

TRABAJO ESPECIAL DE GRADO

ANÁLISIS DE ERRORES DE PRESIÓN Y PVT SOBRE LAS ESTIMACIONES DE BALANCE DE MATERIALES

Presentado ante la Ilustre
Universidad Central de Venezuela
Por el Br. Carlos A. Garcia Z.
Para optar al título
de Ingeniero de Petróleo.

Caracas, Junio 2005

TRABAJO ESPECIAL DE GRADO

ANÁLISIS DE ERRORES DE PRESIÓN Y PVT SOBRE LAS ESTIMACIONES DE BALANCE DE MATERIALES

TUTOR ACADÉMICO: Prof. José R. Villa

Presentado ante la Ilustre
Universidad Central de Venezuela
Por el Br. Carlos A. Garcia Z.
Para optar al título
de Ingeniero de Petróleo.

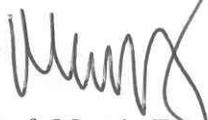
Caracas, Junio 2005

Caracas, Junio de 2005

Los abajo firmantes, miembros del Jurado designado por el Consejo de Escuela de Ingeniería de Petróleo, para evaluar el Trabajo Especial de Grado presentado por el Bachiller Carlos A. Garcia Z., titulado:

“Análisis de errores de presión y PVT sobre las estimaciones de balance de materiales”

Consideran que el mismo cumple con los requisitos exigidos por el plan de estudios conducente al Título de Ingeniero de Petróleo, y sin que ello signifique que se hacen solidarios con las ideas expuestas por el autor, lo declaran APROBADO.



Prof. Martín Essinfeld
Jurado



Prof. Walter Poquioma
Jurado



Prof. José R. Villa
Tutor



Escuela de Ingeniería de Petróleo
Facultad de Ingeniería-UCV

ACTA

Quienes suscriben, integrantes de la totalidad del jurado examinador del Trabajo Especial de Grado, titulado “**Análisis de errores de Presión y PVT sobre las Estimaciones de Balance de Materiales**”, presentado por el Br. Carlos García, titular de la Cédula de Identidad No.15.865.945, para optar al título de Ingeniero de Petróleo, acordaron por unanimidad solicitarle a las autoridades correspondientes de la Facultad de Ingeniería, tengan a bien otorgar **MENCIÓN HONORIFICA** al Trabajo de Grado antes mencionado.

Esta solicitud se hace con base a la originalidad del enfoque utilizado para tratar el tema presentado y a la excelencia, tanto de su contenido como de su presentación formal, atributos que lo convierten en una fuente de referencia útil.

Se levanta la presente Acta en Caracas, a los diez (10) días del mes de junio del año dos mil cinco.

Prof. Martín Essinfeld
Jurado Principal

Prof. Walter Poquioma
Jurado Principal

Prof. José Reinaldo Villa
Tutor Académico

A Dios y mi madre.

Agradecimientos

Al prof. José R. Villa por su guía, sugerencias, y dedicación durante esta investigación.

A mis hermanos Adrián y Carla por su amor y paciencia.

A mis abuelos Vicente y Ana, aunque ya no estén físicamente conmigo, los tengo en mis pensamientos.

A mi madre Coromoto por ser la fuente de inspiración y perseverancia a lo largo de mi carrera.

Carlos Garcia Zurita

García Z. Carlos A.

ANÁLISIS DE ERRORES DE PRESIÓN Y PVT SOBRE LAS ESTIMACIONES DE BALANCE DE MATERIALES

Tutor académico: Prof. José R. Villa. Tesis. Caracas, U.C.V. Facultad de Ingeniería. Escuela de Petróleo. 2005, 359p.

Palabras Claves: Balance de materiales, incertidumbre, presión, PVT.

Resumen. Es bien conocido que los cálculos del POES por medio del balance de materiales están influenciados por la incertidumbre en los datos, y generalmente sus efectos no son estimados. Varias investigaciones han sido publicadas acerca de este tema, sin embargo en lo referente a los datos PVT y los mecanismos de empuje, existen pocas que examinen su impacto en la ecuación de balance de materiales. En este trabajo se cuantifican los efectos de la incertidumbre en los datos de presión y PVT sobre las estimaciones de balance de materiales bajo diversos mecanismos de empuje.

Para evaluar los efectos de la incertidumbre en los datos de presión y PVT se diseñó e implementó un programa que permitiera realizar rápida y eficazmente el análisis estadístico de tales efectos. Esto consistió en la aplicación de dos metodologías diferentes. La primera consistió en generar realizaciones para presión y PVT, y la segunda consistió en implementar la técnica del diseño experimental junto con la técnica estadística de análisis de varianza (ANOVA) para evaluar de manera integrada la influencia de todos los parámetros en forma óptima y a su vez cuantificar su impacto por separado.

Los resultados obtenidos indicaron que las estimaciones de los hidrocarburos en sitio son sensibles al método de balance materiales utilizado, a los errores en los datos de presión y PVT, y a los mecanismos de empuje activos. Para el caso en que se tenga un yacimiento con expansión del petróleo más gas disuelto, la incertidumbre en los datos de presión y PVT es mínima, y además se utilice el método $(F - W_e)$ vs. E_t , se obtiene la mayor exactitud en las estimaciones de balance de materiales. Por el contrario, para en que el principal mecanismo de empuje sea el influjo de agua, exista gran incertidumbre en los datos de presión, y se utilice el método $(F - W_e)/(E_o + E_{fw})$ vs. $(E_g + E_{fw})/(E_o + E_{fw})$ ó el método (F/E_t) vs. (W_e/E_t) , se obtiene la menor exactitud en las estimaciones de balance de materiales.

Índice general

Agradecimientos	II
Introducción	1
Objetivos	3
1. La Ecuación de Balance de Materiales	4
1.1. Principales trabajos relacionados con la Ecuación de Balance de Materiales	4
1.2. Derivación de la Ecuación de Balance de Materiales	8
1.3. Características relevantes de la EBM	13
1.4. Fuentes de error en la aplicación de la EBM	14
1.4.1. Supersaturación de los hidrocarburos líquidos del yacimiento	14
1.4.2. Selección inadecuada de la relación PVT a utilizarse en los cálculos mediante la EBM	14
1.4.3. Presión promedio del yacimiento	14
1.4.4. Errores de medición en los volúmenes de fluidos producidos	14
1.4.5. Acuíferos activos y descensos leves de presión	15
1.4.6. Estimados de m	15
1.4.7. Concepto de petróleo activo	16
1.5. Métodos de Balance de Materiales	17
1.5.1. Método F Vs E_t	17
1.5.2. Método $(F - W_e)/(E_o + E_{fw})$ Vs $(E_g + E_{fw})/(E_o + E_{fw})$	19
1.5.3. Método F/E_t Vs W_e/E_t	20
1.6. Índices de Producción	21
2. Fundamentos de estadística	23
2.1. Poblaciones y muestras	23
2.2. Descripción de datos	24
2.2.1. Descripción numérica de datos	24
2.2.2. Presentación gráfica de datos	29

2.3.	Probabilidades	32
2.3.1.	Teoría de conjuntos	32
2.3.2.	Experimentos y espacios muestrales	33
2.3.3.	Eventos	34
2.3.4.	Probabilidad	35
2.4.	Variable aleatoria y esperanza	36
2.4.1.	Variable aleatoria	36
2.4.2.	Variabes aleatorias distribuidas conjuntamente	38
2.4.3.	Variabes aleatorias independientes	39
2.4.4.	Esperanza matemática	39
2.4.5.	Esperanza de una función de una variable aleatoria	40
2.4.6.	Esperanza de sumas de variables aleatorias	41
2.4.7.	Varianza	41
2.4.8.	Covarianza y correlación	41
2.4.9.	Funciones generadoras de momentos	42
2.5.	Distribución normal	43
2.5.1.	Propiedades de la distribución normal	43
2.5.2.	Media y varianza de la distribución normal	44
2.5.3.	Distribución normal estándar	45
2.6.	Distribución Gamma	46
2.7.	Distribuciones que se derivan de la normal	48
2.7.1.	La distribución chi cuadrada (χ^2)	48
2.7.2.	La distribución t	50
2.7.3.	La distribución F	52
2.8.	Distribuciones de estadísticos muestrales	53
2.8.1.	La media muestral	53
2.8.2.	El teorema del límite central	53
2.8.3.	Distribución aproximada de la media muestral	54
2.8.4.	La varianza muestral	54
2.9.	Estimación de parámetros e intervalos de confianza	56
2.9.1.	Estimación de máxima verosimilitud	56
2.9.2.	Intervalos de confianza	57
2.10.	Pruebas de hipótesis	60
2.10.1.	Nivel de significancia	60
2.10.2.	Pruebas de hipótesis sobre la media de una población normal	61
2.10.3.	Pruebas de hipótesis sobre la varianza de una población normal	64
3.	Herramientas estadísticas a utilizar	66
3.1.	Regresión lineal simple y correlación	66
3.1.1.	Regresión lineal simple	66

