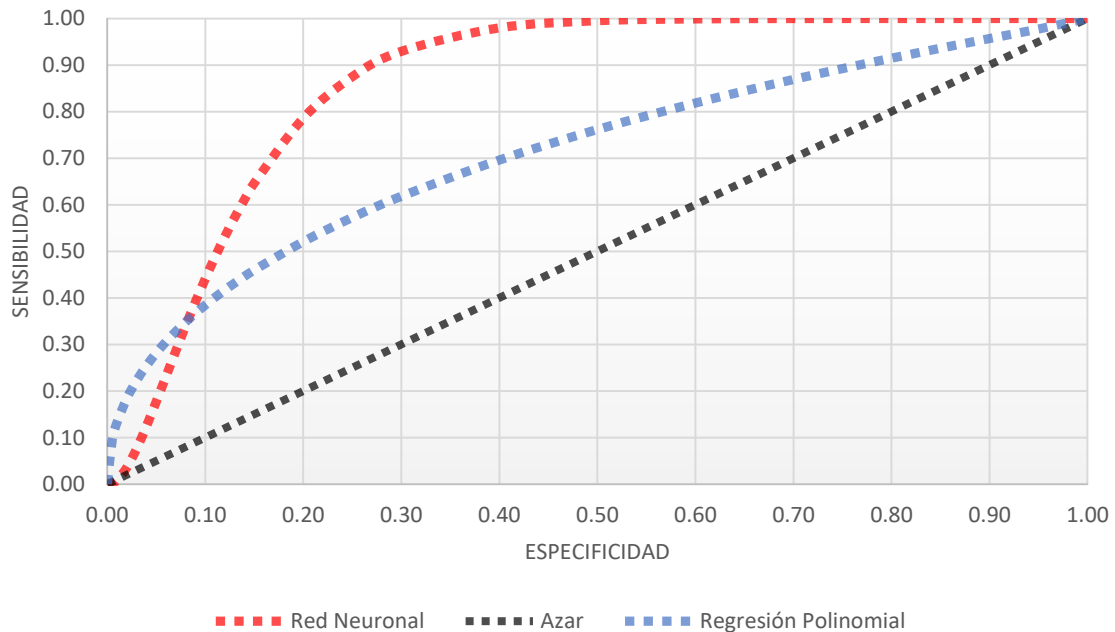


Figura 26. Grafica de los Resultados de la curva ROC

5.4 Análisis complementario de la data espacial

La data espacial en el desarrollo de este trabajo fue fundamental debido a que la ubicación de un nuevo pozo está en función de esta misma. Esto se debe a que cada posible ubicación, posee características diferentes y consecuentemente generan diferentes resultados en términos de producción. Adicionalmente, la ubicación de un nuevo pozo en cada ubicación no únicamente afecta el desempeño de donde se encuentra ubicado, sino también, afecta las ubicaciones próximas que pueden o no estar interconectadas según la geología y propiedades de las rocas presentes en la zona. Para el desarrollo de este trabajo, el manejo de los datos espaciales se limitó a las coordenadas i y j , valores ya predeterminados del modelo estático del campo suministrado por la compañía operadora del campo colombiano. Esta información es representada en puntos con un valor en i desde 1 hasta 117 y un valor en j desde 1 hasta 76, conformando 8892 datos espaciales en total para el sector de interés.

Para poder determinar si existe algún tipo de relación de la data espacial para cada modelo, se recurrió a la medida de asociación espacial I de Morán. El coeficiente I de Morán es una medida global de autocorrelación que calcula la razón entre el producto de la variable de interés y su rezago espacial con el producto de la variable de interés ajustada a su peso espacial. Este coeficiente toma valores entre -1 que indica perfecta dispersión y +1 que indica perfecta autocorrelación. Un valor 0 indica que el patrón espacial es aleatorio (Zhuvok, 2017). La ecuación (6) define el coeficiente global con su respectiva ecuación.

$$I = \frac{v}{\sum_{i=1}^v \sum_{j=1}^v w_{ij}} \frac{\sum_{i=1}^v \sum_{j=1}^v w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^v (y_i - \bar{y})^2} \dots\dots\dots (6)$$

