

TRABAJO ESPECIAL DE GRADO

APLICACIÓN DEL MÉTODO DE MONTE CARLO EN EXPLORACIÓN Y EXPLOTACIÓN DE PETRÓLEO Y GAS NATURAL

Presentado ante la Ilustre
Universidad Central de Venezuela
para optar al Título de:
Ingeniero de Petróleo
Por el Br. Santaniello S. Filippo F.

Caracas, Octubre de 2003

TRABAJO ESPECIAL DE GRADO

APLICACIÓN DEL MÉTODO DE MONTE CARLO EN EXPLORACIÓN Y EXPLOTACIÓN DE PETRÓLEO Y GAS NATURAL

TUTOR ACADÉMICO: Prof. Armando Azpúrua

Presentado ante la Ilustre
Universidad Central de Venezuela
para optar al Título de:
Ingeniero de Petróleo
Por el Br. Santaniello S. Filippo F.

Caracas, Octubre de 2003

Agradecimientos

A Maylin y Luis por el apoyo

A Reinaldo por el ejemplo

A Yosemite por la motivación

A mis familiares por la atención

Y en especial al Prof. Armando Azpúrua por sus
sabios e incondicionales consejos y toda la ayuda prestada

Índice General

Agradecimientos	i
Lista de Tablas	iv
Lista de Figuras	v
Resumen	vi
Introducción	1
I EL PROBLEMA	3
Planteamiento del Problema	4
Objetivo General	6
Objetivos Específicos	6
Justificación	7
Antecedentes	8
II MARCO TEÓRICO	10
Simulación	11
Estadística	12
Incertidumbre	12
El Método de Monte Carlo	13
Simulación utilizando Monte Carlo	14
Distribución de Frecuencias	23
Histograma de Frecuencias	24
Números Aleatorios y Pseudo Aleatorios	30
Generación de Números pseudo aleatorios con Distribución Uniforme utilizando un computador	31
Generación de Números Pseudo Aleatorios con Distribución distinta a la Uniforme utilizando un Computador.	32
Método de la Transformada Inversa	32
Método del Rechazo	34
Método de Composición	35
Ecuación Volumétrica de Petróleo original en sitio	37
Ecuación Volumétrica de Gas original en sitio	39
Ecuación de Presión Hidrostática	40
Ecuación de Índice de Productividad	42

III MARCO METODOLÓGICO	43
Simulación de Monte Carlo utilizando MCP	44
¿Qué es MCP?	44
Instalación en un PC	45
Uso de MCP	45
Cómo se analizan los resultados de MCP	57
Simulación de Monte Carlo utilizando una hoja de cálculo	60
Conclusión sobre éste método	65
 IV CONCLUSIONES Y RECOMENDACIONES	 66
Conclusiones	67
Recomendaciones	69
 Referencias	 70
Apéndices	74
Determinación del número adecuado de simulaciones para obtener resultados satisfactorios con Monte Carlo	74

Lista de Tablas

Tabla 1 - Resultados de simulación	16
Tabla 2 - Distribución de frecuencias	24
Tabla 3 – Cuartiles mostrados por MCP	56
Tabla 4 – Resultados de las simulaciones con varias variables	76
Tabla 5 – Desviaciones estándar para diversos grupos	79
Tabla 6 – Resultados de las simulaciones con diversas desviaciones estándar	80
Tabla 7 – Desviaciones estándar para diversos grupos	82

Lista de Figuras

Figura 1 - Histograma de frecuencia y frecuencia acumulada de simulación	21
Figura 2 – Distribuciones de frecuencia uniforme	26
Figura 3 - Distribución de probabilidad normal	27
Figura 4 - Distribución de probabilidad lognormal	28
Figura 5 - Distribución de probabilidad exponencial.	29
Figura 6 - Distribución de probabilidad triangular.	30
Figura 7 – Representación gráfica de una distribución triangular	35
Figura 8 –Pantalla principal de MCP	46
Figura 9 – Introducción de datos para área en MCP	49
Figura 10 – Introducción de datos en MCP	50
Figura 11 – Histograma de frecuencias de la simulación en MCP	51
Figura 12 – Frecuencias acumuladas en MCP	52
Figura 13 – resultaos de la simulación en tablas de MCP	53
Figura 14 – Cuartiles de las frecuencias acumuladas en MCP	56

Santaniello S. Filippo F.

APLICACIÓN DEL MÉTODO DE MONTE CARLO EN EXPLORACIÓN Y EXPLOTACIÓN DE PETRÓLEO Y GAS NATURAL

TUTOR ACADÉMICO: Prof. Armando Azpúrua

**Tesis. Caracas, Universidad Central de Venezuela, Facultad de Ingeniería.
Escuela de Ingeniería de Petróleo. 2003.**

Palabras Claves: Simulación, Monte Carlo, probabilística, petróleo

Resumen

En el presente trabajo se hace un análisis del método de Monte Carlo como una herramienta para resolver ecuaciones donde las variables necesarias no se conocen con exactitud, pero se conoce un rango donde pueden estar probablemente los valores reales de dichas variables. Se proporciona información sobre temas importantes en relación al método como lo son la generación de números aleatorios, la generación de histogramas de frecuencia y otros. Se hace un análisis del uso de una hoja de cálculo para una simulación de Monte Carlo y se trabaja con MCP, un programa hecho para respaldar este trabajo, el cual utiliza la simulación de Monte Carlo con cuatro ecuaciones importantes en Ingeniería de Petróleo.

Introducción

Tradicionalmente se ha utilizado el método determinístico para hacer cálculos con ecuaciones de uso típico en diversas áreas. Esto es sencillamente asignar un valor a cada variable de la ecuación y hallar así el valor de una variable dependiente, y ha sido muy común dado que en la mayoría de los casos es un proceso rápido y no complicado.

Sin embargo no siempre es así de fácil; algunas veces las variables no se conocen con exactitud porque fueron medidas con métodos de baja precisión, algunas veces las variables no son una característica puntual sino es una propiedad que cambia a lo largo del tiempo o espacio. En el caso de la ingeniería de petróleo es muy común trabajar con datos de yacimientos, pozos, tuberías y otros donde las mediciones se hacen en condiciones extremas y poco favorables por razones de profundidad, temperatura, presión. En muchas ocasiones las mediciones se hacen de manera indirecta, como por ejemplo el área de un yacimiento a través de la sísmica. En estos casos lo que está haciendo es alimentar una ecuación con data errónea, y es muy probable es que el resultado sea erróneo.

Lo que se busca con este Trabajo Espacial de Grado es un método que, sin conocer las variables de la ecuación con exactitud se pueda igualmente trabajar con ella y obtener una conclusión. Se propondrá para ello un método probabilístico, llamado Simulación de Monte Carlo, que pide no el valor de la variable con exactitud sino un rango probable y una distribución de probabilidad en ese rango,

produciendo una distribución probabilística para la solución de la ecuación.

Dado que es necesario el uso de un computador para realizar simulaciones, se hace un estudio de como utilizar esta herramienta con Monte Carlo utilizando software de uso común y se diseñará un software especial que facilite la aplicación del método de Monte Carlo en ciertas ecuaciones útiles en Ingeniería de Petróleo.

I EL PROBLEMA

Planteamiento del Problema

En Ingeniería de Petróleo, cuando se hacen mediciones de longitudes, alturas, y otras dimensiones suele ocurrir que la precisión de las mismas es baja por diversos motivos. El más importante de ellos es la utilización de métodos indirectos dado que físicamente es imposible hacer la medición en forma directa.

También ocurre que cuando se desea saber el valor numérico de propiedades como espesor, densidad de algún fluido, saturación de fluidos, porosidad y muchas otras, estas son constantes en un punto, pero varían de un lugar a otro, haciendo imposible obtener un valor único. En algunos casos se cuenta con valores promedio, pero es muy difícil asegurar la veracidad de los mismos.

En ciertas ocasiones ocurre algo más problemático, que es ante la ausencia de datos para definir una característica o propiedad, y simplemente se asume el valor que se crea correcto.

El problema surge cuando estos valores son usados en una ecuación para hacer el cálculo de una variable en base a la cual se va a tomar una decisión importante. Si se ingresaron datos con baja precisión o datos erróneos, no se puede esperar otra cosa que un resultado con las mismas condiciones.

Por fortuna, en todos los casos anteriores es posible obtener un rango donde muy probablemente esté el valor real de la variable necesaria.

En muchos de los casos se pueden generar ecuaciones probabilísticas y obtener una ecuación analítica para ellas, sin embargo es un proceso que aún con los computadores más poderosos requiere de mucho tiempo y dedicación.

En vista de esto, surge la siguiente pregunta:

¿Existe algún método rápido y sencillo de hacer los cálculos usuales en Ingeniería de Petróleo sin necesidad de proporcionar valores únicos para las variables sino un rango de valores posibles para las mismas?

Objetivo General

Indagar sobre las ventajas y la aplicación del método de Monte Carlo en áreas referentes a la Ingeniería de petróleo, en especial para la exploración y explotación de petróleo y gas natural, y dotar a la Escuela de Ingeniería de Petróleo de la UCV de las herramientas necesarias para utilizar éste método.

Objetivos Específicos

- Determinar y estudiar los conocimientos relacionados con Estadística, Probabilidades y programación necesarios para realizar una simulación de Monte Carlo.
- Determinar el procedimiento de uso de herramientas para realizar simulaciones de Monte Carlo.
- Diseñar una herramienta que permita realizar simulaciones de Monte Carlo con ecuaciones de uso común en la Ingeniería de Petróleo.
- Proveer a la Escuela de Ingeniería de Petróleo de la Universidad Central de Venezuela una base para la realización de Estudios y Trabajos que requieran la simulación de Monte Carlo.

Justificación

Lo que se plantea a continuación es un método probabilístico que permite hacer cálculos en ecuaciones de uso común, utilizando una técnica llamada simulación, que permite generar diversos escenarios y luego organizar esos escenarios para evaluarlos utilizando estadísticas.

Este método, mejor conocido como Simulación de Monte Carlo, permite operar con una ecuación sin tener valores únicos para cada variable, es decir, es posible ingresar un rango de valores para cada variable que lo necesite.

Sobre el resultado se puede hacer un análisis de cuan probable es la ocurrencia de la variable en cuestión, es decir, un análisis probabilístico, lo que permite definir límites de éxito o fracaso y tomar las decisiones en base a esto.

El método se puede considerar de mucha utilidad para la Industria Petrolera dado que es frecuente el uso de ecuaciones con las condiciones descritas anteriormente y permite entregar conclusiones con márgenes de factibilidad.

Antecedentes

Segn  Coss (1999), durante los a os 40, en plena Segunda Guerra Mundial, se llevaba a cabo en EE.UU. el desarrollo del *Proyecto Manhattan*, mejor conocido como el *Proyecto Bomba At mica*. Consist a, en palabras sencillas, en la fisi n de los  tomos de cierto material que daba origen a la explosi n.

Para conocer los resultados de la fisi n, era necesario saber de que manera actuar a la difusi n de neutrones durante el proceso. Pero resultaba casi imposible saber esto debido a que los movimientos de los  tomos dificultaban conocer su posici n en un determinado instante.

En 1955 el matem tico h ngaro americano John von Neuman y el matem tico polaco Stanislaw Ulam resolvieron el problema programando en el Laboratorio de Los Alamos un m todo que tomaba posiciones probables de los  tomos y arrojaba que tan probable era la ocurrencia de la fisi n. A  sta t cnica la llamaron Simulaci n de Monte Carlo.

Sin embargo, esta no fue la primera vez que se utiliz  una simulaci n para resolver un problema. Previamente en el siglo XIX se hab a aproximado el valor de pi haciendo girar al azar una aguja sobre un tablero con l neas paralelas y observando el n mero de intersecciones entre la aguja y las l neas.

Posteriormente, Ulam junto al físico y matemático norteamericano Nicholas Metropolis trabajaron con la ecuación de mecánica ondulatoria del físico austriaco Erwin Schrödinger y obtuvieron una solución para ella utilizando Monte Carlo.

II MARCO TEÓRICO

Simulación

Una *Simulación* es un proceso en el cual se diseña un modelo de un sistema real y se llevan a cabo experiencias con él. También se habla de *Simulación* como la técnica con la cual se imita el ambiente en el cual ocurre un evento para utilizar un modelo análogo y obtener así información del proceso en cuestión.

Podemos sacar de lo antes dicho que el objetivo de una simulación es básicamente comprender el comportamiento de un sistema frente a diversas situaciones, e inclusive evaluar nuevas estrategias, dentro de los límites que se imponen por un criterio o conjunto de ellos.

Es gracias a esto que desde los años 60 se utiliza la simulación como un método para tomar decisiones estratégicas, dada la habilidad de imitar problemas reales y permitir el análisis de estos a medida que cambian las condiciones de entorno.

La simulación nace en el instante en el cual las ecuaciones no son capaces de adaptarse a un modelo real, ya sea por la ausencia de las mismas o por la no disponibilidad de la data para aplicarlas. Es debido a esto que los resultados obtenidos de una simulación no son precisos, más bien otorgan un rango de posibles resultados que pueden ser interpretados de distintas maneras, pero aún así llenan el vacío que existiría si no se contara con esta herramienta.

Estadística

La palabra estadística tiene su origen en el siglo XVIII, cuando se aplicó a “los arreglos políticos de los estados modernos del mundo conocido”. Se refería solamente a una relación existente entre los mismos. Posteriormente, la entrada de datos numéricos en esos arreglos políticos y su estudio le dio a la palabra un carácter cuantitativo. En la actualidad se puede definir la estadística como una metodología científica que trata de la recolección, presentación y agrupación de los datos, así como el análisis, interpretación, proyección e inferencia de ellos. Más adelante se utilizará la estadística para ordenar los resultados de una simulación y generar una conclusión entorno a la misma.

Incertidumbre

“The more precisely the position is determined, the less precisely the momentum is known in this instant, and vice versa”

La cita anterior corresponde al físico alemán Werner Heisenberg, quien en 1927 la postuló bajo el nombre de *Principio de Incertidumbre*.

Esto lo hizo Heisenberg luego de trabajar en el desarrollo de la bomba atómica, no para el *Proyecto Manhattan* que se nombró anteriormente, sino para la contraparte del mismo, en Alemania, cuando se encontró

con el problema de conocer la posición y el momento de una partícula subatómica. Descubrió que mientras más se conocía una de las propiedades, menos se sabía de la otra en ese instante.