3.1.2.	Coeficiente de determinación	70
3.1.3.	Coeficiente de correlación	70
3.1.4.	Inferencias sobre los estimadores de mínimos cuadrados	71
3.2.	Análisis de Varianza	73
3.2.1.	Análisis de varianza de un factor	74
3.2.2.	Análisis de varianza de dos factores	77
3.3.	Diseño de experimentos	82
3.3.1.	Tipos de diseños	82
4.	Antecedentes del problema	100
4.1.	Investigación de McEwen	100
4.2.	Investigación de Tehrani	102
4.3.	Investigación de Fair	105
4.4.	Investigación de Carlson	106
4.5.	Investigación de Wang y Hwan	108
4.6.	Investigación de Walsh	109
4.7.	Investigación de Baker, Regier y Sinclair	110
5.	Error en la medición de la presión y de los parámetros PVT	112
5.1.	Introducción	112
5.2.	Error en la medición de presión	113
5.3.	Error en la medición de la gravedad API del petróleo	114
5.4.	Error en la medición de la gravedad específica del gas	114
5.5.	Error en la medición de la temperatura del yacimiento	114
5.6.	Error en la medición de la relación gas-petróleo en solución	115
6.	El Programa EAM (Error Analysis Module)	116
6.1.	Inicializando la interfase	116
6.2.	Menú File	117
6.3.	Cargando y visualizando los datos	118
6.3.1.	Cargando los datos de producción	119
6.3.2.	Cargando los datos PVT	120
6.3.3.	Cargando el modelo de acuífero	123
6.3.4.	Cargando los datos del yacimiento	129
6.3.5.	Visualización de los datos cargados de producción y PVT	131
6.4.	Realizaciones de presión y PVT	134
6.4.1.	Introduciendo los parámetros para el cálculo de las realizaciones	134
6.4.2.	Visualizando los resultados de las realizaciones	136
6.4.3.	Archivo de salida para el cálculo de realizaciones	149
6.5.	Diseño experimental para los datos presión y PVT	151

6.5.1.	Introduciendo los parámetros para el cálculo del diseño experimental	151
6.5.2.	Visualizando los resultados del diseño experimental	152
6.5.3.	Archivo de salida para los cálculos del diseño experimental	158
7.	Metodología	159
7.1.	Revisión bibliográfica	159
7.2.	Determinación del análisis a utilizar	160
7.3.	Modelos de yacimientos	160
7.4.	Rangos de incertidumbre a evaluar	162
7.5.	Proceso de cálculo	163
7.5.1.	Realizaciones	165
7.5.2.	Diseño experimental	169
7.6.	Procedimiento para el análisis	170
8.	Análisis de resultados	173
8.1.	Realizaciones	173
8.1.1.	Historia de presión	174
8.1.2.	Gravedad API	189
8.1.3.	Gravedad específica del gas	202
8.1.4.	Relación gas petróleo en solución	215
8.1.5.	Temperatura del yacimiento	228
8.1.6.	Diagramas ternarios	241
8.1.7.	Resumen de resultados	256
8.1.8.	Análisis final para las realizaciones	261
8.2.	Diseño experimental	263
8.2.1.	Diseños full factorial	267
8.2.2.	Diseños Box Behnken	283
8.2.3.	Resumen de resultados	299
8.2.4.	Comparación de los diseños full factorial y Box Behnken	303
	Conclusión	307
	Recomendaciones	309
A.	Modelos de yacimientos	313
A.1.	Dimensiones del yacimiento	313
A.1.1.	Número y dimensiones de las celdas utilizados	313
A.1.2.	Dimensiones de la capa de petróleo	313
A.1.3.	Dimensiones de la capa de gas	314
A.2.	Propiedades físicas de la roca y de los fluidos	314
A.2.1.	Porosidad	314

A.2.2. Permeabilidad	314
A.2.3. Varios	314
A.2.4. Permeabilidad relativa y presión capilar	315
A.2.5. Propiedades PVT	316
A.3. Características y localización de los pozos	319
A.4. Descripción de la capa de gas y acuífero	320
B. Ejemplo para la metodología de realizaciones	322
C. Ejemplo para la metodología de diseño experimental	330

Índice de figuras

1.1. Método F Vs E_t mostrando un ajuste perfecto	18
1.2. Método F Vs E_t , influencia del valor de m en la gráfica. Aparece en la referencia [7]	18
1.3. Método $(F - W_e)/(E_o + E_{fw})$ Vs $(E_g + E_{fw})/(E_o + E_{fw})$	19
1.4. Método F/E_t Vs W_e/E_t	20
1.5. Método F/E_t Vs W_e/E_t . Aparece en la referencia [7]	20
1.6. Gráfico basado en los índices de producción	22
2.1. Ejemplo de datos sesgados: (a) hacia la derecha, (b) simétricos, (c) hacia la izquierda. Aparece en [13]	25
2.2. Ejemplos de diferentes coeficientes de correlación muestral. (a) $r < 0$, (b) $r \approx 1$, (c) $r \approx 0$. Aparece en [13]	28
2.3. Ejemplo de histograma	30
2.4. Ejemplo de Diagrama de caja	31
2.5. (a) $A \subset U$, (b) $A \subset B$. Aparece en [15]	32
2.6. (a) $A \cup B$, (b) $A \cap B$. Aparece en [15]	33
2.7. (a) $A \cap C$, (b) $B \cap C$, (c) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$. Aparece en [15]	34
2.8. Función de densidad de probabilidad, $f(x) = \{e^{-x}, x \geq 0\}$	38
2.9. Función de densidad normal con: (a) μ y σ arbitrarias, (b) $\mu = 0$ y $\sigma = 1$	44
2.10. Función de densidad normal estándar	46
2.11. Función de densidad normal estándar, $P\{Z > Z_\alpha\} = \alpha$	46
2.12. Función de densidad chi cuadrada con 8 grados de libertad	49
2.13. Distribución t, $-t_{\alpha,n} = t_{1-\alpha,n}$	50
2.14. Comparación de la función de densidad normal estándar con la densidad T_5	51
2.15. Distribución F	52
2.16. Intervalos de confianza, σ^2 conocida	58
2.17. Intervalo de confianza, σ^2 desconocida	59
2.18. Prueba de hipótesis con σ^2 conocida	63
2.19. Distribución t, $-t_{\alpha,n} = t_{1-\alpha,n}$	64

3.1.	Diseño full factorial de dos niveles 2^3 , con factores X1, X2, y X3. Aparece en [18]	84
3.2.	Esquema de partición para un diseño full factorial de dos niveles 2^3 , usando esquinas alternadas. Tomado de [18]	90
3.3.	Diseño full factorial de tres niveles 3^2 . Tomado de [18]	93
3.4.	Diseño full factorial de tres niveles 3^3 . Tomado de [18]	94
3.5.	Función de información para un diseño cuadrático rotativo. Aparece en [19]	95
3.6.	Función de información para un diseño full factorial 3^2 . Aparece en [19]	95
3.7.	Generando un diseño central composite de dos factores. Tomado de [18]	96
3.8.	Tipos de diseños central composite de tres factores. Tomado de [20]	97
3.9.	Comparando los tres tipos de diseño central composite de dos factores. Tomado de [18]	98
3.10.	Diseño Box-Behnken de tres factores. Tomado de [20]	99
5.1.	Exactitud y precisión. Data Acquisition Toolbox User's Guide. The Mathworks, Inc. 2002. Página 1-29	113
6.1.	Pantalla principal de la interfase	117
6.2.	Ventana Load data-View data	118
6.3.	Ventana Reset load data	118
6.4.	Archivo *.PROD	119
6.5.	Archivo *.PVT	120
6.6.	Caja de dialogo para las correlaciones PVT	121
6.7.	Mensaje emergente preguntando acerca de las condiciones de operación del separador	121
6.8.	Caja de dialogo para las condiciones de operación del separador	122
6.9.	Caja de dialogo para el modelo Pot Acuífer	123
6.10.	Archivo AQ para el modelo Pot Acuífer	123
6.11.	Caja de dialogo para los modelos de Hurst y Schiltuis	124
6.12.	Archivo AQ para los modelos de Schiltuis, y Hurst	124
6.13.	Caja de dialogo para el modelo de Fetkovich	125
6.14.	Archivo AQ para el modelo de Fetkovich	126
6.15.	Caja de dialogo para el modelo de Carter-Tracy	127
6.16.	Archivo AQ para el modelo de Carter-Tracy	128
6.17.	Caja de dialogo para los datos del yacimiento	129
6.18.	Archivo *.RES	130
6.19.	Visualizando los datos de producción	131
6.20.	Visualizando los datos PVT	132