Donde:

v : Variable de interés

y_i : Valor de la variable para la i -ésima observación

\bar{y} : Media de la muestra

w_{ij} : Peso espacial de la conexión entre i y j

La I de Moran es una estadística deductiva, lo que significa que los resultados del análisis siempre se interpretan dentro del contexto de la hipótesis nula. Para la estadística I de Moran global, la hipótesis nula establece que el atributo que se analiza está distribuido en forma aleatoria entre las entidades del área de estudio, es decir, los procesos espaciales que promueven el patrón de valores observado constituyen una opción aleatoria. Cuando el valor P que se obtiene es estadísticamente significativo, se puede rechazar la hipótesis nula, pero cuando el valor P no es estadísticamente significativo, no se puede rechazar la hipótesis nula y es posible que la

distribución espacial de los valores de entidades sea el resultado de procesos espaciales aleatorio (Esri, 2017).

El cálculo del coeficiente de la I de Morán global para el modelo base se realizó con un nivel de significancia del 1%, $\alpha=0.01$, y el resultado obtenido fue de $I=0.136222$ y un valor $P=0$. Dado que el valor P es estadísticamente significativo, es decir, $P < \alpha$, se puede rechazar la hipótesis nula, por tanto, el resultado puede interpretarse como que el patrón espacial del modelo base no es aleatorio, pero dado el índice de correlación tan bajo existente (los datos se correlacionan en alrededor de un 13% a 14%), se genera una respuesta en los modelos altamente no lineal.

6. Conclusiones

- ❖ La alta heterogeneidad de un yacimiento real hace que su modelamiento sea difícil de realizar, es así como, modelos proxy del tipo de regresión polinomial, presentan un muy pobre rendimiento en la replicación de las respuestas del yacimiento dada la no linealidad de los resultados, imposibilitando el uso de esta clase de modelos, pero abriendo las puertas a opciones más sofisticadas como la aplicación de redes neuronales artificiales.

- ❖ Los resultados obtenidos del análisis complementario permiten destacar la no linealidad en la cual se enmarca el comportamiento del modelo de simulación. Este hecho enfatiza la naturaleza pseudo-aleatoria de las respuestas obtenidas por el modelo y es una confirmación de que el uso de modelos proxy convencionales de tipo regresión polinomial, no son una opción viable en estos casos.

- ❖ Los algoritmos de optimización prueban ser métodos efectivos para obtener la ubicación óptima de un nuevo pozo, pero su rendimiento va ser dependiente del tamaño de muestra seleccionado que a su vez determina la cantidad total de simulaciones; por otra parte, no es posible obtener una aproximación de la tendencia del comportamiento global de las variables ya que no se genera, a diferencia de un modelo proxy, una estimación de los datos no simulados.

- ❖ En comparación con los diferentes métodos de optimización aplicados, los modelos proxy presentan un rendimiento superior con respecto al tiempo, así mismo, y, aunque los algoritmos de búsqueda obtuvieron rendimientos cercanos, de los modelos proxy, la red neuronal artificial permitió pronosticar los resultados no simulados de manera acertada, situación que no se presenta con el modelo de regresión polinomial.

- ❖ Es necesario tener en cuenta que los cálculos que conlleva una simulación numérica convencional consumen un cantidad alta de recursos computacionales y tiempo, lo que equivale a la utilización de hardware de última generación, por esta razón, se limitó la optimización a la ubicación de un solo pozo y se decidió no llevar a cabo una optimización de la ubicación de un grupo de pozos, dado que al incrementar las variables a optimizar, estas generan un aumento en el número de los grados de libertad, lo cual se traduce en un incremento exponencial en el número total de simulaciones y un muy considerable aumento en los tiempos empleados.

7. Recomendaciones

- ❖ Investigar y aplicar otras alternativas para validar modelos proxy.

- ❖ Plantear procesos de optimización multi-objetivo por medio de modelos proxy en tareas complejas que permitan la ubicación de varios pozos y la optimización de otras variables de producción de manera simultánea.

- ❖ Utilizar un tipo diferente de red neuronal artificial o redes bayesianas y evaluar su efectividad como modelos proxy.