Posteriormente se utilizó la estadística y el Método de Monte Carlo para obtener una solución, como se explico en la sección *Antecedentes*

En la actualidad la palabra incertidumbre se emplea ampliamente y no sólo en cuanto a partículas subatómicas. Se habla de incertidumbre del clima, incertidumbre de un mercado, incertidumbre política. Y cuando se está en la fase exploratoria de un Yacimiento sea de Petróleo o Gas, las variables como área, profundidad, porosidad, saturación de petróleo y muchas otras suelen estar en incertidumbre debido a la falta de información.

Con este término definido, podemos decir que la simulación es una herramienta que se puede utilizar para resolver ciertos problemas donde existe incertidumbre y las maneras tradicionales no otorgan una solución satisfactoria al mismo.

El Método de Monte Carlo

En la definición de simulación se dijo que era un proceso en el cual se diseña un modelo de un sistema real y se llevan a cabo experiencias con él. Lo que hace el Método de Monte Carlo es eso exactamente,

toma una ecuación (modelo) que imita un sistema real y la aplica en distintos escenarios, realizando las experiencias. Es por esto que este método es también conocido como Simulación de Monte Carlo.

La técnica de ésta simulación consiste en alimentar una ecuación con números aleatorios, algo así como los que se obtienen en una ruleta de un casino. Y es precisamente esta similitud la que le da a este método su nombre, siendo el casino de Monte Carlo en el Principado de Mónaco la inspiración.

Simulación utilizando Monte Carlo

Para hacer una simulación de Monte Carlo se necesita una ecuación la cual será el objeto de la simulación y los datos para la misma. La diferencia de este método con el tradicional para resolver una ecuación es que Monte Carlo toma en cuenta la incertidumbre y los datos que alimentarán las variables de la ecuación no deben ser necesariamente un valor constante sino pueden ser un grupo de ellos.

Al simular, la ecuación se calcula varias veces, y en cada cálculo se escoge un valor al azar para cada valor de las variables independientes, cuidando que estos valores tengan la misma distribución de probabilidad que la variable física. Con esto se obtienen tantos resultados como cálculos hechos, los cuales se utilizan para generar un gráfico conocido como histograma de

frecuencias (del cual se hablará mas adelante) y de allí se realizan una serie de análisis.

A continuación, paso a paso y con un ejemplo se demuestra como se realiza una simulación utilizando el método de Monte Carlo.

Se utilizará la ecuación volumétrica de petróleo original en sitio para un yacimiento. Esta ecuación es:

$$POES = \frac{7758 * \hat{A} * h * \hat{o} * S_i}{\hat{a}_i} \quad \text{Ecuación 1}$$

El objetivo es obtener el *POES*. Desafortunadamente no se tienen algunos de los valores necesarios para calcularlo. Solamente se han perforado algunos pozos y se sabe que el área *A* puede tener como mínimo 100 acres y puede llegar a tener 150 acres. El espesor *h* medido en donde se perforaron los pozos era de 20 pies, pero el geólogo asegura que está entre 15 y 30 pies por los datos de la formación. La porosidad $\hat{o}$ medida fue de 20% pero en arenas similares se ha encontrado que va de 10 a 21%. La saturación de petróleo se sospecha está entre 60 y 70%. El factor volumétrico $\hat{a}_o$ se calculó mediante una prueba PVT y se concluyó que es el valor constante 1,1 BY/BF.

Utilizando el método básico de cálculo sería imposible conocer el *POES* en este caso porque no se tiene un valor único para cada variable sino rangos de valores. Es aquí donde entra Monte Carlo, que

si bien no entrega un valor de *POES*, permite hacer un análisis y entregar un resultado probabilístico.

Se realiza la primera simulación: para empezar es necesario escoger los valores que se usarán en la ecuación. Como se dijo que el área está entre 100 y 150 acres, se escoge un valor al azar dentro de este rango. Utilizando el computador se obtuvo el valor de 114,65 acres. El mismo proceso se hace para el espesor, la porosidad y la saturación de petróleo. El factor volumétrico se toma como su valor constante, 1,1. El primer *POES* que se obtiene es 1,31 MMBbls.

El mismo procedimiento se realiza tantas veces como se considere necesario. En este caso se va a realizar 100 veces. Se obtienen entonces 100 resultados que son los que se muestran en la Tabla 1:

Tabla 1 - Resultados de simulación

A	h	ö	So	âo	POES
114,65	22,89	0,11	0,65	1,1	1,31
149,84	22,17	0,20	0,61	1,1	2,77
122,93	20,28	0,16	0,61	1,1	1,75
102,97	24,03	0,14	0,65	1,1	1,56
119,98	24,05	0,19	0,60	1,1	2,27
146,99	19,00	0,11	0,64	1,1	1,45
105,42	24,49	0,18	0,67	1,1	2,20
114,36	17,62	0,18	0,64	1,1	1,65
140,45	23,20	0,15	0,66	1,1	2,24

113,25	20,46	0,20	0,68	1,1	2,25
148,80	28,66	0,20	0,64	1,1	3,84
135,81	29,53	0,18	0,60	1,1	3,10
133,08	19,21	0,11	0,62	1,1	1,29
135,53	17,34	0,10	0,65	1,1	1,08
130,28	16,96	0,11	0,62	1,1	1,01
148,65	28,40	0,19	0,62	1,1	3,45
105,59	27,35	0,16	0,67	1,1	2,16
103,47	16,48	0,17	0,69	1,1	1,42
141,11	22,56	0,11	0,62	1,1	1,58
135,81	27,16	0,13	0,69	1,1	2,41
103,21	26,11	0,14	0,61	1,1	1,68
144,39	29,31	0,19	0,66	1,1	3,81
115,89	24,22	0,13	0,62	1,1	1,63
133,12	19,84	0,17	0,66	1,1	2,06
145,74	21,48	0,18	0,62	1,1	2,48
122,76	26,86	0,16	0,63	1,1	2,32
112,90	16,41	0,12	0,60	1,1	0,92
120,33	25,05	0,21	0,67	1,1	2,93
140,90	24,16	0,16	0,64	1,1	2,43
129,46	24,02	0,11	0,62	1,1	1,56
100,02	16,23	0,20	0,67	1,1	1,54
133,54	15,22	0,12	0,61	1,1	1,06
120,49	17,28	0,14	0,67	1,1	1,35
133,97	20,37	0,20	0,65	1,1	2,43
141,18	18,15	0,11	0,62	1,1	1,21
119,94	29,57	0,21	0,64	1,1	3,35
145,38	23,81	0,15	0,70	1,1	2,55
114,75	23,85	0,18	0,64	1,1	2,24

121,73	23,15	0,10	0,66	1,1	1,32
101,62	20,84	0,11	0,68	1,1	1,09
137,74	24,04	0,12	0,60	1,1	1,71
142,41	27,36	0,11	0,69	1,1	2,17
135,90	15,09	0,18	0,66	1,1	1,74
100,28	17,53	0,11	0,65	1,1	0,89
112,08	18,53	0,14	0,61	1,1	1,26
102,87	24,01	0,14	0,69	1,1	1,70
121,47	20,87	0,13	0,70	1,1	1,62
101,26	17,74	0,19	0,61	1,1	1,45
142,85	18,18	0,12	0,63	1,1	1,45
138,33	16,81	0,21	0,65	1,1	2,23
107,53	28,07	0,16	0,68	1,1	2,36
101,25	25,52	0,15	0,63	1,1	1,74
126,05	26,93	0,20	0,62	1,1	2,97
149,24	22,95	0,13	0,62	1,1	1,91
144,74	26,95	0,13	0,64	1,1	2,25
129,92	25,90	0,11	0,65	1,1	1,71
138,97	22,76	0,10	0,63	1,1	1,46
117,51	23,86	0,19	0,62	1,1	2,35
140,13	15,79	0,11	0,60	1,1	0,99
116,78	29,43	0,16	0,69	1,1	2,71
114,14	22,02	0,17	0,64	1,1	1,96
107,76	22,09	0,18	0,64	1,1	1,99
102,70	16,64	0,15	0,64	1,1	1,13
137,05	16,33	0,17	0,67	1,1	1,78
112,23	27,92	0,15	0,67	1,1	2,16
132,55	29,66	0,15	0,64	1,1	2,65
114,30	22,77	0,12	0,63	1,1	1,44

103,09	29,84	0,11	0,69	1,1	1,57
140,90	15,61	0,18	0,67	1,1	1,84
135,10	22,22	0,15	0,63	1,1	2,05
143,67	29,62	0,17	0,70	1,1	3,48
111,20	21,87	0,12	0,62	1,1	1,24
127,58	21,07	0,11	0,67	1,1	1,44
148,37	23,59	0,16	0,65	1,1	2,56
117,63	17,35	0,10	0,64	1,1	0,97
142,69	21,09	0,15	0,69	1,1	2,25
148,74	16,96	0,11	0,70	1,1	1,33
111,82	28,77	0,19	0,61	1,1	2,66
132,35	29,52	0,10	0,68	1,1	1,89
133,07	22,39	0,20	0,64	1,1	2,67
118,50	26,65	0,16	0,65	1,1	2,35
129,79	28,07	0,14	0,67	1,1	2,48
135,17	17,35	0,14	0,66	1,1	1,49
114,35	24,61	0,20	0,61	1,1	2,44
128,96	23,45	0,17	0,61	1,1	2,28
142,91	22,87	0,14	0,62	1,1	2,07
135,10	26,73	0,11	0,65	1,1	1,83
126,11	27,22	0,21	0,61	1,1	3,11
130,75	18,37	0,17	0,64	1,1	1,82
114,92	28,09	0,20	0,67	1,1	3,01
110,47	23,09	0,19	0,63	1,1	2,21
125,42	25,31	0,19	0,64	1,1	2,67
138,50	15,37	0,14	0,62	1,1	1,27
122,80	20,04	0,20	0,70	1,1	2,39
140,55	17,46	0,18	0,61	1,1	1,91
148,90	15,11	0,13	0,60	1,1	1,20

100,24	21,18	0,12	0,67	1,1	1,15
117,38	16,95	0,15	0,61	1,1	1,28
114,10	29,79	0,21	0,64	1,1	3,16
144,43	25,35	0,20	0,65	1,1	3,29

El siguiente paso es analizar los valores obtenidos.

En primer lugar se sumarán todos los valores de *POES* obtenidos y el resultado se dividirá entre 100 para obtener el promedio, el cual es 1,9989 MMBBls. Esto indica cual es la tendencia central del valor a obtener.

Luego se observa que el menor valor es 0,89 MMBBls y el mayor es 3,84 MMBBls. Estos serán llamados límites e indican el rango de los valores dentro de los cuales puede oscilar el valor que tratamos de calcular que es el *POES*.

Adicionalmente podemos calcular la varianza que es 0,4726 y la desviación estándar que es 0,6874. Estos valores indican que tan dispersos están los distintos valores de *POES* obtenidos.

Estos resultados son útiles en una simulación, pero se puede obtener mucho más. Se hará a continuación una división de todos los resultados de *POES* obtenidos en subgrupos de igual tamaño. La cantidad de subgrupos a usar es flexible y ha de ser escogida por el usuario. Algunos autores sugieren que es conveniente utilizar tantos subgrupos como la raíz cuadrada del número de muestras, es por eso

que se utilizaran diez subgrupos. Lo que se obtendrá es un histograma de frecuencias, y es el que se muestra en la Figura 1

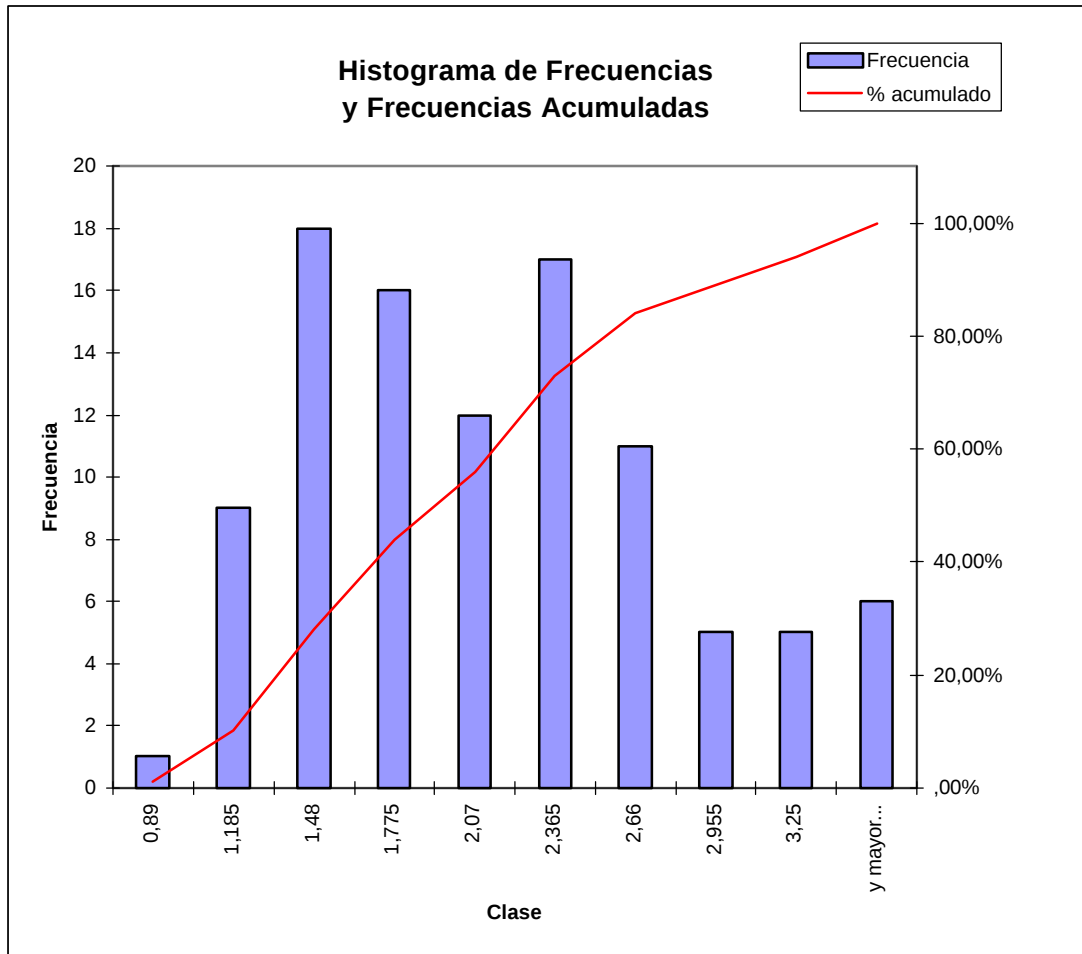


Figura 1 - Histograma de frecuencia y frecuencia acumulada de simulación

Éste gráfico (Figura 1) se obtuvo de una manera sencilla utilizando una hoja de cálculo, posteriormente se explicará con detalle la división de los valores en grupos, el gráfico y el proceso con la hoja de cálculo.