6.21. Visualizando el perfil de presiones del yacimiento	133
6.22. Ventana Realizations	134
6.23. Caja de dialogo Parameter standard deviation	135
6.24. Caja de dialogo Number of Realizations	135
6.25. Ventana mostrando el tiempo remanente de espera y el porcentaje de cálculos realizados	136
6.26. Ventana Realizations - Results Plots	136
6.27. Tabla y diagrama de caja para One-Way ANOVA	137
6.28. Ventana de selección para escoger los métodos que se van a graficar	138
6.29. Métodos de balance de materiales mostrando todas las realizaciones calculadas	138
6.30. Ventana de selección con la opción Energy Plot, disponibles para los submenús Minimum y Maximum RMSE	139
6.31. Gráfico de los métodos de balance de materiales junto con su corres- pondiente gráfico de los índices de producción	140
6.32. Ejemplo del gráfico pdf's-CDF	141
6.33. Ejemplo del gráfico de realizaciones de presión	142
6.34. Ejemplo del gráfico de realizaciones de las propiedades PVT	143
6.35. Ventana Error plots	144
6.36. Caja de dialogo para el POES estimado	144
6.37. Gráfico del Error en el POES vs desviación estándar del parámetro escogido	145
6.38. Gráfico del Error en el POES vs el número de realizaciones	146
6.39. Ventana de selección para el gráfico ternario	147
6.40. Ejemplo de gráfico ternario	148
6.41. Archivo *.MBC	149
6.42. Archivo *.MBC, continuación	150
6.43. Ventana DoE	151
6.44. Ventana Experimental design - Results plots	152
6.45. Ventana de visualización del diseño experimental	153
6.46. Ejemplo de diagrama 1-D	154
6.47. Ejemplo de diagrama 2-D	154
6.48. Ejemplo de diagrama 3-D	155
6.49. Ejemplo de diagrama 4-D	155
6.50. Ejemplo de una tabla N-Way ANOVA	156
6.51. Ejemplo de una tabla N-Way ANOVA para el método ($F-W_e$) vs E_t	157
6.52. Archivo *.MBC para diseño experimental	158
7.1. Diagrama mostrando la configuración del tamaño de la capa de gas y el tamaño relativo del acuífero	161
7.2. Diagrama mostrando el esquema del proceso de cálculo	164

7.3.	Cálculo de las realizaciones de presión	166
7.4.	Cálculo de las realizaciones PVT	167
7.5.	Gráfico ternario para la superficie de error	169
8.1.	Gráfico del error en el POES vs. la desviación estándar de la historia de presión para el caso 1	175
8.2.	Gráfico del error en el POES vs. el número de realizaciones para el caso 1	175
8.3.	Gráfico del error en el POES vs. la desviación estándar de la historia de presión para el caso 5	178
8.4.	Gráfico del error en el POES vs. el número de realizaciones para el caso 5	179
8.5.	Gráfico del error en el POES vs. la desviación estándar de la historia de presión para el caso 11	182
8.6.	Gráfico del error en el POES vs. el número de realizaciones para el caso 11	183
8.7.	Gráfico del error en el POES vs. la desviación estándar de la historia de presión para el caso 15	186
8.8.	Gráfico del error en el POES vs. el número de realizaciones para el caso 15	187
8.9.	Gráfico del error en el POES vs. la desviación estándar de la gravedad API para el caso 1	190
8.10.	Gráfico del error en el POES vs. el número de realizaciones para el caso 1	190
8.11.	Gráfico del error en el POES vs. la desviación estándar de la gravedad API para el caso 5	192
8.12.	Gráfico del error en el POES vs. el número de realizaciones para el caso 5	193
8.13.	Gráfico del error en el POES vs. la desviación estándar de la gravedad API para el caso 11	196
8.14.	Gráfico del error en el POES vs. el número de realizaciones para el caso 11	197
8.15.	Gráfico del error en el POES vs. la desviación estándar de la gravedad API para el caso 15	199
8.16.	Gráfico del error en el POES vs. el número de realizaciones para el caso 15	200
8.17.	Gráfico del error en el POES vs. la desviación estándar de la gravedad específica del gas para el caso 1	203
8.18.	Gráfico del error en el POES vs. el número de realizaciones para el caso 1	203

8.19. Gráfico del error en el POES vs. la desviación estándar de la gravedad específica del gas para el caso 5	205
8.20. Gráfico del error en el POES vs. el número de realizaciones para el caso 5	206
8.21. Gráfico del error en el POES vs. la desviación estándar de la gravedad específica del gas para el caso 11	209
8.22. Gráfico del error en el POES vs. el número de realizaciones para el caso 11	210
8.23. Gráfico del error en el POES vs. la desviación estándar de la gravedad específica del gas para el caso 15	212
8.24. Gráfico del error en el POES vs. el número de realizaciones para el caso 15	213
8.25. Gráfico del error en el POES vs. la desviación estándar de la relación gas petróleo en solución para el caso 1.	216
8.26. Gráfico del error en el POES vs. el número de realizaciones para el caso 1	216
8.27. Gráfico del error en el POES vs. la desviación estándar de la relación gas petróleo en solución para el caso 5	218
8.28. Gráfico del error en el POES vs. el número de realizaciones para el caso 5	219
8.29. Gráfico del error en el POES vs. la desviación estándar de la relación gas petróleo en solución para el caso 11	222
8.30. Gráfico del error en el POES vs. el número de realizaciones para el caso 11	223
8.31. Gráfico del error en el POES vs. la desviación estándar de la relación gas petróleo en solución para el caso 15	225
8.32. Gráfico del error en el POES vs. el número de realizaciones para el caso 15	226
8.33. Gráfico del error en el POES vs. la desviación estándar de la temperatura del yacimiento para el caso 1	229
8.34. Gráfico del error en el POES vs. el número de realizaciones para el caso 1	229
8.35. Gráfico del error en el POES vs. la desviación estándar de la temperatura del yacimiento para el caso 5	231
8.36. Gráfico del error en el POES vs. el número de realizaciones para el caso 5	232
8.37. Gráfico del error en el POES vs. la desviación estándar de la temperatura del yacimiento para el caso 11	235
8.38. Gráfico del error en el POES vs. el número de realizaciones para el caso 11	236

8.39. Gráfico del error en el POES vs. la desviación estándar de la temperatura del yacimiento para el caso 15	238
8.40. Gráfico del error en el POES vs. el número de realizaciones para el caso 15	239
8.41. Gráfico ternario para la historia de presión. $\sigma = 1$ psi	242
8.42. Gráfico ternario para la historia de presión. $\sigma = 100$ psi	243
8.43. Gráfico ternario para la gravedad API. $\sigma = 0,05$ °API	245
8.44. Gráfico ternario para la gravedad API. $\sigma = 0,25$ °API	246
8.45. Gráfico ternario para la gravedad específica del gas. $\sigma = 6e - 05$.	248
8.46. Gráfico ternario para la gravedad específica del gas. $\sigma = 0,0062$. .	249
8.47. Gráfico ternario para la relación gas petróleo en solución. $\sigma = 0,738$ SCF/STB	251
8.48. Gráfico ternario para la relación gas petróleo en solución. $\sigma = 14,76$ SCF/STB	252
8.49. Gráfico ternario para la temperatura del yacimiento. $\sigma = 0,02$ °F .	254
8.50. Gráfico ternario para la temperatura del yacimiento. $\sigma = 0,45$ °F .	255
8.51. Clasificación de los parámetros de acuerdo a la incertidumbre en los POES calculados y los métodos utilizados. Gráfico A, incluyendo la presión. Gráfico B, excluyendo la presión	262
A.1. Localización de los pozos	319
A.2. Diagrama mostrando la configuración del tamaño de la capa de gas y el tamaño relativo del acuífero	321
B.1. Realizaciones para los métodos de balance de materiales	323
B.2. Realización con el máximo RMSE para el ajuste	324
B.3. Realización con el mínimo RMSE para el ajuste	325
B.4. Gráficos de distribución de probabilidad para los POES calculados	327
B.5. Prueba ANOVA para determinar los efectos de los métodos de balance de materiales	328
B.6. Realizaciones de la historia de presión para $\sigma = 1$ psi	329
C.1. Gráfico de todas las corridas para los métodos de balance de materiales	332
C.2. Corrida con el máximo RMSE para el ajuste	333
C.3. Corrida con el mínimo RMSE para el ajuste	334
C.4. Representación gráfica del diseño utilizado	335
C.5. Gráficos de distribución de probabilidad para los POES calculados	336
C.6. Cuadro ANOVA para el método (F-We vs. Et)	337
C.7. Cuadro ANOVA para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	337

Índice de cuadros

3.1.	Número de corridas para un diseño full factorial 2^k	83
3.2.	Diseño full factorial 2^3 en orden estándar	84
3.3.	Matriz de análisis para un diseño full factorial 2^3	85
3.4.	Configuraciones alta (+1), baja (-1), y estándar (0) para una operación de pulido	86
3.5.	Diseño full factorial 2^3 repetido dos veces y presentado en orden estándar	87
3.6.	Diseño full factorial 2^3 repetido dos veces y presentado en orden aleatorio	88
3.7.	Diseño full factorial 2^3 repetido dos veces y presentado en orden aleatorio con puntos centrales	89
3.8.	Diseño full factorial 2^3 en dos bloques	91
3.9.	Diseño full factorial 3^3	94
8.1.	Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero	174
8.2.	Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero	176
8.3.	Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$	180
8.4.	Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento, $r_e D = \infty$	184
8.5.	Desviación estándar de los POES para la realizaciones de la historia de presión	188
8.6.	Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero	189
8.7.	Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero	191
8.8.	Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$	194
8.9.	Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento, $r_e D = \infty$	198
8.10.	Desviación estándar de los POES para la realizaciones de la gravedad API	201