Referencias bibliográficas

- Aghajan, H., Delgado, R. L.-C., & Augusto, J. C. (2010). *Human-Centric Interfaces for Ambient Intelligence*. Elsevier.
- Al Dossary, M. A., & Nasrabadi, H. (2015). Well Placement Optimization Using Imperialist Competition Algorithm. *SPE International*, 14.
- Al-Mudhafar, W. J., Al-Jawad, M. S., & Al-Shamma, D. A. (2010). Using Optimization Techniques for Determining Optimal Locations of Additional Oil Wells in South Rumaila Oil Field. *SPE International*, 1-20.
- Al-Mudhafer, W. J. (2013). Coupling Programming of an Adaptive Genetic Algorithm with Reservoir Flow Simulation for Infill Oil Well Locations Optimization. *SPE International*, 1-12.
- Beers, K. J. (2007). *Numerical Methods for Chemical Engineering - Applications in MatLAB®*. Cambridge University Press.
- Belegundu, A. D., & Chandrupatla, R. (2011). *Optimization Concepts and Applications in Engineering (2nd Edition)*. Cambridge University Press.
- Chen, C.-H., & Lee, L. (2011). *Stochastic Simulation Optimization - An Optimal Computing Budget Allocation*. World Scientific.
- Computer Modelling Group Ltd. (2015). CMG USER GUIDE.

- Cullick, A., Johnson, D., & Shi, G. (2006). Improved and More.Rapid History Matching With a Nonlinear Proxy and Global Optimization. *SPE International*.
- Dawson, G. F. (2009). Interpretación fácil de la bioestadística. Madrid: ELsevier.
- Esri. (2017). *ArcGIS Pro*. Obtenido de Cómo funciona Autocorrelación espacial (I de Moran global): <https://pro.arcgis.com/es/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm>
- Kuehl, R. O. (2001). *Diseño de Experimentos*. México: THOMSON LEARNING.
- Li, B., & Friedmann, F. (2005). Novel Multiple Resolution Design of Experiment/Response Surface Methodology for Uncertainty Analysis of Reservoir Simulation Forecasts. *SPE International*.
- Minitab Inc. (28 de Julio de 2017). *ASISTENTE DE MINITAB 18*. Obtenido de Prueba t de 2 muestras: https://support.minitab.com/es-mx/minitab/18/Asistente_Prueba_t_de_2_muestras.pdf
- Minitab, Inc. (2017). *Soporte de Minitab 17*. Obtenido de ¿Qué es un valor p?: <http://support.minitab.com/es-mx/minitab/17/topic-library/basic-statistics-and-graphs/introductory-concepts/p-value-and-significance-level/what-is-pvalue/>
- Nikraves, M., & Zadeh, L. (2003). *Soft Computing and Intelligent Data Analysis in Oil Exploration*. Elsevier.
- Osorio Gallego, G. (2002). *Notas sobre simulación numérica de yacimientos*. Medellín: Universidad Nacional, Sede Medellín.

- Pallav, S., & Wen, H. C. (2008). Efficient Well Placement Optimization with Gradient-Based Algorithms and Adjoint Models. *SPE International*, 1-18.
- Sacks, J., Welch, W., Mitchell, T., & Wynn, H. (1989). Design and Analysis of Computer Experiments. *Statistica Science* 4, 409-435.
- Schlumberger. (2017, Enero 18). *Oilfield Glossary*. Retrieved from Schlumberger: http://www.glossary.oilfield.slb.com/Terms/i/infill_drilling.aspx
- Temizel, C., Purwar, S., Urrutia, K., Abdullayev, A., Md Adnan, F., Agarwal, A., & Garcia, A. (2014). Optimization of Well Placement in Real-Time Production Optimization of Intelligent Fields with Use of Local and Global Methods. *SPE International*, 1-12.
- Zhuvok, Y. (30 de 06 de 2017). *Applied Spatial Statistics in R*. Obtenido de <http://scholar.harvard.edu/zhukov/classes/applied-spatial-statistics-r>
- Zubarev, D. (2009). Pros and Cons of Applying Proxy-Models as a Substitute for Full Reservoir Simulations. *SPE International*, 1-23.