Se puede observar que existen dos datos en él, las barras azules que son el histograma de frecuencias de los valores obtenidos para el *POES*. De éste histograma se puede obtener una importante información, que es cuan probable es obtener un resultado en un rango dado. Por ejemplo, en la barra que corresponde a 1,48, lo que quiere decir que corresponde al subgrupo que va de 1,48 a 1,775, se puede observar que hay 18 valores, lo que indica que 18 de cada 100 valores está en ese rango o lo que es equivalente, hay un 18 por ciento de probabilidad de que el *POES* tenga un valor que va de 1,48 a 1,775 MMBBl.

La línea roja representa las frecuencias acumuladas. Es sumamente útil y en especial para tomar decisiones de orden económico. Supóngase por ejemplo que esta simulación responde a un análisis económico de aceptación para explotar un yacimiento y se sabe que para que sea rentable ha de tener un *POES* mayor a 1,48 MMBBl de petróleo. Cómo no se cuenta con el valor real de *POES* por razones descritas en la justificación de este trabajo de grado, se toma la decisión por probabilidades. De la curva de frecuencias acumuladas obtenemos que para 1,48 MMBBl o menos hay un 30% de probabilidad, es decir, existe un 70% de probabilidad de que el *POES* sea mayor a 1,48 MMBBl. Corresponderá entonces al analista decidir en base a esta simulación.

Por fortuna, el método de Monte Carlo no llega hasta aquí. Cada una de las variables que se tomó en cuenta para simular no tiene que ser sencillamente tomada al azar en un rango dado con la misma probabilidad para cada valor. El método permite la entrada de

variables con zonas de mayor probabilidad y menor probabilidad. Esto se llama distribución de probabilidad.

Antes de continuar es necesario aclarar un punto que es la correlación entre las variables independientes. Tal como se ha planteado, el Método de Monte Carlo supone que los datos de entrada al modelo son totalmente independientes entre sí. Sin embargo, esto no es siempre así; en algunas ocasiones existe una fuerte correlación entre dos o más de estas variables. En esos casos es conveniente utilizar estas correlaciones al generar cada variable de manera aleatoria en función de la otra variable de la cual ésta es dependiente. Esto no es tomado en cuenta por considerarse fuera del alcance de este Trabajo.

Distribución de Frecuencias

Anteriormente se procedió a dividir los valores obtenidos en diez subgrupos, de manera que se contabilizó la cantidad de valores que estaban en el grupo que va de 0.89 a 1.185, luego se hizo lo mismo para los que estaban entre 1.185 y 1.48 y así sucesivamente. El resultado se muestra en la tabla 2

Tabla 2 - Distribución de frecuencias

Rango	Incidencias
0,89 a 1,185	1
1,185 a 1,48	9
1,48 a 1,775	18
1,775 a 2,07	16
2,07 a 2,365	12
2,365 a 2,66	17
2,66 a 2,955	11
2,955 a 3,25	5
3,25 a 3,545	5
3,545 a 3,84	6

A esta tabla se le llama Distribución de Frecuencias, y se define como una disposición tabular de los datos por clases junto con las correspondientes frecuencias de clase. A cada subgrupo se le llama clase y al el número de veces que se repite un valor dentro de una clase se le llama frecuencia.

Histograma de Frecuencias

Un histograma de frecuencias es un gráfico de barras que muestra las frecuencias para distintas clases.

Spiegel (1998) afirma que un histograma de frecuencias suele tener entre 5 y 20 clases. Otros expertos indican que el número de clases

conveniente es aproximadamente la raíz cuadrada del número total de datos. Aunque ambas propuestas son bastante buenas en la práctica, es posible tomar una de ellas, generar el histograma de frecuencias y posteriormente cambiar el número de clases.

Si cada frecuencia del histograma representara una probabilidad podríamos obtener una *Distribución de Probabilidades*; para ello sería necesario que la suma de las frecuencias fuera uno, lo que se logra dividiendo cada frecuencia entre la suma de todas. Dado que son números discretos, se obtiene *Distribución de Probabilidades Discreta*. Una *Distribución de Probabilidades Continua* sería aquella en la que las clases son infinitamente pequeñas, quedando en lugar de gráfico de barras una curva.

La *Distribución de Probabilidades* es un tema muy amplio y el cual no es el objetivo de este trabajo, sin embargo puede ser consultado en cualquier texto de probabilidades.

Las *Distribuciones de Probabilidades Continuas* tienen una definición matemática que las caracteriza, alguna de ellas son muy comunes. Una básica es la conocida como *Distribución de Probabilidad Uniforme*. Se define como:

$$f(x) = 1/(b-a) \text{ en } a \leq x \leq b \text{ y } f(x) = 0 \text{ en otro rango} \quad \text{Ecuación 2}$$

donde b es el límite superior en el que existe la función y a es el límite inferior. A esta función se le conoce como *función de densidad de probabilidad*.

En base a esta definición en la Figura 3 se observan tres ejemplos de distribución uniforme con distintos límites en cada una.

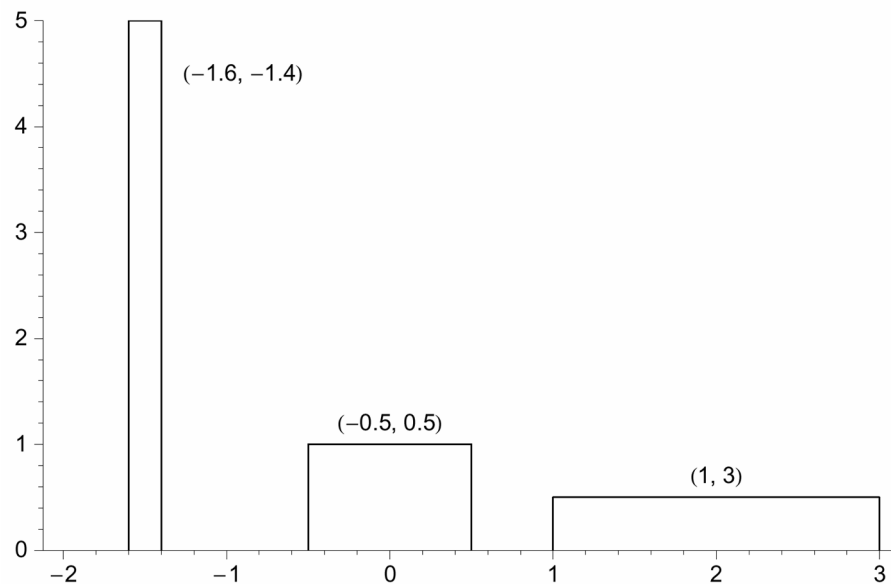


Figura 2 – Distribuciones de frecuencia uniforme
Tomado de McLaughlin (1999)

McLaughlin (1999) afirma que la función de densidad de una distribución de probabilidad muy común llamada normal es la siguiente:

$$f(x) = \frac{1}{\sqrt{2\sigma^2}} * e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{en } -\infty < x < \infty \quad \text{Ecuación 3}$$

Siendo $\bar{x}$ el promedio de los valores de la distribución y σ^2 la varianza de dicha distribución. La Figura 4 muestra dos ejemplos de esta distribución de probabilidad

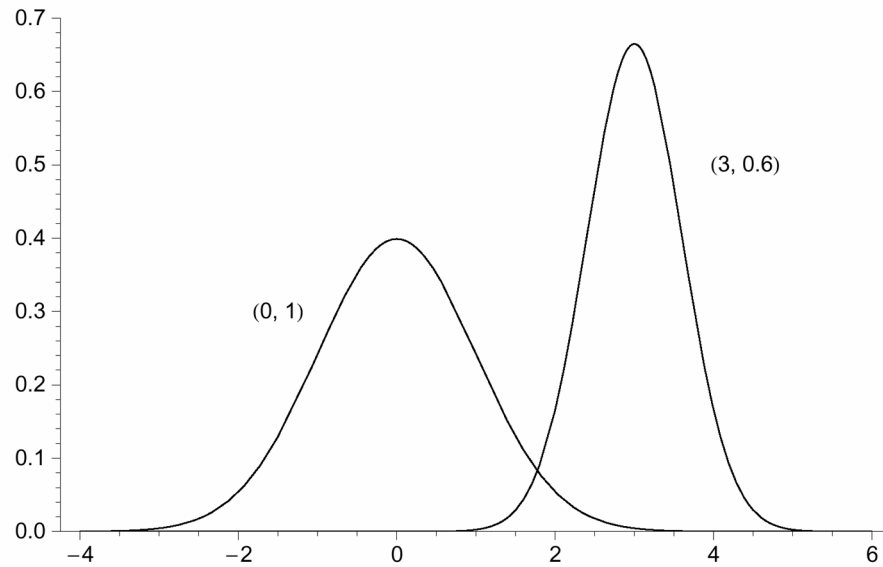


Figura 3 - Distribución de probabilidad normal
Tomado de McLaughlin (1999)

Lo que se observa en la Figura 4 son dos distribuciones de probabilidad normal; la de la izquierda con media cero y desviación estándar 1 y la de la derecha con media tres y desviación estándar 0,6.

Una tercera distribución muy común es la distribución de probabilidad lognormal. Se le llama así porque si bien las frecuencias no se distribuyen normalmente, si lo hacen los logaritmos de estas. De McLaughlin (1999)

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} * e^{\frac{-(\log(x)-\mu)^2}{2\sigma^2}}$$

Ecuación 4

Siendo nuevamente μ el promedio de los valores de la distribución y σ la varianza. La forma que toma esta distribución de probabilidad se ve en la Figura 5 la cual muestra dos ejemplos.

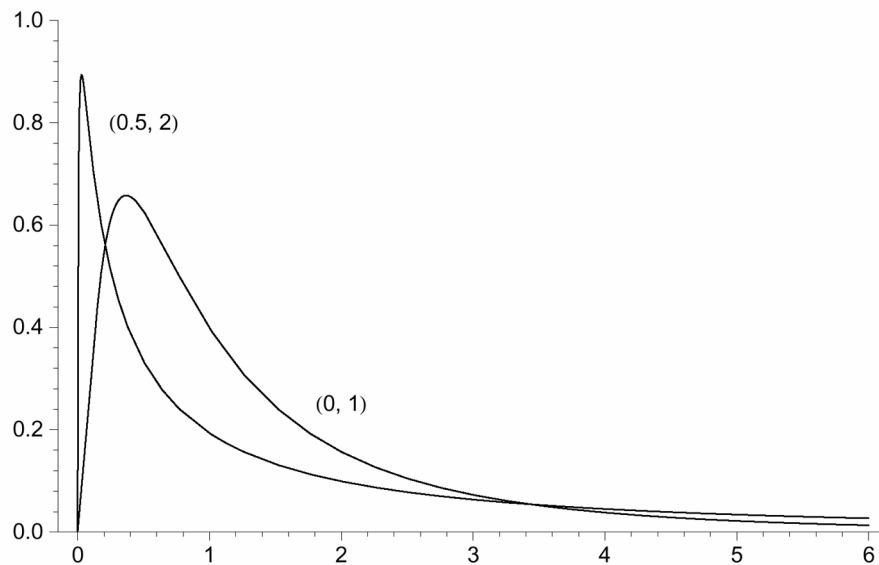


Figura 4 - Distribución de probabilidad lognormal
Tomado de McLaughlin (1999)

Coss (1999) muestra la función de densidad de la distribución de probabilidad exponencial, que es la siguiente:

$$f(x) = \lambda e^{-\lambda x} \text{ para } x \geq 0 \text{ y } f(x) = 0 \text{ para } x < 0$$

Ecuación 5

siendo λ la media de la distribución. Dos ejemplos de esta distribución de probabilidad exponencial se muestran en la Figura 6

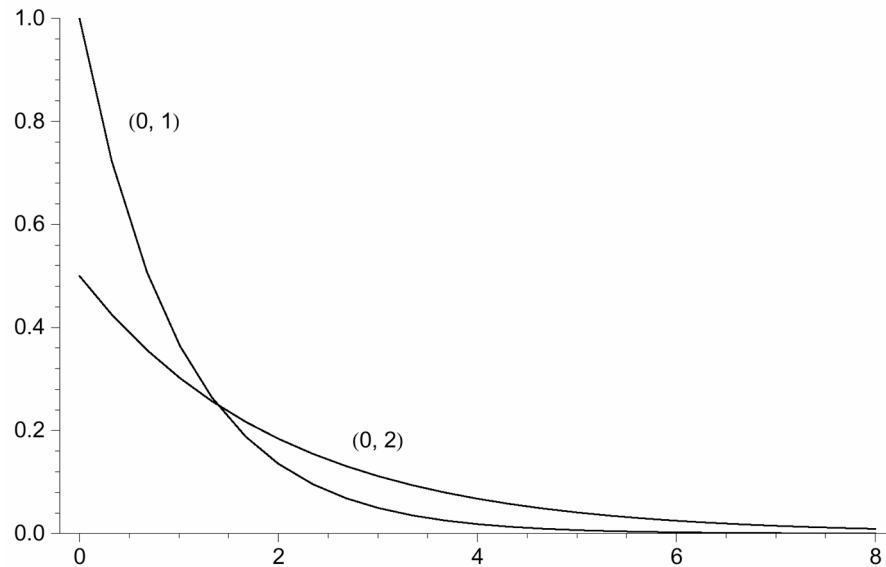


Figura 5 - Distribución de probabilidad exponencial.
Tomado de McLaughlin (1999)

La quinta distribución de probabilidad de la que se hablará es la triangular, cuya función de densidad es:

$$f(x) = (2(x-a)) / ((c-a)(b-a)) \text{ para } x < c \quad y$$

$$f(x) = (2(x-c)) / ((c-a)(c-b)) \text{ para } x > c$$

$$f(x) = 0 \text{ para } a > x > c$$

Ecuación 6

Su definición es compuesta por dos secciones porque la su función de densidad son dos líneas que parten desde a y b , los límites inferior y superior respectivamente e interceptándose en c , el valor más probable dentro de esos límites. El ejemplo se muestra en la Figura 7

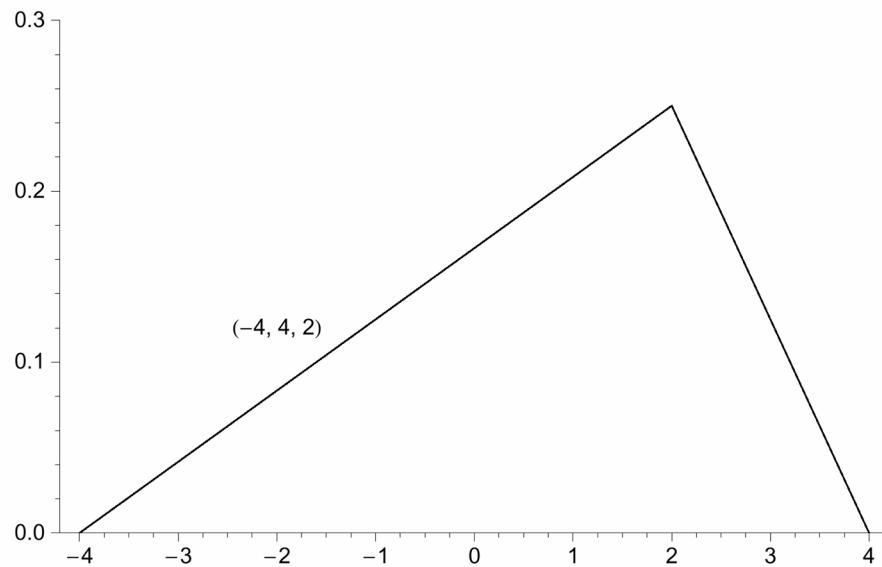


Figura 6 - Distribución de probabilidad triangular.
Tomado de McLaughlin (1999)

Números Aleatorios y Pseudo Aleatorios

La Enciclopedia Encarta describe algo aleatorio como “relativo al juego de azar” o “dependiente de algún suceso casual”. Es así que se puede aplicar este adjetivo a un evento que ocurre sin ser esperado. Es importante destacar que es posible saber que el evento ocurrirá y aún así seguir siendo aleatorio. Un ejemplo de esto es la caída de un rayo. Los pronósticos meteorológicos pueden indicar una zona propensa a la caída de rayos. La ubicación exacta y el momento ocurren al azar. Para ser mas exacto en la definición de algo aleatorio es necesario incluir que el evento ocurra sin intervención del hombre.