8.11. Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero . . .	202
8.12. Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero . . .	204
8.13. Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$	207
8.14. Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento, $r_e D = \infty$	211
8.15. Desviación estándar de los POES para la realizaciones de la grave- dad específica del gas	214
8.16. Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero . . .	215
8.17. Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero . . .	217
8.18. Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$	220
8.19. Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento, $r_e D = \infty$	224
8.20. Desviación estándar de los POES para la realizaciones de la relación gas petróleo en solución	227
8.21. Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero . . .	228
8.22. Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero . . .	230
8.23. Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$	233
8.24. Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento, $r_e D = \infty$	237
8.25. Desviación estándar de los POES para la realizaciones de la tem- peratura del yacimiento	240
8.26. Historia de presión. $\sigma = 1$ psi	256
8.27. Gravedad API. $\sigma = 0.05$ °API	256
8.28. Gravedad específica del gas. $\sigma = 5e-05$	257
8.29. Relación gas-petróleo en solución. $\sigma = 0.738$ SCF/STB	257
8.30. Temperatura del yacimiento. $\sigma = 0.02$ °F	258
8.31. Historia de presión. $\sigma = 100$ psi	258
8.32. Gravedad API. $\sigma = 0.25$ °API	259
8.33. Gravedad específica del gas. $\sigma = 0.00615$	259
8.34. Relación gas-petróleo en solución. $\sigma = 14.76$ SCF/STB	260
8.35. Temperatura del yacimiento. $\sigma = 0.45$ °F	260
8.36. Clasificación de los parámetros de acuerdo a la incertidumbre en los POES calculados	261
8.37. Notación para las matrices de los diseños experimentales	263
8.38. Matriz de diseño full factorial para todos los parámetros (5 factores)	264

8.39. Matriz de diseño full factorial para los parámetros PVT	265
8.40. Matriz de diseño Box Behnken para todos los parámetros (5 factores)	265
8.41. Matriz de diseño Box Behnken para los parámetros PVT	266
8.42. Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero . . .	267
8.43. Resultados de la matriz de diseño de los parámetros PVT para el caso 1	268
8.44. Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero . . .	269
8.45. Resultados de la matriz de diseño de los parámetros PVT para el caso 5	271
8.46. Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$	274
8.47. Resultados de la matriz de diseño de los parámetros PVT para el caso 11	277
8.48. Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento, $r_e D = \infty$	279
8.49. Resultados de la matriz de diseño de los parámetros PVT para el caso 15	281
8.50. Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero . . .	283
8.51. Resultados de la matriz de diseño de los parámetros PVT para el caso 1	284
8.52. Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero . . .	285
8.53. Resultados de la matriz de diseño de los parámetros PVT para el caso 5	287
8.54. Caso 11. Tamaño de la capa de gas, $m=0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$	290
8.55. Resultados de la matriz de diseño de los parámetros PVT para el caso 11	293
8.56. Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento, $r_e D = \infty$	295
8.57. Resultados de la matriz de diseño de los parámetros PVT para el caso 15	297
8.58. Caso 1	299
8.59. Caso 5	299
8.60. Caso 11	300
8.61. Caso 15	300
8.62. Caso 1	301
8.63. Caso 5	301
8.64. Caso 11	302
8.65. Caso 15	302

8.66. Valores de la desviación estándar del POES para los métodos de balance de materiales. Los diseños toman en cuenta la historia de presión	305
8.67. Valores de la desviación estándar del POES para los métodos de balance de materiales. Los diseños sólo involucran los parámetros PVT	306
A.1. Permeabilidad relativa al agua y presión capilar como función de la saturación de agua	315
A.2. Permeabilidad relativa al gas y presión capilar como función de la saturación de gas	315
A.3. Permeabilidad relativa al petróleo para los casos de agua-petróleo y agua connata-gas-petróleo	316
A.4. Propiedades de la roca	317
A.5. Densidades de los fluidos a condiciones estándar (lb/ft^3)	317
A.6. Propiedades PVT del agua	317
A.7. Propiedades PVT del gas	317
A.8. Propiedades PVT del petróleo	318
C.1. Matriz de diseño para el caso ejemplo	331

Introducción

Para realizar cálculos de balance de materiales, se necesitan un conjunto de datos, como por ejemplo: datos de producción, presión, PVT, etc. Es por ello que la calidad de estos datos, es decir su incertidumbre, tendrá influencia en los estimados del balance de materiales. La incertidumbre en los datos incluye el error en la medición de las propiedades PVT del laboratorio, datos de producción del campo y presiones promedio del yacimiento. Las ganancias de las industrias del petróleo y gas están basadas en las cantidades de petróleo y gas que producen, por tanto se puede decir razonablemente que la producción de petróleo y gas se mide generalmente con gran precisión y por consiguiente los errores en los datos de producción son pequeños, más no es así con los datos de presión y PVT.

Investigaciones previas ya han examinado los efectos de alguna de estas fuentes de incertidumbre por separado. McEwen [1] estudió los efectos de la incertidumbre en la presión para yacimientos con influjo de agua, los resultados mostraron que métodos de balance de materiales no eran confiables incluso bajo pequeñas incertidumbres en la presión. Tehrani [2] rearrregló la ecuación de balance de materiales en diferentes formas y las aplicó a un yacimiento con influjo de agua bajo pequeñas incertidumbres en la presión, encontró que no todas las ecuaciones presentaban la misma robustez frente a la incertidumbre en la presión, es más algunas de ellas mostraron gran sensibilidad a la poca incertidumbre en la presión. Wang y Hwan [3] estudiaron la influencia de los mecanismos de empuje sobre los estimados de hidrocarburos en sitio, sus conclusiones fueron que para yacimientos con influjo de agua la determinación de los estimados de hidrocarburos en sitio era más difícil cuanto mayor fuera el influjo de agua. Resultados similares se encontraron para yacimientos con capa inicial de gas, mientras más grande fuera la capa inicial de gas mayor es la incertidumbre en las estimaciones de balance de materiales. Walsh [4] estudió los efectos de la incertidumbre en la presión sobre distintos métodos de balance de materiales, sus principales conclusiones fueron que los métodos de Havlena-Odeh para yacimientos con capa inicial de gas no tenían tolerancia frente a la incertidumbre en la presión, además de que la linealidad de los gráficos de los métodos de balance de materiales no era un indicador confiable de la robustez de estos métodos. Baker, Regier y Sinclair [5] estudiaron los efectos de la incertidum-

bre del PVT sobre las estimaciones de balance de materiales, encontraron que el impacto de los errores en el PVT sobre las estimaciones de balance de materiales puede ser significativo si la declinación de presión es muy pequeña, o si se utiliza una correlación PVT sin ajustarla a los datos de campo.

Sin embargo, en ninguna de estas investigaciones se muestra un tratado unificado donde se evalúen los efectos combinados de los errores en la presión y PVT sobre las estimaciones de balance de materiales bajo un amplia gama de mecanismos de empuje. En este trabajo se estudian los efectos de la incertidumbre en los datos de presión y PVT en los estimados de balance de materiales bajo diversos mecanismos de empuje. Para ello se utilizan dos métodos, el primero evalúa el efecto de la incertidumbre de cada parámetro sobre los estimados de balance de materiales utilizando el método de realizaciones, el segundo evalúa los efectos combinados de las incertidumbres de todos los parámetros, utilizando el método de diseño experimental. Para ello se desarrolló un programa que permitiera aplicar esta metodología y que aplicara un análisis estadístico a los resultados.

Los resultados indicaron que las estimaciones de hidrocarburos en sitio son sensibles al método de balance de materiales utilizado, errores en los datos de presión y PVT, y al principal mecanismo de empuje que posea el yacimiento. Para yacimientos donde el principal mecanismo de empuje fue la expansión del petróleo más gas disuelto, y que existía poca incertidumbre en los datos de presión y PVT, además que se utilizó el método $(F-W_e$ vs. $E_t)$ se obtuvieron los estimados más exactos de balance de materiales de todos los casos estudiados. El peor caso, donde se obtuvieron los estimados menos exactos de balance de materiales, fue aquel donde el principal mecanismo de empuje fue por influjo de agua, existía una gran incertidumbre en los datos de presión, y se utilizó el método $(F-W_e)/(E_o+E_{fw})$ vs. $(E_g+E_{fw})/(E_o+E_{fw})$ ó el método $(F/E_t$ vs. $W_e/E_t)$ para los cálculos.

El trabajo de tesis de grado comienza por explicar los conceptos fundamentales de balance de materiales en el capítulo 1, en el capítulo 2 (Fundamentos de estadística) se exponen los conceptos necesarios para aplicar las herramientas estadísticas que se utilizarán en la metodología de esta investigación, el capítulo 3 explica las herramientas estadísticas utilizadas, entre ellas regresión lineal, análisis de varianza y diseño experimental. El capítulo 4 presenta un resumen de las principales investigaciones previas relacionadas al problema de esta investigación, el capítulo 5 presenta los errores actuales en la medición de la presión y los parámetros PVT, el capítulo 6 expone la documentación del programa desarrollado en este trabajo de grado, el capítulo 7 explica la metodología implementada a lo largo del trabajo, el capítulo 8 presenta un análisis de todos los resultados obtenidos de la implementación de la metodología, y por último, se presentan las conclusiones y recomendaciones.

Objetivos

Objetivo general

Evaluar la incertidumbre en los cálculos de balance de materiales debido a los errores en los datos de presión y PVT bajo diversos mecanismos de empuje.

Objetivos específicos

- Realizar un análisis estadístico de los efectos de la incertidumbre en los datos de presión y PVT sobre las estimaciones de balance de materiales bajo diversos mecanismos de empuje.
- Diseñar una interfase en la plataforma de programación MATLAB que permita evaluar los rangos de incertidumbre obtenidos en los cálculos de balance de materiales afectados por el mecanismo de empuje e incertidumbre en los datos de presión y PVT.

Capítulo 1

La Ecuación de Balance de Materiales

En este capítulo, primero se expone una breve reseña histórica acerca de los trabajos realizados relacionados con la Ecuación de Balance de Materiales (EBM), posteriormente se desarrollará la EBM Generalizada. También se establecen los usos y limitaciones de la EBM, así como las suposiciones en las cuales está basada. Por último, se exponen los diferentes métodos de balance de materiales que se utilizarán en este trabajo, así como también se definen los índices de producción.

1.1. Principales trabajos relacionados con la Ecuación de Balance de Materiales

Existen diversos trabajos publicados referente a los métodos de estimación de las reservas de hidrocarburos, como lo mencionan Omole y Ojo [6]. La mayoría de los trabajos se enfocan en varios aspectos del tema, que van desde las leyes y principios fundamentales que gobiernan la extracción de petróleo del yacimiento, derivación de ecuaciones definiendo la relación entre la cantidad de hidrocarburos producidos con las propiedades del yacimiento, entre otros.

Dentro de los primeros trabajos realizados acerca del tema se encuentra el de Coleman, Wilde y Moore ¹. Su estudio se basó en la declinación de la presión del yacimiento posterior a la producción de petróleo y gas. Presentaron una ecuación que define una relación entre la presión del yacimiento, la cantidad de petróleo y gas producido, la cantidad de gas en el yacimiento y las propiedades de los fluidos del yacimiento.

Schiltuis ², presentó una forma modificada de la ecuación de Coleman, Wilde y Moore. La ecuación de Schiltuis se puede describir como un balance volumétrico entre las cantidades de petróleo, gas y agua producida, con la declinación de presión del yacimiento que acompaña a la producción, la cantidad total de agua que pudo haber entrado al yacimiento y la cantidad total de petróleo y gas del yacimiento. La ecuación de Coleman, Wilde y Moore está basada en las leyes de gases perfectos y soluciones perfectas, a diferencia de ésta, la ecuación de Schiltuis usa la relación entre la presión y el volumen obtenido en el laboratorio a partir de muestras de petróleo y gas del yacimiento, resultando que la ecuación sea aplicable al estudio de yacimientos de alta presión. Otra mejora de la ecuación de Schiltuis sobre la de Coleman, Wilde y Moore es una simplificación del procedimiento de cálculo involucrado, la mayoría de los términos usados en la ecuación de Schiltuis pueden ser leídos directamente de curvas provenientes del laboratorio. Schiltuis, para la derivación de la ecuación, supuso que existe un estado de equilibrio instantáneo en el yacimiento. Esta suposición de equilibrio es tal que el yacimiento se comporta como si contuviera cantidades más pequeñas de petróleo y gas de las que realmente contiene. Tal estado de equilibrio nunca se alcanza. Una consecuencia de esto es que el contenido de hidrocarburos calculado mediante la ecuación de Schiltuis es siempre menor que el contenido real. Se cree que la cantidad calculada de petróleo es esa porción del petróleo contenida en la parte permeable e interconectada que contribuye activamente al mantenimiento de la presión del yacimiento. Schiltuis llamó a este fenómeno “petróleo activo”. La ecuación de Schiltuis no toma en cuenta la disminución en el volumen poroso debido al efecto combinado de la expansión del agua connata y la reducción en el volumen poroso del yacimiento. Schiltuis también propuso un modelo de flujo de agua el cual expresa la tasa de flujo de agua dentro del yacimiento a un tiempo cualquiera, como proporcional a la diferencia de presión entre, la presión original del yacimiento y la presión en el yacimiento en un instante dado.

¹S.P. Coleman, H.D. Wilde y T.V. Moore, *Quantitative effects of GOR on decline of average rock pressure*. TRANS AIME(1930). 86. 174. (citado por Omole-Ojo. 1993. pp. 6.)

²R.J. Schiltuis, *Active Oil Reservoir Energy*. TRANS AIME(1936). 118. 32. (citado por Omole-Ojo. 1993. pp. 6.)

El trabajo de Old ³ expuso el uso simultáneo de la EBM y la ecuación de Hurst de flujo de fluidos, aplicado a la determinación de las reservas de hidrocarburos. Estudió el comportamiento de un yacimiento de petróleo y evaluó las fuerzas naturales que actuaban en el yacimiento. Old afirmó que un uso importante de éste método de análisis consiste en determinar el comportamiento de presión.

Woods y Muskat ⁴ presentaron un procedimiento de análisis de mínimos cuadrados para resolver la EBM y su aplicación para estimar el petróleo en sitio a partir de observaciones de campo. El estudio concluyó que el balance de materiales por si mismo no puede, con seguridad, proporcionar una determinación única de las características físicas básicas del petróleo que se produce de un yacimiento. Sin embargo, el método proporciona una herramienta útil para estimar la intrusión de agua o para predecir el comportamiento futuro del yacimiento, cuando existen datos de control determinados independientemente, tales como valores de petróleo inicial en sitio y volumen de gas.

Everdingen, Timmerman y McMahon ⁵ presentaron una forma modificada de la EBM aplicable a yacimientos con empuje parcial de agua. El método combinó la EBM con la ecuación de influjo de agua de Hurst-Van Everdingen, para obtener valores confiables del petróleo activo original en sitio y una evaluación cuantitativa del influjo de agua acumulado. El método de solución usa el método de mínimos cuadrados para obtener dos ecuaciones normales a partir de un cierto número de ecuaciones de balance de materiales. El método de desviación estándar fue utilizado para determinar el valor de petróleo en sitio asociado con el valor más confiable de los intervalos de tiempo reducidos.

Hawkins ⁶ presentó una extensión de la EBM para aplicarse a yacimientos volumétricos subsaturados por encima del punto de burbujeo mediante la inclusión de un término que toma en cuenta la presencia de agua intersticial y su compresibilidad.

³R. F. Jr. Old, *Analyzing of Reservoir Performance*. TRANS AIME(1943). 151. 86. (citado por Omole-Ojo. 1993. pp. 8.)

⁴R. E. Woods y M. Muskat, *An Analysis of MB Calculations*. TRANS AIME(1943). 151. 73. (citado por Omole-Ojo. 1993. pp. 9.)

⁵A.F. Van Everdingen, E.H. Timmerman y J.J. McMahon, *Application of the MBE to a Partial Water Drive Reservoirs*. TRANS AIME(1953). 198. 51. (citado por Omole-Ojo. 1993. pp. 9.)

⁶M. F. Jr. Hawkins, *Material balance in expansion type reservoir above bubblepoint*. TRANS AIME(1953). 204. 267. (citado por Omole-Ojo. 1993. pp. 10.)

Tracy ⁷ presentó una forma simplificada de la EBM de Schiltuis. En la ecuación los términos de petróleo producido acumulado, el gas producido acumulado y el influjo neto de agua, se multiplican por diferentes factores de presión. El método estima tasas gas-petróleo instantáneas junto con producción incremental de petróleo.

Havlena y Odeh [7] presentaron un método en el cual la EBM fue expresada como la ecuación de una línea recta. El método consiste en graficar un conjunto de variables versus otro, dependiendo en el mecanismo de empuje del yacimiento. Este método proporciona un tercer y quizás un criterio necesario que sólo una solución exitosa de la EBM debería satisfacer. El método fue aplicado e ilustrado mediante varios casos de campo. El método ha demostrado ser el mejor en términos de la interpretación de los cálculos de balance de materiales.

Dake [8] consideró la disminución en el volumen poroso de hidrocarburos debido al efecto combinado de la expansión del agua connata y la reducción en el volumen poroso la cual no fue tomada en cuenta por Schiltuis al derivar la forma general de la EBM.

⁷G. W. Tracy, *Simplified form of the MBE*. SPE reprint series no 3. 1970. pp 62. (citado por Omole-Ojo. 1993. pp 11.)

1.2. Derivación de la Ecuación de Balance de Materiales

Primero es necesario definir el factor m , el cual corresponde al tamaño inicial de la capa de gas de un yacimiento con respecto al tamaño inicial de la zona de petróleo, ambos volúmenes expresados a condiciones de yacimiento, es decir, a la presión y temperatura del yacimiento. De la definición se puede observar que el valor de m es adimensional, y dado que m se definió como una proporción entre los volúmenes *iniciales* de la capa de gas y la zona de petróleo, este valor no cambia con el tiempo. También se puede deducir que el valor mínimo para m es cero.