Tomando en cuenta esto se puede entonces definir números aleatorios como números que se toman al azar de un grupo. Un

ejemplo de esto sería el número de frutos que carga una planta de manzanas.

Si se trabaja en un computador, es posible generar números aleatorios. Para ello es necesario un “input” aleatorio del entorno. De lo contrario los números generados serán pseudo aleatorios, es decir, números que sin ser aleatorios tienen sus propiedades.

Generación de Números pseudo aleatorios con Distribución Uniforme utilizando un computador

La implementación del Método de Monte Carlo de este trabajo de grado se hará utilizando el lenguaje de programación Delphi 6, y es por esto que se enfocará la generación de números pseudo aleatorios con este compilador.

La técnica para generar un número pseudo aleatorio es a través de la función *random*. Ésta devuelve números de distribución uniforme entre cero y uno generados a partir de secuencias predefinidas. Sin embargo, aunque los números de la secuencia son efectivamente números pseudo aleatorios uniformemente distribuidos, son siempre los mismos, así que no se pueden considerar pseudo aleatorios como tal.

Lo que resuelve esto es que existen muchas secuencias, exactamente 2^{32} (aunque este número cambia según la versión del compilador). Utilizando el reloj del computador y haciendo operaciones con la hora, minuto, segundo y milisegundo en que se hace la operación, la función *randomize* escoge una de estas secuencias y entrega los números pseudo aleatorios uniformemente distribuidos.

Generación de Números Pseudo Aleatorios con Distribución distinta a la Uniforme utilizando un Computador.

Ya que se cuenta con números aleatorios uniformemente distribuidos entre 0 y 1 proporcionados por el computador, es posible obtener otras distribuciones utilizando éstos y un método de conversión. Existen básicamente tres métodos con los cuales se obtiene la mayoría de dichas distribuciones. A continuación se hablará de ellos, sin embargo es común la aparición constante de nuevos métodos que suelen cubrir necesidades específicas.

Método de la Transformada Inversa

Este método utiliza la distribución acumulada de los números aleatorios. Coss (1999) indica que consiste en asumir que la transformada del número aleatorio es igual a un número aleatorio con distribución uniforme entre 0 y 1, es decir:

$$F(x) = R$$

Ecuación 8

Aplicando luego la transformada inversa se tiene que:

$$x = F^{-1}(R)$$

Ecuación 9

Este método es útil para obtener números aleatorios exponencialmente distribuidos. Partiendo de la conocida función de distribución de probabilidad exponencial

$$f(x) = \lambda e^{-\lambda x} \text{ para } x \geq 0 \text{ y } f(x) = 0 \text{ para } x < 0$$

Ecuación 10

y obteniendo la distribución acumulada de la misma

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$$

Ecuación 11

Esto es equivalente a un número aleatorio R con distribución uniforme entre 0 y 1

$$1 - e^{-\lambda x} = R$$

Ecuación 12

Finalmente despejando x

$$x = -(1/\lambda) \ln(R)$$

Ecuación 13

obteniéndose x , cuya distribución es exponencial. $\bar{x}$ es el valor promedio de los números de la distribución.

Dado que los números aleatorios con distribución uniforme que provee el computador están entre 0 y 1, es posible obtener un nuevo rango utilizando este mismo método. Se obtiene de la misma manera la función de distribución acumulada y se siguen los mismos pasos. Finalmente se tiene

$$x = a + (b - a)R$$

Ecuación 14

siendo en este caso a y b los límites inferior y superior de la distribución uniforme. Estos dos casos son utilizados en *MCP* para generar los números aleatorios necesarios.

Método del Rechazo

Coss indica que este método basa su funcionamiento en analizar un par de números aleatorios uniformemente distribuidos entre 0 y 1. Si cumplen con la condición impuesta sobre los límites de la distribución deseada se aceptan, de lo contrario son rechazados y se genera un nuevo par de ellos. De allí su nombre.

Se utiliza este método en la distribución del tipo triangular. Con los datos de los puntos del triángulo se obtienen los rangos de operación y se generan 2 números aleatorios R_1 y R_2 de distribución uniforme entre 0 y 1. Se evalúa si ambos valores entran en el rango anteriormente nombrado, si es así se pasa a calcular el nuevo número aleatorio con distribución triangular. De lo contrario, se descartan

ambos valores aleatorios de distribución uniforme y se generan dos nuevos hasta que el valor de los mismos sea satisfactorio.

Es evidente que la eficacia de este método es baja por la repetitiva búsqueda de valores de los cuales muchos de ellos no serán utilizables y serán necesarios otros de ellos. Afortunadamente existe otro método que resuelve este problema, y del cual se habla a continuación.

Método de Composición

Este método consiste en dividir la función de distribución de probabilidad deseada en sub-áreas y obtener la distribución acumulada a partir de esto como una sumatoria de las mismas. Posteriormente con el número aleatorio R_1 y utilizando la transformada inversa se escoge la sub-área sobre la cual se trabajará y con el número aleatorio R_2 se obtiene el nuevo valor aleatorio con la distribución deseada.

La representación gráfica de la distribución triangular se muestra en la Figura 8:

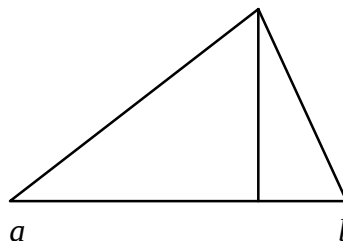


Figura 7 – Representación gráfica de una distribución triangular

donde a y c son el mínimo y el máximo valores posibles de la distribución y b el valor más probable. Dado que es una función de distribución de probabilidad, el área del triángulo es 1. Luego su altura es $2/(c-a)$.

Se toman dos sub-áreas las cuales serán de a hasta b y de b hasta c . La función de probabilidad se expresa entonces en base a las dos sub-áreas, esto es:

$$f(x) = (2(x-a)) / ((c-a)(b-a)) + (2(x-c)) / ((c-a)(c-b)) \quad \text{Ecuación 15}$$

Se generan los números aleatorios R_1 y R_2 y se determina si se cumple la condición

$$R_1 < (b-a)/(c-a) \quad \text{Ecuación 16}$$

de ser cierto el nuevo numero aleatorio será

$$x = a + (b-a) R_2 \quad \text{Ecuación 17}$$

es decir, pertenece a la primera sub-área del triángulo. De lo contrario se tiene

$$x = c - (c-b) R_2 \quad \text{Ecuación 18}$$

para pertenecer a la segunda sub-área.

Ecuación Volumétrica de Petróleo original en sitio

Las trampas en el subsuelo generalmente se forman por características estructurales y estratigráficas que se presentan en las partes más porosas y permeables de los estratos de la corteza terrestre. Estos están conformados principalmente por arenas, areniscas, calizas y dolomitas, las cuales presentan espacios porosos debido a diaclasas, fracturas o efectos de disolución del material rocoso, inclusive debido a una mezcla de estos.

Aquella parte de una trampa que contiene petróleo, gas o ambos como un solo sistema hidráulico conectado se llama Yacimiento.

Una manera de cuantificar la cantidad de petróleo que existe en un Yacimiento es a través de un Balance Volumétrico, es decir, cuantificar el volumen neto de petróleo en base al volumen del yacimiento.

El primer punto a considerar es el volumen bruto del yacimiento, es decir, todo el volumen del mismo. Aunque existen diversas maneras de hacer esto, para un yacimiento se hará suponiendo que tiene forma de paralelepípedo, de manera que el volumen estará dado por:

$$V = A * h$$

Ecuación 19

siendo V el volumen del yacimiento, A el área de la base y h la altura. Si se considera que solo la parte porosa de esta área puede ser ocupada por petróleo, se puede obtener el volumen máximo ocupable por fluidos en esas condiciones. Luego:

$$V_o = A * h * \phi$$

Ecuación 20

donde V_o es el volumen ocupable y ϕ la porosidad expresada como fracción. Dado que no todo el volumen ocupable se encuentra lleno de petróleo, es necesario incluir la fracción de volumen que si lo está. Esto se llama saturación de petróleo y se representará como S_o . Entonces se tiene:

$$V_{op} : A * h * \phi * S_o$$

Ecuación 21

donde V_{op} es el volumen ocupado por petróleo y el cual está a condiciones de yacimiento, es decir, a la presión y temperatura del mismo. Para conocer el volumen de petróleo a condiciones de superficie se utiliza el factor o $\hat{\alpha}_o$, el cual es una relación entre los volúmenes de petróleo en el yacimiento y en superficie y se expresa en barriles de yacimiento por cada barril fiscal (BY/BF). Aplicándolo a la ecuación anterior es posible saber el volumen de petróleo en superficie o $POES$ (petróleo original en sitio) que es:

$$POES : A * h * \phi * S_o / \hat{\alpha}_o$$

Ecuación 22

Dado que el área se suele expresar en acres y la altura en pies, el *POES* se obtendría en acre pie, pero dado que el petróleo generalmente se expresa en barriles, se utiliza el factor de conversión de acre pie a barril, obteniéndose:

$$POES : 7758 * A * h * \phi * S_o / \alpha_o$$

Ecuación 23

Ecuación Volumétrica de Gas original en sitio

En el punto anterior se definió el volumen máximo ocupable por fluidos de la siguiente manera:

$$V_o = A * h * \phi$$

Ecuación 24

siendo *A* el área del yacimiento, *h* la altura y ϕ la porosidad. La parte de este volumen ocupada por gas se obtiene con la saturación de gas, *S_g*, quedando:

$$V_{og} : A * h * \phi * S_g$$

Ecuación 25

siendo *V_{og}* el volumen ocupado por gas. De nuevo para saber este volumen a condiciones de superficie, se utiliza el factor volumétrico α_g en pies cúbicos de yacimiento por cada pie cúbico fiscal, y se obtiene el *GOES* o gas original en sitio:

$$GOES : A * h * \bar{o} * Sg / \hat{a}g$$

Ecuación 26

Finalmente para llevar el *GOES* de acre pie a pies cúbicos se utiliza el factor de conversión 43560

$$GOES : 43560 * A * h * \bar{o} * Sg / \hat{a}g$$

Ecuación 27

Es importante destacar lo siguiente con respecto a los factores volumétricos $\hat{a}g$ y $\hat{a}o$: las unidades en las que se expresen estos factores no alteran de forma alguna el valor del mismo, es decir, si el $\hat{a}o$ de cierto tipo de petróleo es 1,2 BY/BF, si se desea en litros de yacimiento por litro fiscal, será también 1,2 debido a que estamos hablando del porcentaje de variación de un volumen cuando cambian sus condiciones y al cambiar el volumen ese porcentaje se mantendrá igual. La razón por la que el $\hat{a}o$ y el $\hat{a}g$ se utilizan en BY/BF y PCY/PCF respectivamente es para que tengan las mismas unidades que el resto de las variables. Sin embargo hay que aclarar que si se expresan estos factores en forma inversa, este valor si ha de variar, es decir, si se desea el $\hat{a}g$ en PCF/PCY en lugar del inverso, es necesario invertir el valor anterior de $\hat{a}g$.

Ecuación de Presión Hidrostática

De Haby (s. f.) la ecuación diferencial hidrostática es la siguiente:

$$dP / dh = - r * g$$

Ecuación 28

Siendo dP la variación de presión, dh la variación de altura, r la densidad del fluido y g la aceleración de gravedad. Resolviendo esta ecuación se tiene:

$$\int_0^P dP = \int_0^h -\tilde{r} g dh$$

Ecuación 29

e integrando

$$P = - r g h$$

Ecuación 30

expresando la aceleración de gravedad en sistema ingles y haciendo los cambios de unidades requeridos se tiene

$$P = 0,052 r h$$

Ecuación 31

y es ésta la ecuación que se utiliza para calcular la presión hidrostática en libras por pulgada cuadrada en un pozo de profundidad h en pies lleno con fluido de densidad r en libras por galón.

Ecuación de Índice de Productividad

En la Industria Petrolera se suele medir la eficiencia de un pozo a través de un término conocido como Índice de Productividad (IP). Nind (1964) indica que este índice se expresa como una relación entre la tasa (Q) de producción de líquido de un pozo y el *drawdown* que se genera para esa tasa.

El *drawdown* a su vez se define como la diferencia entre la presión estática (P_s) del yacimiento y la presión en el fondo del pozo mientras éste produce fluido (conocida también como presión de fondo fluyente P_{wf}).

Es así que el índice de productividad (IP) puede ser medido con la siguiente ecuación:

$$IP = Q / (P_s - P_{wf})$$

Ecuación 32

III MARCO METODOLÓGICO

Simulación de Monte Carlo utilizando MCP

¿Qué es MCP?

MCP son la iniciales de la frase “Monte Carlo para Petróleo” y es una herramienta que facilita la utilización del Método de Monte Carlo en algunas ecuaciones de Ingeniería de Petróleo, proporcionando un medio sencillo, rápido y directo para hacer cálculos con variables probabilísticas.

MCP se presenta a través de un programa para el sistema operativo Windows, diseñado, desarrollado y compilado utilizando Delphi, un lenguaje visual de bajo nivel de última generación.

Al poner MCP en funcionamiento se presenta una ventana en la cual el usuario ingresa los valores probabilísticos de las variables que se manejan en la ecuación en uso y MCP realiza un numero elevado de simulaciones, ordena los resultados y los presenta a través de gráficos y tablas. Sobre este tema tenemos más en el próximo punto.

Instalación en un PC

MCP se presenta en un disco compacto con la característica de auto instalación. Lo que debe realizarse es introducir el disco en una unidad lectora de una computadora con sistema operativo *Microsoft Windows 95 o superior* y la instalación comenzará automáticamente, solo necesita seguir las instrucciones en pantalla.

Uso de MCP

El diseño de MCP fue enfocado en ser sencillo de usar presentando de entrada una ventana que pide los datos para la simulación. La imagen que se muestra en la figura 9 corresponde a la pantalla principal de MCP. Para se explicado su uso de una manera sencilla, se indican en la figura seis secciones que van de la *A* a la *E* y que se comentaran con detalle a continuación

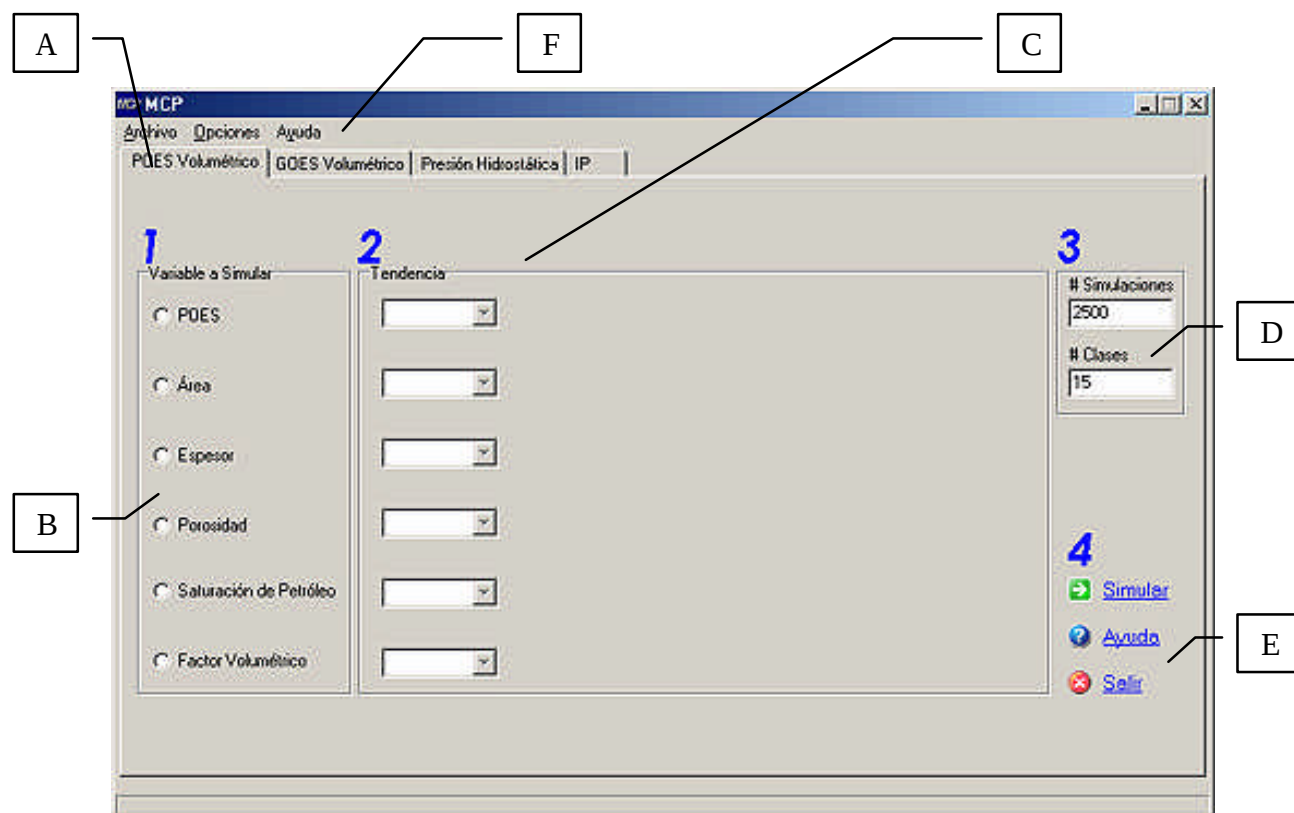


Figura 8 –Pantalla principal de MCP

Sección A: Esta sección corresponde a las pestañas de la parte superior de la ventana. Se utilizan para seleccionar la fórmula con la que se desea trabajar; el MCP fue diseñado con cuatro de ellas que son el cálculo volumétrico del petróleo original en sitio (POES), el cálculo volumétrico del gas original en sitio (GOES), el cálculo de presión hidrostática debida a una columna de fluido y el cálculo de el índice de productividad (IP).

Sección B: En esta sección se escoge cual variable se desea conocer a partir de las demás. Es necesario seleccionar con el ratón una de ellas antes de seguir con el siguiente paso. Se muestra un número uno en azul que indica que es el primer paso a realizar para proceder con la simulación de la fórmula seleccionada.

Sección C: En esta sección se debe escoger la tendencia de las variables. Se entiende por tendencia a la distribución de probabilidad de las variables involucradas en la simulación y los parámetros para definirla, es decir, los límites, promedio, desviación y otros dependiendo del caso. Las Distribuciones posibles son: constante, uniforme, triangular, normal, lognormal y exponencial. Tiene el número dos que indica que es el segundo paso.

Sección D: Aquí se escogen el número de simulaciones y de clases a utilizar. Está Identificada con el número tres.

Sección E: Luego de pasar por cada una de las secciones anteriores en ésta se procede a realizar las simulaciones haciendo clic sobre el texto *Simular*. En esta sección también es posible entrar en la ayuda del programa al hacer clic sobre *ayuda* o cerrar MCP haciendo clic sobre *salir*. Está identificada con el número cuatro.

Sección F: En esta sección se encuentra el menú desplegable del programa donde hay opciones como guardar los datos para simular, cargar esos datos, salir de MCP, acceder la configuración del programa y obtener ayuda.

El proceso de uso de MCP se explicará con un ejemplo en el cual se desea calcular el POES volumétricamente para un yacimiento. Se usaran los siguientes datos:

Área entre 100 y 150 acres. Se escogerá una distribución uniforme.

Espesor con distribución normal, promedio 22 pies, con desviación de 2.

La porosidad triangular, que va de 10 a 21%, con el más probable en 20%.

Saturación de petróleo normal, promedio 65% y desviación 3.

Factor volumétrico constante de 1,1 BY/BF.

Lo primero que se hace es seleccionar en el MCP la pestaña que corresponde al cálculo volumétrico de POES. Luego en la sección B se selecciona con el ratón la opción de POES. En la sección C se escoge junto a *Área* la opción *Uniforme*. Se abren a la derecha 2 espacios. El primero dice *Mínimo* y el segundo *Máximo*, indicando que son para los valores máximo y mínimo de la distribución *Uniforme*. Indican también que los valores que se solicitan deben estar en *Acre*s. Se ingresan 100 y 150 y queda como se muestra en la Figura 10

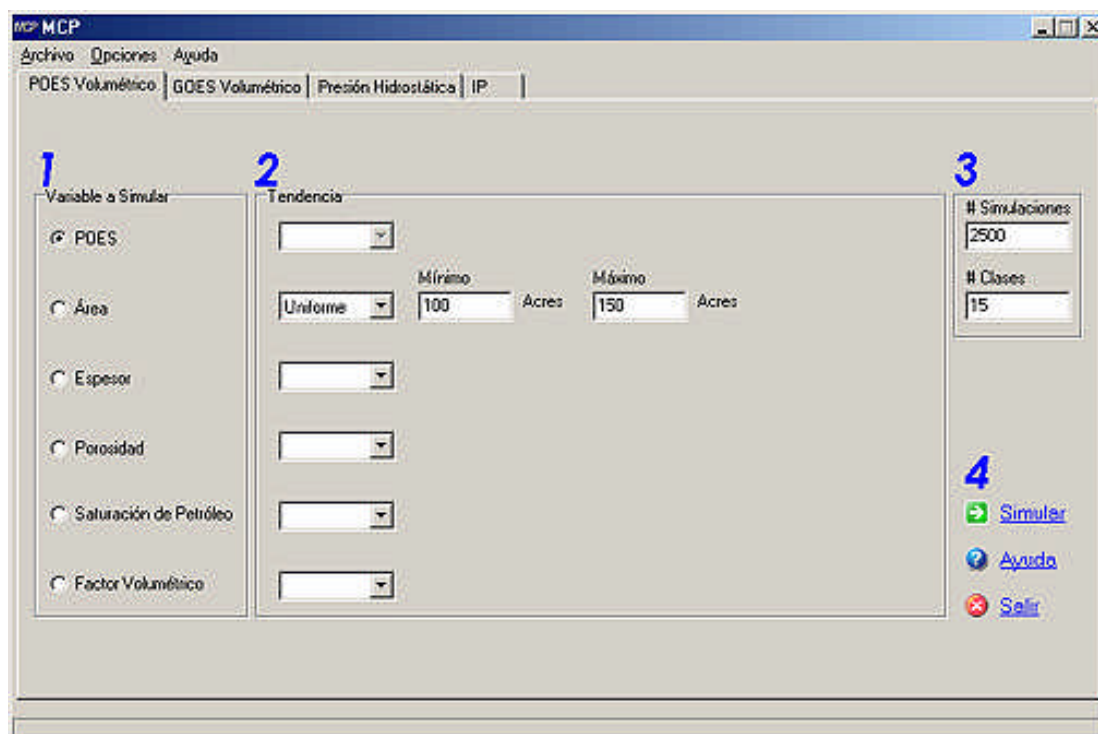


Figura 9 – Introducción de datos para área en MCP

Se repite este mismo proceso para cada una de las variables restantes, *Espesor*, *Porosidad*, *Saturación de Petróleo* y *Factor Volumétrico*. El resultado se muestra en la Figura 11

The screenshot shows the MCP software interface with the following components:

- 1 Variable a Simular:** A list of variables to simulate, including POES (selected), Área, Espesor, Porosidad, Saturación de Petróleo, and Factor Volumétrico.
- 2 Tendencia:** A section for selecting the trend type and its parameters. The selected trend is 'Uniforme'. The parameters are:

Parámetro	Valor	Unidad
Mínimo	100	Acres
Máximo	150	Acres
Media	22	Pies
Desviación	2	
Mínimo	10	%
Más probable	20	%
Máximo	21	%
Media	65	%
Desviación	3	
Constante	1.1	BY/BF
- 3 # Simulaciones:** A field for the number of simulations, set to 2500.
- # Clases:** A field for the number of classes, set to 15.
- 4 Botones:** A set of buttons including 'Simular' (Simulate), 'Ayuda' (Help), and 'Salir' (Exit).

Figura 10 – Introducción de datos en MCP

El siguiente paso es escoger el número de simulaciones y de clases en la sección D. MCP sugiere 2500 simulaciones y 15 clases, y aunque es posible cambiarlo, se van a escoger estos valores. Luego en la sección E se hace clic con el ratón sobre *Simular*. MCP muestra una ventana indicando el progreso de la simulación y cuatro segundos después muestra la ventana que se presenta en la Figura 12

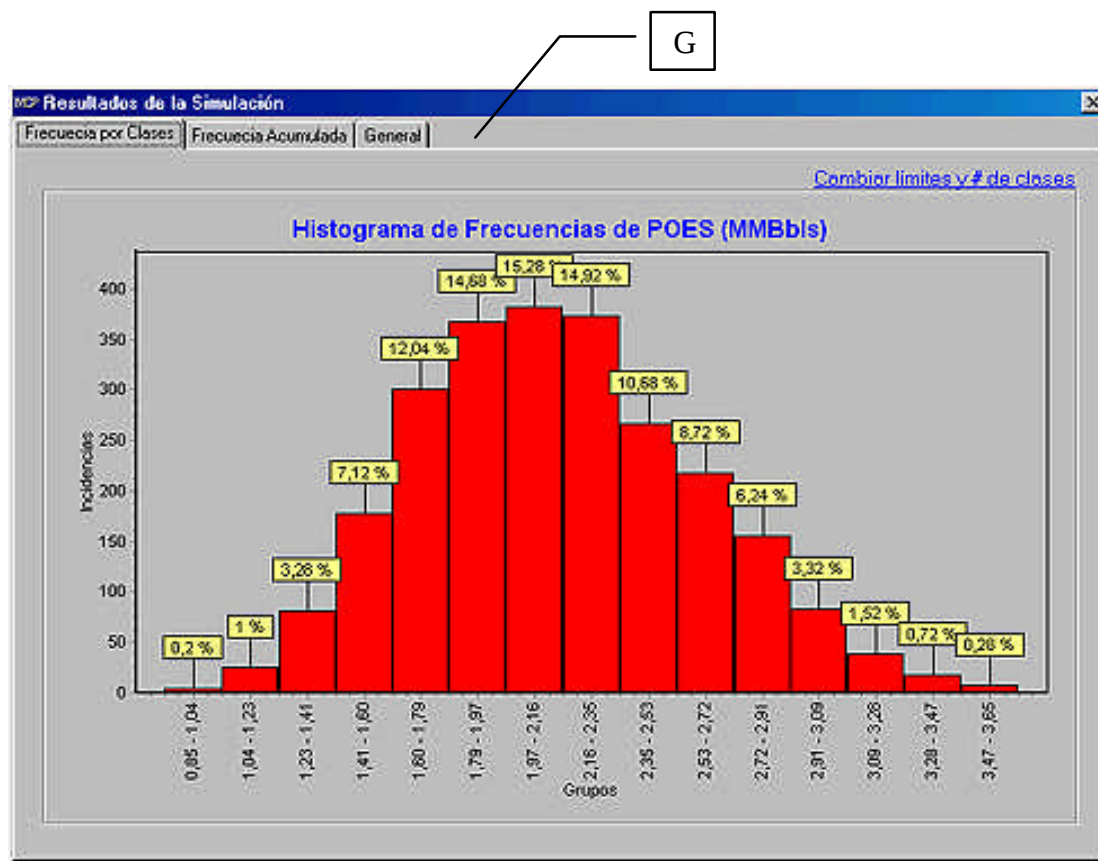


Figura 11 – Histograma de frecuencias de la simulación en MCP

Luego de haber hecho las 2500 simulaciones, recogido los resultados y ordenado los mismos, MCP presenta el *Histograma de Frecuencias* de los mismos en tantas clases como se escogió anteriormente., donde es posible ver como se distribuyen los datos. El eje inferior del gráfico indica los valores límites de cada clase en MMBbls. El eje lateral muestra la frecuencia de dichas clases. Si por ejemplo se desea saber la frecuencia de la clase que está entre 2,07 y 2,26 MMBbls, se tiene gracias al eje lateral que es aproximadamente 400. En la parte superior se puede hacer clic sobre *cambiar límites y # de clases* para modificar la visualización del histograma. La **Sección G** está compuesta por una pestaña donde es posible seleccionar el tipo de resultado que se desea ver. En este instante esta seleccionado