La EBM representa un balance volumétrico aplicado a un volumen de control, definido como los límites iniciales de aquellas zonas ocupadas por hidrocarburos. La suma algebraica de todos los cambios volumétricos que ocurren en cada una de las zonas definidas dentro del volumen de control es igual a cero. Para el análisis volumétrico se definirán tres zonas: la zona de petróleo, la zona de gas y la zona de agua que existe dentro del volumen de control. Una de las principales suposiciones es que las tres fases (petróleo, gas y agua) estarán siempre en un equilibrio instantáneo dentro del yacimiento. Los cambios de volúmenes ocurrirán a partir de un tiempo $t=0$ a un tiempo $t=t$ cualquiera. Primero se procederá a definir los volúmenes iniciales en cada una de las zonas, luego los volúmenes remanentes al tiempo $t=t$, y por último la diferencia entre éstos representará la disminución en cada zona. Posteriormente seguirá una serie de manipulaciones matemáticas para llegar a la EBM Generalizada. Todo los volúmenes están expresados a condiciones de yacimientos.

Cambio volumétrico en la zona de petróleo

El volumen inicial de petróleo es:

$$N * B_{oi} \tag{1.1}$$

El volumen remanente de petróleo a un tiempo t es:

$$(N - N_p) * B_o \tag{1.2}$$

Por ende la reducción de volumen en la zona de petróleo es:

$$N * B_{oi} - (N - N_p) * B_o \tag{1.3}$$

Cambio volumétrico en la zona de la capa de gas

El factor m se definió como:

$$m = \frac{\text{gas inicial libre}}{\text{petróleo inicial}}, \text{ ambos a condiciones de yacimiento} \quad (1.4)$$

$$m = \frac{G_i B_{gi}}{N B_{oi}} \quad (1.5)$$

de aquí se obtiene:

$$G_i B_{gi} = m N B_{oi}, \text{ volumen inicial de gas libre} \quad (1.6)$$

Para determinar el volumen de gas libre dentro del volumen de control a un tiempo t , se realiza un balance volumétrico a *todo* el gas dentro del volumen de control (gas libre más el gas en solución).

$$\begin{aligned} \text{Gas libre al tiempo } t &= \text{Gas libre} + \text{Gas en solución}_{\text{inicial}} \\ &\quad - \text{Gas producido al tiempo } t \\ &\quad - \text{Gas en solución al tiempo } t \end{aligned} \quad (1.7)$$

expresando cada término en la nomenclatura apropiada se tiene:

$$\text{Gas libre al tiempo } t = \frac{m N B_{oi}}{B_{gi}} + N R_{si} - N_p R_p - (N - N_p) R_s \quad (1.8)$$

en donde:

$$R_p = \frac{G_p}{N_p} = \frac{\text{todo el gas neto producido}}{\text{todo el petróleo neto producido}} \quad (1.9)$$

El gas libre al tiempo t es:

$$\text{Gas libre al tiempo } t = B_g \left[\frac{m N B_{oi}}{B_{gi}} + N R_{si} - N_p R_p - (N - N_p) R_s \right] \quad (1.10)$$

Por ende la disminución volumétrica de la zona de gas libre al tiempo t es:

$$mNB_{oi} - B_g \left[\frac{mNB_{oi}}{B_{gi}} + NR_{si} - N_p R_p - (N - N_p)R_s \right] \quad (1.11)$$

Cambio volumétrico del agua dentro del volumen de control

El volumen inicial de agua dentro del volumen de control es:

$$W \quad (1.12)$$

El volumen final de agua dentro del volumen de control es:

$$W + W_e - W_p B_w \quad (1.13)$$

El aumento en el volumen de agua dentro del volumen de control es:

$$W_e - W_p B_w \quad (1.14)$$

Igualando la disminución en el volumen de las zonas de petróleo y de gas libre al aumento de agua dentro del volumen de control se tiene:

$$\begin{aligned} N * B_{oi} - (N - N_p) * B_o + mNB_{oi} - B_g \left(\frac{mNB_{oi}}{B_{gi}} \right) \\ + NR_{si} - N_p R_p - (N - N_p)R_s = W_e - W_p B_w \end{aligned} \quad (1.15)$$

Sumando y restando el término $N_p B_g R_s$ al lado izquierdo de la igualdad:

$$\begin{aligned} N * B_{oi} - NB_o + N_p B_o + mNB_{oi} - mNB_{oi} B_g / B_{gi} \\ - NB_g R_{si} + N_p R_p B_g + NB_g R_s - N_p B_g R_s + N_p B_g R_{si} \\ - N_p B_g R_{si} = W_e - W_p B_w \end{aligned} \quad (1.16)$$

Reagrupando términos se tiene:

$$\begin{aligned} N * B_{oi} + mNB_{oi} - N[B_o + (R_{si} - R_s)B_g] \\ + N_p [B_o + (R_{si} - R_s)B_g] + N_p B_g [R_p - R_s] \\ - mNB_{oi} B_g / B_{gi} = W_e - W_p B_w \end{aligned} \quad (1.17)$$

dado que:

$$B_t = B_o + (R_{si} - R_s)B_g \quad (1.18)$$

$$B_{ti} = B_{oi} \quad (1.19)$$

entonces la ecuación queda:

$$N * B_{ti} - NB_t + N_p B_t + N_p B_g [R_p - R_s] + mNB_{ti}(1 - B_g/B_{gi}) = W_e - W_p B_w \quad (1.20)$$

Reagrupando términos se tiene:

$$N * (B_{ti} - B_t) + N_p (B_t + B_g [R_p - R_s]) + mNB_{ti}(1 - B_g/B_{gi}) = W_e - W_p B_w \quad (1.21)$$

Reordenando la ecuación para reconocer los diferentes procesos que están ocurriendo en el yacimiento:

$$N_p (B_t + B_g [R_p - R_s]) + W_p B_w = N * (B_{ti} - B_t) + mNB_{ti}(1 - B_g/B_{gi}) + W_e \quad (1.22)$$

Se puede observar que el vaciamiento, el lado izquierdo de la ecuación, es igual a la expansión de las zonas de petróleo y gas libre, y al influjo de agua. Además se pueden identificar cada uno de los procesos que ocurren en el yacimiento (las producciones son acumuladas):

$$\text{Expansión en la zona de petróleo: } N * (B_t - B_{ti}) \quad (1.23)$$

$$\text{Expansión de la zona de gas libre: } mNB_{ti}(B_g/B_{gi} - 1) \quad (1.24)$$

$$\text{Influjo de agua: } W_e \quad (1.25)$$

$$\text{Producción de la zona de petróleo: } N_p B_t \quad (1.26)$$

$$\text{Producción de gas en la zona de gas: } N_p [R_p - R_s] B_g \quad (1.27)$$

$$\text{Producción de agua: } W_p B_w \quad (1.28)$$

Tomando en cuenta el cambio en el volumen poroso para hidrocarburos debido a la expansión del agua connata y la reducción en el volumen poroso, la EBM queda:

$$\begin{aligned}
N_p * (B_o + (R_p - R_s) * B_g) &= N * B_{oi} * ((B_o - B_{oi}) + (R_{si} - R_s) * B_g) / B_{oi} \\
&+ m * N * B_{oi} * (B_g / B_{gi} - 1) \\
&+ (1 + m) * N * B_{oi} * \left(\frac{C_w * S_{wc} + C_f}{1 - S_{wc}} * \Delta p \right) \\
&+ W_e - W_p * B_w \tag{1.29}
\end{aligned}$$

En su libro, Essenfeld y Barberii [9], derivan la EBM despreciando la expansión del agua connata y de la roca, explicando que es debido a las diferencias en magnitud de las compresibilidades del gas y del petróleo con respecto a la del agua y de la roca.

1.3. Características relevantes de la EBM

Se puede observar ciertas características acerca de la EBM que a primera vista pueden pasar desapercibidas, como menciona Dake [8], estas son:

- Es cero dimensional, sólo se evalúa en un punto del yacimiento.
- Muestra independencia del tiempo, aunque en algunos modelos de influjo de agua se muestra dependencia explícita del tiempo.
- Aunque la presión aparece sólo explícitamente en el término de la expansión de la roca y el agua connata, se encuentra implícita en los parámetros PVT, (B_o , R_s , y B_g), los cuales son dependientes de la presión. También es de hacer notar que los cálculos de influjo de agua son dependientes de la presión.
- No tiene forma diferencial, la EBM fue derivada comparando los volúmenes actuales a la presión p , con los volúmenes iniciales a la presión p_i .

1.4. Fuentes de error en la aplicación de la EBM

Essenfeld y Barberii [9] plantean varias situaciones posibles en las cuales no se cumplen los supuestos utilizados en la derivación de la EBM, esto se debe principalmente a que la suposición de equilibrio total e instantáneo entre las fases es bastante ideal y generalmente no ocurre.

1.4.1. Supersaturación de los hidrocarburos líquidos del yacimiento

Existen ciertos casos que al caer la presión en un yacimiento que contiene crudo saturado, sale gas de solución pero en un volumen inferior al pronosticado al análisis PVT efectuado bajo condiciones de equilibrio, es decir, se encuentra supersaturado con gas. Este efecto causa que la presión del yacimiento sea más baja de lo que sería si el equilibrio se hubiera alcanzado.