Frecuencia por Clases, donde se ve el histograma de frecuencia. Haciendo clic en *Frecuencia Acumulada* se tiene lo que se muestra en la Figura 13

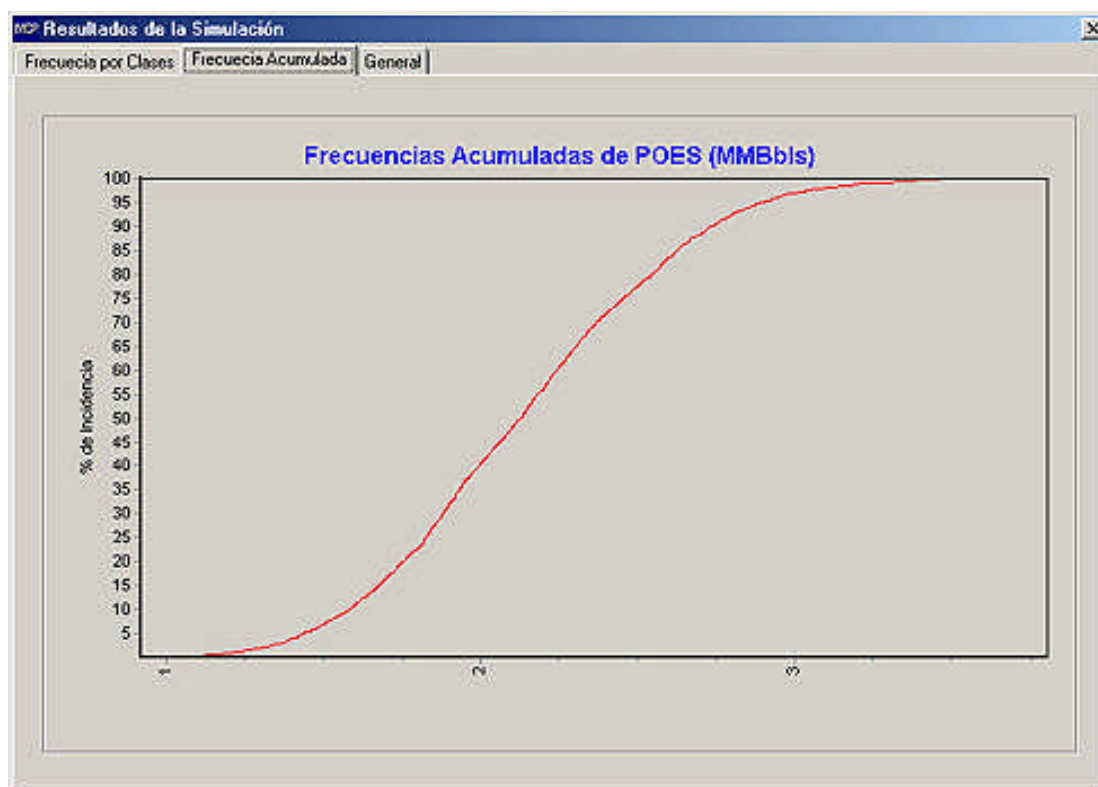


Figura 12 – Frecuencias acumuladas en MCP

Esta curva es la *Frecuencia Acumulada*. Su eje inferior muestra el POES en MMBbls y el eje lateral la incidencia para esa cantidad de POES. Como es acumulada, es posible saber que incidencia de los valores hasta un POES determinado. Supóngase que se desea saber que porcentaje de incidencia representan los valores de POES inferiores o iguales a 2 MMBbls. Lo que se hace es observar que porcentaje de incidencia muestra la curva para 2 MMBbls, que es en este caso 40%.

Al hacer clic sobre la última pestaña, llamada *General*, se muestra la ventana que se ve en la Figura 14

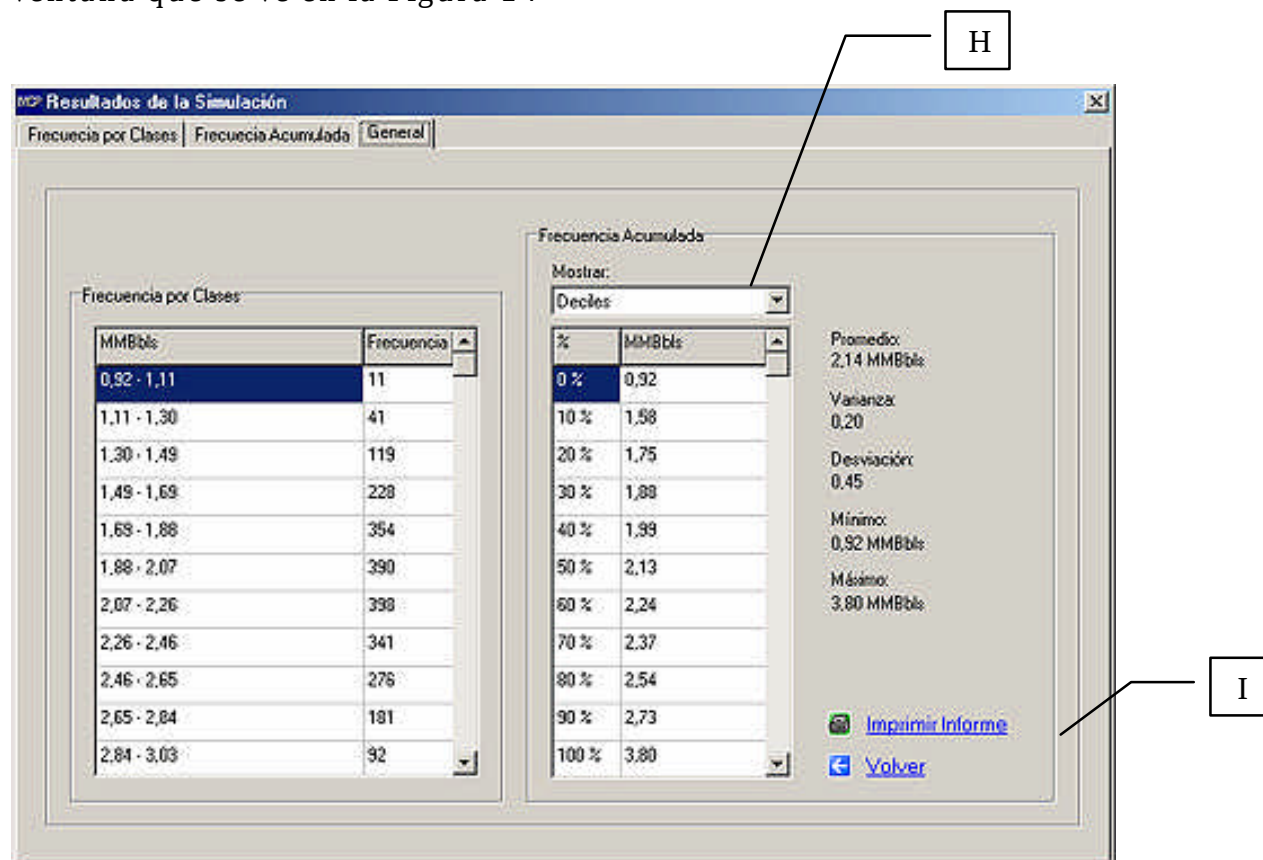


Figura 13 – resultados de la simulación en tablas de MCP

Lo que se tiene en este caso es la información de las gráficas anteriores mostrada en cuadros. A la izquierda de la ventana se tiene un cuadro con el título *Frecuencia por Clases*. Está construido con los valores con los que se realizó el *Histograma de Frecuencias* y muestra claramente los valores que tendrían que aproximarse en éste último. Por ejemplo, en el histograma se quiso saber la frecuencia de la clase que tenía como límites 2,07 y 2,26; se aproximó a 400, pero en este cuadro se puede ver que es exactamente 398.

A la derecha de la ventana se tiene un segundo cuadro con el título *Frecuencia Acumulada*. Este, como el cuadro anterior, muestra claramente los valores pero en este caso, del gráfico de frecuencias acumuladas.

Más a la derecha MCP muestra los siguientes datos:

- *Promedio*: Es la suma algebraica de todos los valores obtenidos dividido entre el número de ellos. En este caso es 2,14 MMBbls. Su definición matemática es la siguiente:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{Ecuación 33}$$

siendo x_i cada uno de los resultados de la simulación y n el número de ellos.

- *Varianza*: es una medida de dispersión de los valores con respecto al promedio. Se define matemáticamente de la siguiente manera:

$$V = \frac{\sum_{i=1}^n (\bar{X} - x_i)^2}{n-1} \quad \text{Ecuación 34}$$

- *Desviación*: es otra medida de dispersión de los valores y se define de la siguiente manera:

$$S = \sqrt{V} \quad \text{Ecuación 35}$$

Es decir, es la raíz cuadrada de la varianza o también:

$$S = \sqrt{\frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n-1}} \quad \text{Ecuación 36}$$

- *Mínimo y Máximo*: son los valores extremos que se obtienen al ordenar los resultados, es decir, el menor y el mayor de todos los valores.

Existe una **Sección H** que en este instante dice *Mostrar: deciles*. Los deciles son los valores que dividen una distribución de frecuencia acumulada en 10 partes de igual cantidad de incidencias. Haciendo clic sobre esto se puede escoger entre *Cuartiles, Deciles y Percentiles*. Los Cuartiles son los valores que dividen la distribución acumulada en cuatro partes de igual incidencia y los percentiles en cien partes.

Haciendo clic y escogiando cuartiles se obtiene la siguiente ventana que se muestra en la Figura 15:

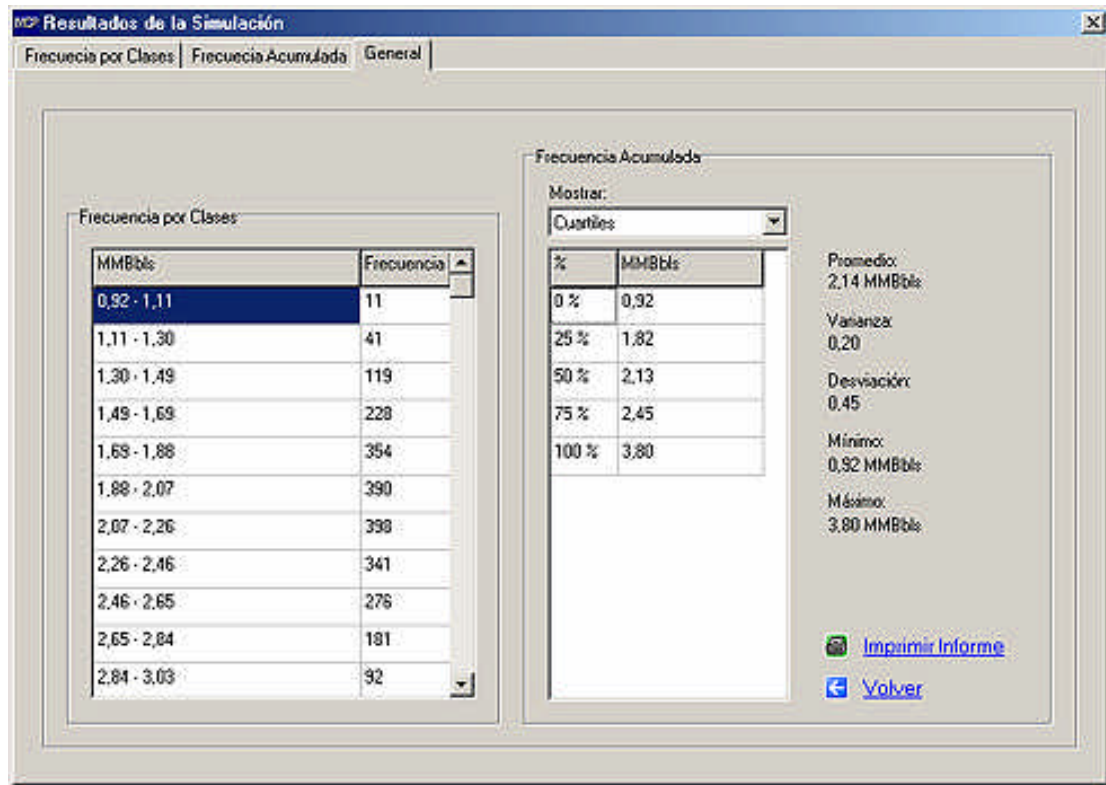


Figura 14 – Cuantiles de las frecuencias acumuladas en MCP

Lo que muestra MCP es un cuadro que se presenta en la Tabla 3

Tabla 3 – Cuantiles mostrados por MCP

%	MMBbIs
0%	0,92
25%	1,82
50%	2,13
75%	2,45
100%	3,80

En la columna de la izquierda están los porcentajes, en la de la derecha el POES en MMBbbs. Para 0% y 100% se tienen los límites inferior y superior de los resultados obtenidos (0,92 y 3,80 MMBbbs respectivamente). Los otros 3 restantes son los cuartiles (25%, 50% y 75% con 1,82, 2,13 y 2,45 MMBbbs)). Son tres valores porque dividen como se dijo antes en 4 partes con igual cantidad de incidencias cada una.

La sección que se indica como **Sección I** muestra dos opciones que se pueden escoger en este instante; la superior llamada *Imprimir Informe* que envía a la impresora un resumen de la simulación realizada, luego de solicitar ciertos datos referentes a la impresora y al opción inferior o *Volver*, que permite regresar a la ventana principal y realizar cambios y simular de nuevo o salir de MCP.

Cómo se analizan los resultados de MCP

Una manera muy sencilla y rápida de obtener un resultado con MCP es a través del promedio que entrega. En el ejemplo anterior es posible decir que la tendencia central de POES es de 2,14 MMBbbs. Si se desea incluir información sobre cuan preciso puede ser éste resultado, se puede recurrir a la varianza, ya que esta es una medida de la dispersión de los resultados y mientras menor sea este valor más cercanos están los resultados obtenidos al promedio. También es posible reportar la desviación, que es directamente proporcional a la varianza.