1.4.2. Selección inadecuada de la relación PVT a utilizarse en los cálculos mediante la EBM

Al usar la EBM es fundamental seleccionar un análisis PVT para los parámetros B_o , R_s , y B_g que a diferentes presiones represente apropiadamente, en su totalidad, la secuencia de fenómenos que han actuado en la producción de los fluidos medidos, desde el yacimiento pasando por el pozo hasta el separador. En las referencias [10, 11] se comenta que errores asociados a los datos PVT pueden producir grandes errores en los cálculos de los hidrocarburos en sitio.

1.4.3. Presión promedio del yacimiento

Debido a la naturaleza cero dimensional de la EBM y recordando la suposición del equilibrio total e instantáneo, el yacimiento se comporta como una celda o tanque ubicado en un “volumen de control”. De allí la suposición que todos los hidrocarburos para un momento dado se encuentran a la misma presión. Por tanto se debe tener en mente que las presiones utilizadas en la EBM deben ser representativas del sistema, y siempre y cuando sea factible debe utilizarse ponderación volumétrica de las presiones medidas.

1.4.4. Errores de medición en los volúmenes de fluidos producidos

Una de las principales fuentes de error en la aplicación de la EBM son las cifras erróneas de la producción de fluidos. Se sabe que para yacimientos con cru-

do subsaturado que tienen errores de medición, los estimados de N y W_e dan valores muy altos. Jones-Parra [12] expone muy bien esta situación: el petróleo fiscal no se mide necesariamente por yacimiento, se mide por estaciones de flujo y luego se prorratea al yacimiento. Cuando se prueba un pozo, se pasa de un separador de producción a determinadas presión y temperatura a un separador de prueba en el que las condiciones de presión y temperatura no son necesariamente las mismas. Una vez probados todos los pozos que fluyen a una estación se suma su tasa de producción para obtener una producción teórica por estación y determinar la fracción que cada pozo contribuye. Esta fracción luego se multiplica por la tasa de producción real de la estación para determinar el petróleo que se debe tomar como el volumen producido del yacimiento correspondiente. La producción de gas está sujeta a un control aún menos efectivo. Generalmente se hacen pruebas mensuales a la relación gas - petróleo, promediándose los valores obtenidos y multiplicándose por la producción de petróleo para obtener el volumen de gas producido. El volumen de agua que se produce también se mide en pruebas periódicas; pero como el agua no tiene ningún valor comercial se mide con muy poca precisión. Tomando en cuenta la incertidumbre en las mediciones de los volúmenes producidos, con frecuencia es necesario rectificar las cifras reportadas. El gas producido a veces se calcula multiplicando el volumen de petróleo producido por la relación gas - petróleo de la última prueba y se debe volver a calcular multiplicando por la relación promedio entre dos pruebas consecutivas.

1.4.5. Acuíferos activos y descensos leves de presión

Cuando el acuífero es muy activo o la capa de gas es muy grande, los cambios de presión a través del yacimiento son muy leves y esto acarrea grandes dificultades en la aplicación de la EBM, principalmente debido a que las diferencias de las propiedades PVT no son significativas y también influye la precisión con que se hayan medido en el laboratorio B_o , R_s , y B_g .

1.4.6. Estimados de m

La EBM supone que todo el gas libre del yacimiento se encuentra en la capa de gas y que todo el petróleo en la zona de petróleo. Sin embargo, en algunas oportunidades ocurre que hay saturación de petróleo en la capa de gas y hay gas en la zona de petróleo. En esos casos el valor de m debe ser calculado utilizando todo el gas libre y todo el petróleo en estado líquido, independientemente donde quiera que estén.

1.4.7. Concepto de petróleo activo

Existen casos en los cuales los descensos de presión causados por la producción y/o inyección de fluidos no afectan la totalidad de hidrocarburos contenidos en el yacimiento. Esto ocurre bajo diferentes circunstancias, por ejemplo: cuando el yacimiento es muy grande y ha habido poca producción; cuando en el yacimiento existen zonas con bajas permeabilidad y difusividad las cuales no han sido afectadas por los descensos de presión que hay en aquellas zonas mas permeables (con una mejor difusividad); etc. Cuando ocurren estas situaciones existen en el sistema, para ese momento, dos valores de N ; una fracción llamada petróleo activo (N activo) y la otra denominada petróleo inactivo (N inactivo). Se puede notar que la suma del petróleo activo y el inactivo conforman el petróleo total en sitio (N). Se debe tener en mente que el petróleo original en sitio no cambia, mas si lo hace la relación del volumen activo al inactivo con el tiempo, mas aún, el volumen de petróleo activo crece con el tiempo mientras el volumen del petróleo inactivo disminuye con el tiempo, hasta llegar al punto que todo el petróleo activo es igual al petróleo original en sitio. Por tanto para las situaciones antes mencionadas (gran tamaño, zonas de baja difusividad, etc.), los resultados de los cálculos con la EBM generan valores de N que corresponden al volumen de petróleo activo y no al petróleo original en sitio, por esta razón es que a medida que pasa el tiempo y se repite el cálculo, el valor de N calculado va creciendo debido a que representa el volumen de petróleo activo.

1.5. Métodos de Balance de Materiales

Los métodos a desarrollar a continuación son similares a aquellos que se encuentran en el trabajo de Havlena y Odeh [7], la diferencia entre los métodos que se expondrán y los de Havlena y Odeh reside en que en su trabajo, ellos despreciaron la contribución de la expansión de la roca y el agua connata mientras que aquí se tomarán en cuenta. El método de Havlena y Odeh consiste en agrupar ciertos términos en la EBM y graficar un conjunto de variables con respecto a otro. Dependiendo del mecanismo principal de empuje, se grafican diferentes conjuntos de términos contra otros, resultando que si el mecanismo de empuje elegido es el correcto, al igual que otros parámetros (como por ejemplo modelo de acuífero), se observará una relación lineal entre las variables graficadas. Esto permite la estimación de otros parámetros como por ejemplo N , m , y/o W_e , a partir del comportamiento lineal observado.

1.5.1. Método F Vs E_t

Primero es necesario definir los siguientes términos:

$$\begin{aligned}
 F &= N_p * (B_o + (R_p - R_{si}) * B_g) + W_p * B_w \\
 E_o &= (B_o + (R_{si} - R_s) * B_g) - B_{oi} \\
 E_g &= B_o / B_{gi} * (B_g - B_{gi}) \\
 E_{fw} &= B_{oi} (C_w * S_{wc} + C_f) / (1 - S_{wc}) * \Delta p \\
 E_t &= N * (E_o + m * E_g + (1 + m) * E_{fw})
 \end{aligned}$$

Por tanto la EBM queda de la siguiente forma:

$$F = N * E_t + W_e$$

Consecuentemente una gráfica de $(F - W_e)$ Vs E_t producirá una línea recta con pendiente N e intercepto 0, es decir, pasa por el origen del sistema de coordenada. Este método asume que el valor de m es correcto o cercano al verdadero, al igual que los valores de W_e , así como todas las otras suposiciones intrínsecas a la EBM. Por lo general, el término correspondiente a la expansión de la roca y el agua connata es despreciado, si además, el yacimiento no tiene capa de gas, se desprecia el término correspondiente a la capa de gas, y si no existe influjo de agua, la EBM desarrollada anteriormente queda: $F = N * E_o$, que es la forma que se dedujo en el trabajo de Havlena y Odeh. Si el valor de m es muy grande o muy pequeño, el gráfico se desviará por encima o por debajo, respectivamente, de la línea recta teórica para el valor correcto de m .

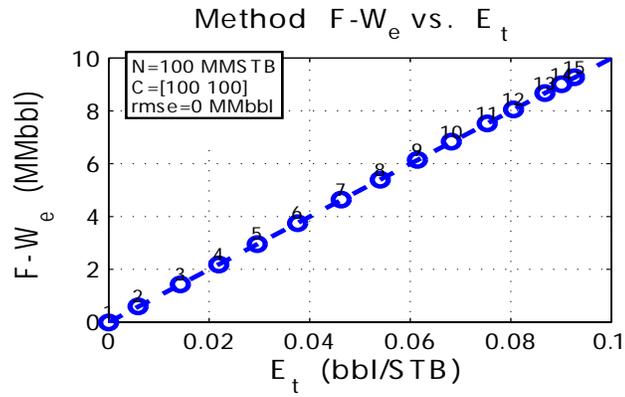


Figura 1.1: Método F Vs E_t mostrando un ajuste perfecto

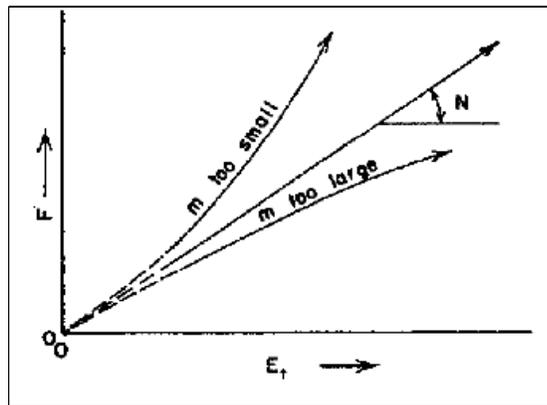


Figura 1.2: Método F Vs E_t , influencia del valor de m en la gráfica. Aparece en la referencia [7]

1.5.2. Método $(F - W_e)/(E_o + E_{fw})$ Vs $(E_g + E_{fw})/(E_o + E_{fw})$

Este método permite calcular simultáneamente los valores de N y m . Graficando $(F - W_e)/(E_o + E_{fw})$ Vs $(E_g + E_{fw})/(E_o + E_{fw})$ producirá una línea recta cuyo intercepto con el eje Y es N , y cuya pendiente es $m * N$. De nuevo, si tenemos un yacimiento donde se puede despreciar el término E_{fw} y no existe influjo de agua, el gráfico resultante será: F/E_o Vs E_g/E_o , nuevamente esta fue la forma desarrollada por Havlena y Odeh en su trabajo. También se puede observar que si no existe capa de gas, el gráfico resultante sería una línea horizontal con intercepto N .