Otra manera de entregar un resultado es a través de los valores mínimo y máximo, indicando en el ejemplo que el POES puede estar entre 0,92 y 3,80 MMBbbs para este caso. Sin embargo sería muy útil incluir además la información del promedio y la dispersión.

Una tercera manera puede concretarse con un análisis un poco más profundo; supóngase que la simulación que se hizo corresponde a un estudio económico para explotar un yacimiento de petróleo. Luego de proyectar el campo, se sumaron los valores cronológicos de cada uno de los gastos necesarios. Teniendo un estimado del precio al cual se puede vender el petróleo de ese yacimiento, de el factor de recobro esperado y otros factores, se llega a la conclusión de que es necesario que existan cuando menos 1,8 MMBbbs. Luego, ya sea en la grafica de frecuencia acumulada o en el cuadro, se lee la incidencia de 1,8 MMBbbs que es 23%. Lo que se puede concluir de esto es que hay un 23% de probabilidades de que haya menos de 1,8 MMBbbs de petróleo original en sitio o igual cantidad. Recíprocamente el 77% restante es que haya más de esa cantidad en barriles. Una excelente conclusión al estudio sería: “hay 77% de probabilidades de éxito según la simulación de Monte Carlo”.

Es posible utilizar éste método no para valores fijos solamente, sino también para intervalos, es decir, se puede preguntar: ¿Cual es la probabilidad de que el valor del POES esté entre dos y tres MMBbbs? Simplemente se obtiene de la tabla de frecuencia acumulada la incidencia para dos MMBbbs que es 39% y para tres MMBbbs, que es 96%. Como lo que se pregunta es la posibilidad de que el POES esté entre los dos valores mencionados, se restan las incidencias y se

puede decir que: “existe un 57% de probabilidades de que el valor del POES esté entre dos y tres MMBbls”

Simulación de Monte Carlo utilizando una hoja de cálculo

MCP es una herramienta diseñada para la aplicación de Monte Carlo con cuatro ecuaciones de uso muy común en Ingeniería de Petróleo. Existen múltiples razones por las que fueron solo estas cuatro las escogidas, entre ellas tenemos las siguientes:

- Se vio anteriormente que Monte Carlo nació a razón de que o fuera muy difícil obtener un valor único para una variable envuelta en una ecuación o fuera muy costoso; esta razón sigue siendo válida y se escogen estas cuatro ecuaciones a causa de que poseen involucradas variables que por lo general se miden indirectamente (como el área que se mide con un método sísmico o con perforación de muchos pozos) o a causa de que incluyen variables que varían en la zona de estudio y es muy difícil obtener un valor promedio acertado (como la densidad de un fluido que cambia por razones de homogeneidad, temperatura, presión y otras).
- Se deseaba obtener un programa sencillo que fuera fácil de instalar, operar y que ocupara pocos recursos del sistema en un computador. Todo esto con la finalidad de hacerlo amigable al usuario, y poner muchas ecuaciones implicaba ir en contra de lo deseado.

En el caso de ser necesario trabajar con una ecuación distinta a las que incluye MCP, se puede utilizar una hoja de cálculo. El proceso es más lento y no tan directo como MCP, pero permite incluir cualquier ecuación que se necesite.

Existen muchas hojas de cálculo actualmente en el mercado, comercializadas por distintas casas de software. A continuación se explicara como realizar una simulación de Monte Carlo con la hoja de cálculo *Excel 2002*, de la casa de software *Microsoft*.

1. Primero que nada es necesario cargar el archivo *SIM5032.DLL* en el sistema, esto se hace copiando el archivo con ese nombre que se encuentra en el directorio XLSIM del CD de MCP al directorio *Windows\System* del computador.
2. Luego se necesita cargar una plantilla que genera números aleatorios, para esto haga doble clic sobre le archivo *SIM.XLA* que se encuentra en el mismo CD. Esto hará que se inicialice la hoja de cálculo.
3. Si su hoja de cálculo posee un nivel de seguridad medio, solicitará su aprobación para cargar la plantilla, para esto se hace clic sobre le botón *Habilitar macros*. Si el nivel de seguridad es alto se le notificará que los macros han sido deshabilitados, si ese es el caso haga clic sobre el botón aceptar y luego en el menú *Herramientas*, en el sub menú *Macro*, en la opción *Seguridad*, escoja el nivel se seguridad medio, cierre *Excel* y repita desde el paso 2.

4. Ahora para verificar que la herramienta para generar el histograma está instalada, haga clic en el menú *Herramientas* y luego en *Análisis de datos*, debe aparecer la opción *Histograma*. Si no existe la opción *Análisis de datos* haga clic sobre *Complementos* y luego sobre *Herramientas para el análisis*, esto debe activar *Análisis de datos*. Si no existe esa opción la causa es que no fue instalado en el equipo cuando se instaló el *Excel*, por lo que se hace necesario utilizar el disco de instalación original.
5. Abra una hoja de cálculo nueva. Estando en la casilla de la primera fila y primera columna escriba el signo igual (=) y luego haga clic en el cuadro de *funciones*, en la parte superior de la hoja. Escoja *Más funciones* y navegue hasta la función *gen_binomial*. De allí en adelante existen 12 funciones que generan números aleatorios. Para explicarlo con un ejemplo escoja *gen_normal*. Esto generará números aleatorios normalmente distribuidos. Al hacer clic en *Aceptar* el programa pide que se introduzca un valor llamado *Mean* que es la media y *StDev* que es la desviación estándar deseada. Introduzca 18 y dos respectivamente y aparecerá un número aleatorio en esa casilla. Copie la casilla en las casillas inferiores desde la A2 hasta la A30. Aparecerán 30 números aleatorios distintos distribuidos normalmente y con media 18 y desviación estándar dos. En la columna siguiente, en la casilla B1 escoja *gen_uniform* (que genera números aleatorios con distribución uniforme) e introduzca los valores tres y nueve como límites inferior y superior para generar números aleatorios con

distribución uniforme. Copie de nuevo, desde la casilla *b2* hasta la casilla *B30*.

6. En la columna *C*, en la casilla *C1*, irá la ecuación con la que se desea simular. Por ser este un ejemplo se utilizará una ecuación muy sencilla la cual es:

$$C = A * B$$

Ecuación 37

Por lo que en la casilla *C1* irá la ecuación:

$$= A1 * B1$$

Ecuación 38

Copie nuevamente, hasta la casilla *C30*. Aparecerán los 30 resultados de la simulación.

7. Ahora se procederá a definir los límites del histograma y su número de clases. En la simulación que se acaba de realizar, se observa que el menor resultado de *C* es 51,57 (que variará cada vez que se realice porque se generan otros números aleatorios) y debido a esto se tomará el límite inferior como 50. Para simulaciones de muchos valores se puede utilizar una función de *Excel* llamada *Min* que devuelve el mínimo valor de un grupo. Si se desea un número definido de clases se puede utilizar la función *Max* (que devuelve el máximo), y calcular el ancho de cada clase como:

$$\text{Ancho de clase} = (\text{Máximo} - \text{Mínimo}) / \text{numero de clases}$$

Para hacer más simple este proceso se escogerá un ancho de clase de 10 y el número de clases vendrá determinado por el mayor valor. En la columna *D*, en la casilla *D1* se introduce el límite superior de la primera clase que es 60. En la siguiente casilla (*D2*) el límite inferior de la siguiente clase que sería 70. Como el mayor es 155,55 la última casilla será la que tenga como valor 150, porque la siguiente clase agrupa todos los valores mayores de 150.

8. Ahora en el menú *Herramientas*, luego en *Análisis de Datos*, haga clic en *histograma*. En la ventana que se presenta, en rango de entrada introduzca los valores que alimentaran el histograma, es decir la columna *C* (*\$C\$1:\$C\$30*). En rango de clase los mínimos de cada clase, es decir, la columna *D* (*\$D\$1:\$D\$10*). Active las opciones *Porcentaje acumulado* y *Crear gráfico* y haga clic
9. *Excel* muestra en una hoja nueva la distribución de frecuencias de la simulación y el histograma de frecuencias con la curva de frecuencias acumuladas. Esto se puede hacer con cualquier fórmula que se necesite con sólo hacer los cambios necesarios.

Conclusión sobre éste método

Utilizar una hoja de cálculo para realizar una simulación de Monte Carlo resuelve efectivamente el problema de utilizar la ecuación que se desee, además de ser una herramienta de uso común para los ingenieros. Esto es gracias a que no funciona como programa compilado el cual necesita las ecuaciones antes de generar un archivo ejecutable sino que interpreta las operaciones que se indican con cada casilla. Sin embargo, la ventaja que se logra con esto hace al *Excel* débil en el sentido de que se necesitan muchos pasos y chequeos antes de obtener una simulación exitosa.

El mismo ejemplo se trató de realizar con 3000 simulaciones y el resultado se obtuvo luego de 22 segundos. Una simulación equivalente se obtiene en MCP en sólo tres segundos, y todo esto sin incluir los tiempos necesarios para preparar la hoja de cálculo de manera adecuada.

Todo esto indica que la simulación de Monte Carlo utilizando una hoja de cálculo debe realizarse cuando no se cuenta con una herramienta diseñada específicamente para ello o dicha herramienta no cuente con las ecuaciones que se desean simular.

IV CONCLUSIONES Y RECOMENDACIONES

Conclusiones

El estudio realizado permite establecer las siguientes conclusiones:

1. El método de Monte Carlo permite obtener soluciones probabilísticas para ecuaciones donde no es satisfactorio utilizar un método determinístico por falta de precisión en los datos de entrada.
2. Es muy recomendable hacer uso de un computador para aplicar el método de Monte Carlo. Éste puede realizar las cientos o miles de simulaciones necesarias en pocos segundos y con la cantidad de cifras decimales necesarias.
3. Es muy importante escoger el software indicado para realizar la simulación. Los software de amplio uso como *Microsoft Excel* permiten realizar simulación de Monte Carlo con muchas ecuaciones distintas, sin embargo la preparación previa es un proceso engorroso y simular muchas veces es lento. Por otro lado los software de uso exclusivo como *MCP* que se diseñó para este Trabajo tienen definidas las ecuaciones con las que trabajan, sin embargo son rápidos en la preparación previa y en las simulaciones. Existen también paquetes de software de nivel intermedio, que toman ventajas de los dos anteriores, aunque

en algunos casos es necesaria la capacitación previa del usuario.

4. El análisis de resultados puede ser tan preciso como el usuario lo desee, desde simplemente tomar la media o los límites, hasta escoger un resultado con probabilidades de ocurrencia de distintos rangos de la variable simulada.

Recomendaciones

Con base en los resultados de éste estudio y sus conclusiones, surgen las siguientes recomendaciones:

1. Considerar el uso de una simulación de Monte Carlo siempre que no se cuente con información precisa de las variables de una ecuación para obtener una mejor apreciación de los posibles resultados.
2. Escoger el software que se utilizará para realizar la simulación considerando la ecuación a utilizar, el número de simulaciones que se desean y el tiempo con el que se dispone.
3. Escoger el número de simulaciones de acuerdo a los datos de entrada, mientras más variables tenga la ecuación y mientras mas dispersas estén las mismas, más simulaciones serán necesarias.
4. No abusar del uso de Monte Carlo. Debe tenerse siempre en cuenta que esta es una herramienta probabilística y los análisis que se hagan con ella deben reflejar que se trata de una probabilidad.
5. Enriquecer el acervo de la Escuela en este campo, instrumentando aplicaciones adicionales del método de Monte Carlo en Ingeniería de Petróleo.

Referencias

Abelson, Robert P. (1998). *La estadística razonada, reglas y principios*. Ediciones Paidós, Barcelona, España. Primera edición. pp. 19 – 21, 39 – 60, 109 – 136, 199 – 212.

American Institute of Physics (s. f.). *Quantum mechanics. The uncertainty principle*. [Documento en línea]. Disponible: <http://www.aip.org/history/heisenberg/p08.htm> [Consulta: 2003, Mayo 15]

Analycorp Consulting (s. f.). *Xlsim*. [Documento en línea]. Disponible: <http://www.analycorp.com/dl-xlsim.htm> [Consulta: 2003, Abril 27]

Craft, B. C., Hawkins, M. F. (1959). *Applied Petroleum Reservoir Engineering*. Prentice Hall, Usa. pp. 26 – 35, 105 – 114.

Coss, Raúl (1999). *Simulación, un enfoque práctico*. Editorial Limusa, México. pp.11, 12, 17, 18, 31 – 59.

Haby, Jeff (s. f.) *Hydrostatic Equation*. [Documento en línea]. Disponible: <http://www.theweatherprediction.com/habyhints/62/> [Consulta: 2003, Junio 04]

Enciclopedia Microsoft® Encarta® 98 © 1993-1997 Microsoft Corporation

Kalos, Malvin (1986). *Monte Carlo methods*. John Wiley & Sons, New York

McLaughlin, Michael P. (1999). *A Compendium of Common Probability Distributions*. [Documento en línea]. Disponible:

http://www.causascientia.org/math_stat/Dists/Compendium.pdf

[Consulta 2003, Mayo 28]

Murtha, J. A. (1997, Abril). *Monte Carlo simulation: its status and future*. Journal of Petroleum Technology. pp. 361 – 373

Nawrocki, David (2003). *The problems with Monte Carlo simulation*.

[Documento en línea]. Journal of financial planning. Disponible:

http://www.fpanet.org/journal/articles/2001_Issues/jfp1101-art12.cfm

[Consulta: 2003, Abril 27]

Naylor, Thomas H. (1977). *Técnicas de simulación en computadoras*.

Editorial Limusa, México. pp.15 – 141.

Nind, T. (1964) *Principles of oil well production*. McGraw Hill. pp. 56, 57

Parra, Juan Jones (1994). *Elementos de Ingeniería de Yacimientos*.

Editorial Innovación Tecnológica, Universidad Central de Venezuela.

pp. 4.4 – 4.5

Pllana, Sabri (s. f.). *History of Monte Carlo method*. [Documento en

línea]. Disponible:

<http://stud4.tuwien.ac.at/~e9527412/history.html> [Consulta: 2003, Abril 24]

Scheaffer, Richard (1990). *Probabilidad y estadística para Ingeniería*. Grupo Editorial Iberoamérica, México. pp. 2 – 5, 66 – 113, 124 – 171.

Soto Negrín, Armando (1990). *Iniciación a la estadística*. Expediente editorial José Martí, Caracas. pp. 35 – 160, 211 – 248

Spiegel, Murray (1998). *Estadística*. McGraw Hill, Caracas. pp. 1 - 59, 129 – 185.

Stockton, J. R. (2002, Mayo 02). *Borland Pascal random page*. [Documento en línea]. Disponible:
<http://www.merlyn.demon.co.uk/pas-rand.htm> [Consulta: 2003, Abril 18]

The Rig Site (2002) *Hydostatic Pressure*. [Documento en línea]. Disponible:
<http://www.workover.co.uk/well%20control1/Well%20control%20online%20training%20manual%20hydrostatic%20pressure.htm>
[Consulta: 2003, Junio 03]

Torres de Forzanti, Liana (s. f.). *Metodología para la Elaboración del Plan del Trabajo Final de Grado en Ingeniería*. [Documento en línea]. Disponible:
<http://www.asee.org/international/intertech2002/840.pdf> [Consulta: 2003, Mayo 03]

University of Sydney, School of Information Technologies (s. f.).
Generating random variables. [Documento en línea]. Disponible:
<http://www.cs.usyd.edu.au/~deveritt/networksimulation/rv.pdf>
[Consulta: 2003, Abril 18]

Apéndices

Determinación del número adecuado de simulaciones para obtener resultados satisfactorios con Monte Carlo

El número de simulaciones que se realizan cuando se aplica el método de Monte Carlo puede ser cambiado a gusto del usuario, ya sea con una hoja de cálculo o con el MCP. Un número bajo de simulaciones hace el proceso sumamente rápido pero entrega resultados no muy convincentes, es decir, la curva de frecuencia acumulada dista de ser suave y si se realiza nuevamente la misma simulación los valores del promedio, la varianza, la desviación estándar, el máximo y el mínimo varían con respecto a los resultados anteriores, debiendo ser iguales. Esto se resuelve paulatinamente a medida que se incrementa el número de simulaciones, sin embargo también se incrementa el tiempo de ejecución y es incómodo pasar mucho tiempo frente al computador esperando un resultado.

Este problema es mas complicado de lo que se desearía, dado que son muchos los factores que influyen en él, como lo son la ecuación con la que se trabaja, que tipo de distribución de frecuencia tiene cada variable, que dispersión tiene cada variable (que puede definirse por los límites, la varianza u otro), el número de variables de la ecuación y la operación que se realiza con cada una.

Evidentemente es necesario un estudio exhaustivo para aclarar todas las dudas, pero dado que ese no es el objetivo de éste trabajo y no se cuenta con las herramientas necesarias para realizar ese estudio en este momento, se propone a continuación una técnica experimental que pretende dar una idea del número adecuado de simulaciones a utilizar.

Se estudiarán los siguientes casos:

- Si el número de variables de la ecuación influye y como lo hace.
- Si la dispersión de los datos de entrada influye y como lo hace.

Para el primer caso se tomará una ecuación sencilla, del tipo:

$$a = b * c * d * e$$

Ecuación 39

Se simulará tomando los valores de b , c , d y e con distribución normal, de valor promedio uno y desviación de 0,1. Se hará tres veces, la primera considerando que la ecuación sólo posee las variables a , b y c , la segunda incluyendo la variable d y la tercera con la variable e incluida. En cada caso se usaran distintos números de simulaciones y se anotarán los resultados de la media con los cuales se hará el análisis.

A continuación se presenta la Tabla 4, que contiene las medias de las simulaciones realizadas:

Tabla 4 – Resultados de las simulaciones con varias variables

Número de Simulaciones	3 variables	4 variables	5 variables
100	1,0055	1,0078	1,0014
300	0,9982	1,0043	0,9988
500	0,9967	0,9962	1,0067
700	0,9971	0,9977	0,9928
900	0,9982	0,9984	1,0090
1100	1,0040	0,9972	0,9957
1300	0,9984	0,9967	1,0032
1500	0,9988	1,0066	1,0044
1700	0,9984	1,0049	1,0030
1900	0,9971	1,0024	0,9934
2100	1,0009	1,0019	1,0033
2300	1,0037	0,9969	1,0015
2500	1,0035	0,9996	0,9980
2700	1,0019	1,0035	0,9984
2900	0,9990	0,9963	1,0033
3100	1,0032	0,9990	0,9954
3300	0,9991	0,9977	1,0012
3500	1,0007	1,0019	0,9997
3700	0,9989	1,0027	0,9995
3900	1,0031	0,9985	0,9995
4100	0,9982	1,0002	1,0025
4300	1,0025	1,0036	1,0040
4500	1,0022	1,0038	0,9969
4700	1,0019	0,9996	1,0007
4900	0,9991	0,9970	1,0016

5100	1,0002	1,0015	1,0018
5300	0,9998	1,0017	1,0035
5500	0,9991	0,9988	0,9992
5700	0,9997	0,9980	1,0039
5900	1,0016	1,0022	1,0011
6100	0,9993	0,9999	0,9999
6300	0,9989	0,9991	0,9994
6500	1,0012	0,9999	0,9996
6700	0,9995	1,0022	0,9978
6900	1,0000	0,9998	1,0010
7100	1,0002	1,0003	0,9996
7300	0,9998	0,9994	0,9997
7500	1,0000	0,9993	1,0004
7700	0,9994	0,9998	0,9993
7900	0,9998	0,9993	0,9997
8100	0,9994	0,9995	0,9996

Con estos datos se construyó el gráfico que se muestra a continuación:

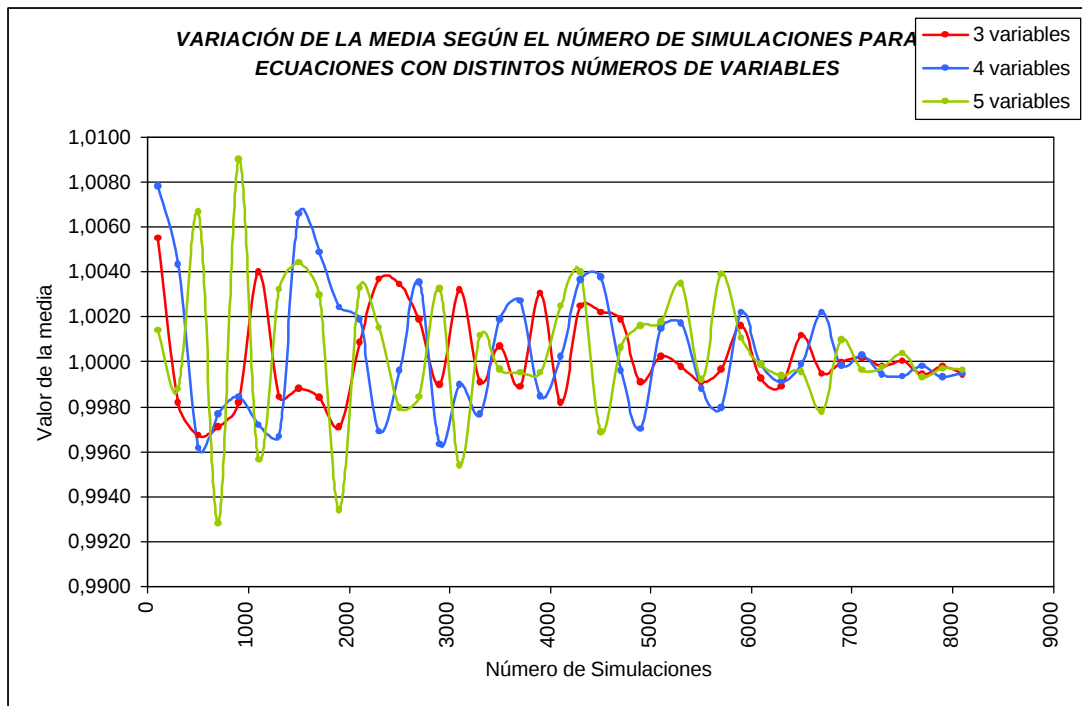


Figura 15 – Variación de la media para las distintas simulaciones con diversos números de variables

Se puede observar lo siguiente:

1. Las medias para la ecuación de tres variables ($a = b * c$) tiende a tener valores mas distantes de uno (lo que debería ser el resultado de la media) que para la de cuatro variables ($a = b * c * d$) y esta a su vez menor que para la de 5 variables ($a = b * c * d * e$) para cualquier número de simulaciones.
2. Observando la dispersión de los valores de la curva, se podrían definir 3 áreas dentro del cuadro, una que va de cero a 2500 simulaciones, una de 2500 a 7000 simulaciones, y la última para más de 7000 simulaciones. Haciendo un análisis de la desviación estándar de estos tres grupos se tiene lo siguiente:

Tabla 5 – Desviaciones estándar para diversos grupos

Grupos	Desviación estándar 3 variables	Desviación estándar 4 variables	Desviación estándar 5 variables
De 0 a 2500	0,0030	0,0042	0,0051
De 2500 a 7000	0,0016	0,0022	0,0023
7000 y más	0,0003	0,0004	0,0004

con este cuadro se puede asegurar que de cero a 2500 simulaciones los resultados se dispersan mucho, mientras que de 2500 a 7000 simulaciones la dispersión baja, haciéndose muy baja para más de 7000 simulaciones.

En segundo caso se estudiará como influye la dispersión en los datos de entrada de la simulación. Se utilizará de nuevo la ecuación:

$$a = b * c * d * e$$

Ecuación 40

Se simulará tomando los valores de *b*, *c*, *d* y *e* con distribución normal, de valor promedio uno pero con tres desviaciones distintas que serán 0,05, 0,1 y 0,15. En cada caso se usaran distintos números de simulaciones y se anotarán los resultados de la media con los cuales se hará el análisis.

A continuación se presenta la tabla que contiene las medias de las simulaciones realizadas:

Tabla 6 – Resultados de las simulaciones con diversas desviaciones estándar

Número de Simulaciones	<i>Desviación 0,05</i>	<i>Desviación 0,1</i>	<i>Desviación 0,15</i>
100	0,9948	1,0024	1,0097
300	1,0034	0,9892	0,9970
500	0,9978	1,0058	0,9869
700	0,9969	0,9936	1,0082
900	0,9975	1,0042	1,0061
1100	0,9998	1,0003	0,9933
1300	1,0030	1,0013	0,9907
1500	0,9978	0,9930	0,9946
1700	0,9963	0,9957	1,0130
1900	0,9988	1,0048	1,0082
2100	0,9974	0,9940	0,9925
2300	1,0000	1,0001	1,0026
2500	1,0020	1,0008	0,9978
2700	1,0001	0,9988	0,9912
2900	0,9995	1,0067	0,9963
3100	0,9977	0,9938	0,9988
3300	1,0024	1,0032	1,0014
3500	0,9989	1,0017	0,9968
3700	0,9996	0,9989	1,0067
3900	0,9983	0,9976	1,0085
4100	0,9992	1,0035	0,9968
4300	1,0012	0,9987	0,9952
4500	0,9998	0,9950	0,9972
4700	1,0009	0,9962	1,0072

4900	0,9989	0,9992	1,0071
5100	0,9982	0,9977	0,9948
5300	0,9998	1,0020	0,9986
5500	0,9994	0,9999	1,0040
5700	0,9999	1,0034	1,0043
5900	1,0009	1,0003	0,9942
6100	1,0005	0,9988	0,9995
6300	1,0007	0,9978	0,9983
6500	0,9992	1,0013	1,0041
6700	0,9996	1,0012	0,9992
6900	0,9998	1,0001	0,9972
7100	0,9997	0,9989	1,0024
7300	1,0002	0,9993	1,0012
7500	0,9996	0,9989	1,0057
7700	1,0001	1,0012	0,9982
7900	0,9999	0,9988	0,9979
8100	0,9998	1,0003	1,0016

Y con estos datos se construye de nuevo un gráfico que es el siguiente:

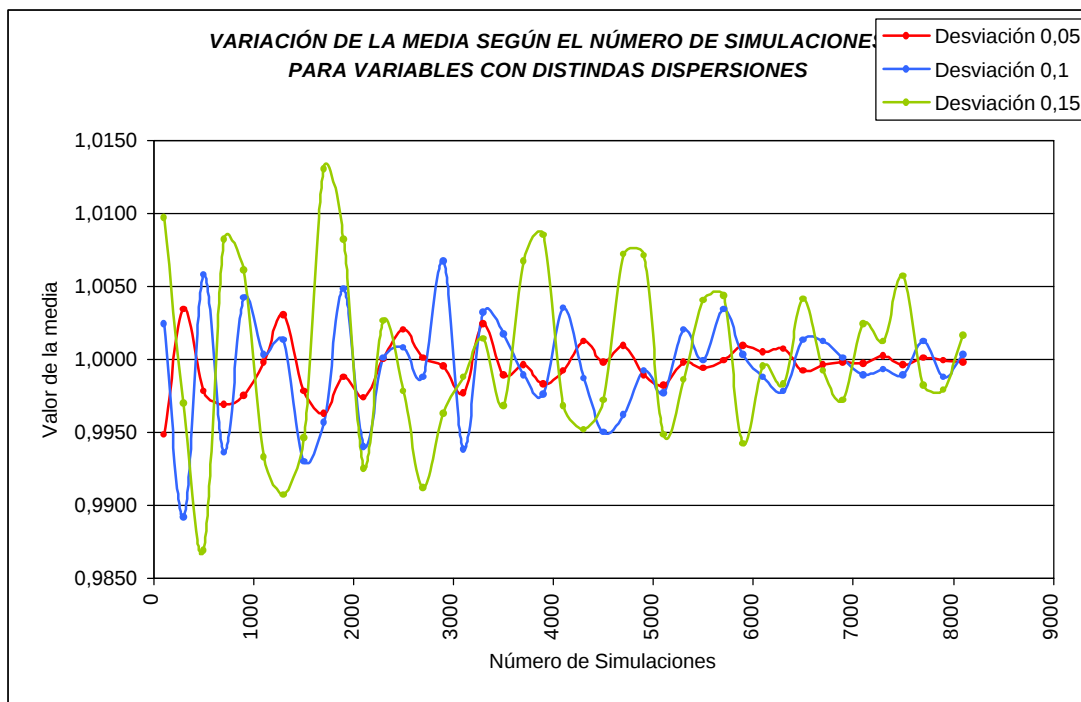


Figura 16 – Variación de la media para las distintas simulaciones con diversas desviaciones estándar

Lo que se hace a continuación es dividir los datos en las tres áreas que se usaron en el caso anterior y verificar la dispersión de los mismos con la desviación estándar.

Tabla 7 – Desviaciones estándar para diversos grupos

Grupos	Desviación estándar para datos con desviación 0,05	Desviación estándar para datos con desviación 0,1	Desviación estándar para datos con desviación 0,15
De 0 a 2500	0,0030	0,0042	0,0051
De 2500 a 7000	0,0016	0,0022	0,0023
7000 y más	0,0003	0,0004	0,0004

Lo que se puede concluir en este caso es que mientras mas grande es la dispersión de los datos de entrada, más dispersa será la salida de la simulación y serán necesarias más simulaciones.