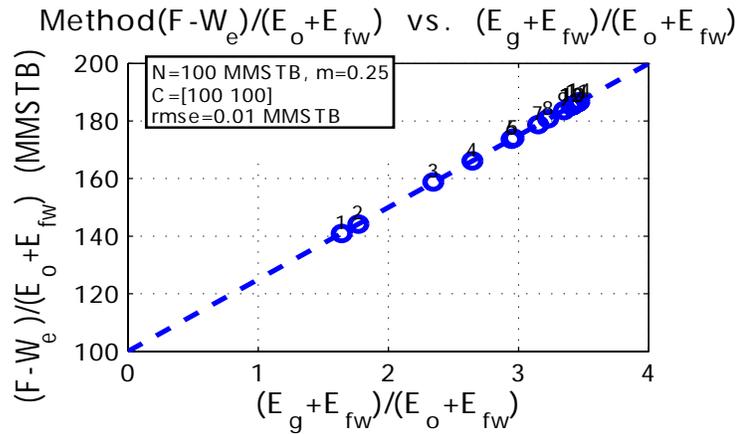


Figura 1.3: Método $(F - W_e)/(E_o + E_{fw})$ Vs $(E_g + E_{fw})/(E_o + E_{fw})$

1.5.3. Método F/E_t Vs W_e/E_t

Este método permite calcular N además de que pone una restricción adicional, aparte de un comportamiento lineal; estos es, la pendiente de la línea recta debe ser igual a 1. Si existen valores erróneos para el término de W_e , se verá reflejado en que el comportamiento se aleja de la tendencia lineal; específicamente, si el W_e asumido es demasiado grande, la tendencia es hacia debajo del comportamiento lineal; si el W_e asumido es demasiado pequeño, la tendencia es hacia arriba. Si la tendencia es en forma de S entonces se podría obtener un mejor ajuste de los datos si se asume un influjo de agua lineal.

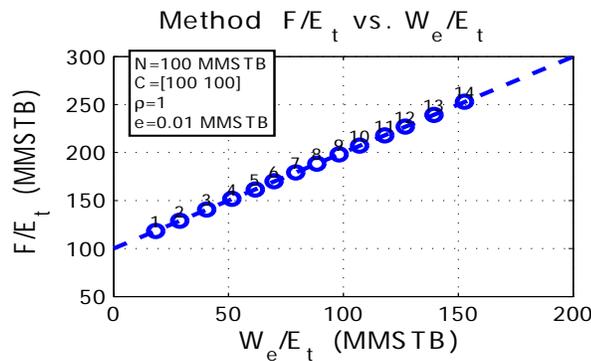


Figura 1.4: Método F/E_t Vs W_e/E_t

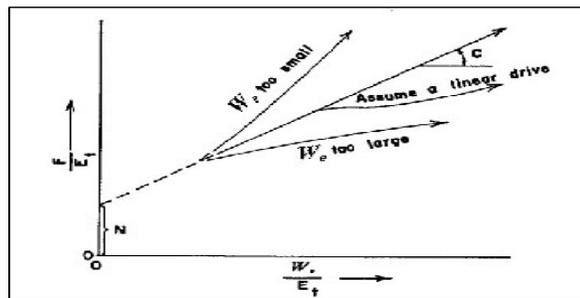


Figura 1.5: Método F/E_t Vs W_e/E_t . Aparece en la referencia [7]

1.6. Índices de Producción

Un concepto de interés práctico, en yacimientos con empujes simultáneos, para determinar la contribución relativa de los principales mecanismos que favorecen a la producción, son los índices de producción. Con este fin se reordena la EBM para obtener varias fracciones, cuya suma debe ser uno, éstas son: I_o , Índice de Producción por Agotamiento; I_g , Índice de Producción por Expansión de la Capa de Gas; I_w , Índice de Producción por Influjos de Agua; I_{fw} , Índice de Empuje por Expansión de la Roca y Agua Connata.

Cuando existe capa de gas generalmente el término debido a la expansión de la roca y el agua connata (I_{fw}) puede ser despreciado, quedando sólo tres principales mecanismos que contribuyen a la producción del petróleo y gas del yacimiento. Dividiendo la EBM (ec.1.29), por el término del lado izquierdo, es decir, el vaciamiento, queda:

$$\begin{aligned} & \frac{N * B_{oi} * ((B_o - B_{oi}) + (R_{si} - R_s) * B_g) / B_{oi}}{N_p * (B_o + (R_p - R_s) * B_g)} \\ & + \frac{m * N * B_{oi} * (B_g / B_{gi} - 1)}{N_p * (B_o + (R_p - R_s) * B_g)} \\ & + \frac{(1 + m) * N * B_{oi} * (\frac{C_w * S_{wc} + C_f}{1 - S_{wc}} * \Delta p)}{N_p * (B_o + (R_p - R_s) * B_g)} \\ & + \frac{W_e - W_p * B_w}{N_p * (B_o + (R_p - R_s) * B_g)} = 1 \end{aligned}$$

Definiendo cada uno de los términos de la ecuación anterior en la siguiente forma:

$$\begin{aligned} A &= N * B_{oi} * ((B_o - B_{oi}) + (R_{si} - R_s) * B_g) / B_{oi} \\ B &= m * N * B_{oi} * (B_g / B_{gi} - 1) \\ C &= (1 + m) * N * B_{oi} * (\frac{C_w * S_{wc} + C_f}{1 - S_{wc}} * \Delta p) \\ D &= W_e - W_p * B_w \\ E &= N_p * (B_o + (R_p - R_s) * B_g) \end{aligned}$$

los términos representan: A, las expansiones del petróleo más gas disuelto; B, de la capa de gas; C, la expansión de la roca y el agua connata; D, el término del influjo neto de agua y, E, el vaciamiento del yacimiento.

Rearreglando en forma de fracción de la energía total se tiene:

$I_o = A/E$, Índice de Producción por Agotamiento

$I_g = B/E$, Índice de Producción por Expansión de la Capa de Gas

$I_w = C/E$, Índice de Producción por Influjo de Agua

$I_{fw} = D/E$, Índice de Producción por Expansión de la Roca y Agua Connata

Lo anterior puede ser expresado como:

$$I_o + I_g + I_w + I_{fw} = 1$$

Es muy importante resaltar que los índices de producción, así como se presentaron en las ecuaciones anteriores, representan la contribución acumulada a través de toda la historia de producción del yacimiento. El gráfico a continuación es de utilidad para visualizar la importancia relativa de los mecanismos de empuje que actúan para un yacimiento dado.

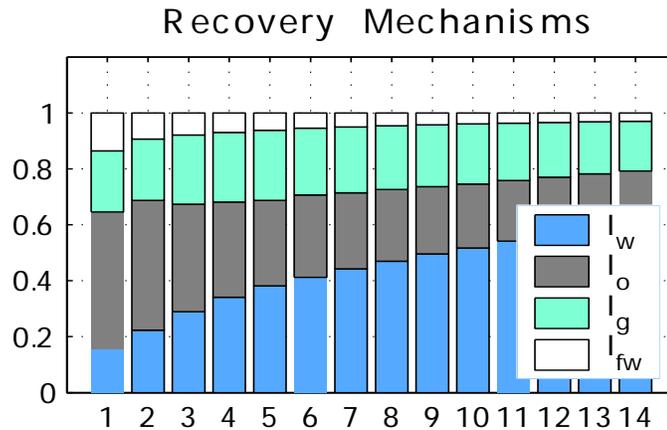


Figura 1.6: Gráfico basado en los índices de producción

Capítulo 2

Fundamentos de estadística

Este capítulo es un compendio de los conceptos más importantes relacionados con las herramientas estadísticas a utilizar en el presente trabajo. Las referencias [13, 14, 15] presentan información adicional acerca de estos temas.

2.1. Poblaciones y muestras

Según el diccionario enciclopédico Larousse [16] la estadística se define como: “ciencia cuyo objeto es reunir una información cuantitativa concerniente a individuos, grupos, series de hechos, etc. y deducir de ella, gracias al análisis de estos datos, unos significados precisos o unas previsiones para el futuro”. Es decir, la estadística está directamente relacionada con la recolección de datos, su descripción y análisis, a partir de los cuales se sacarán conclusiones. Se puede decir que la estadística se divide en dos ramas: la estadística *descriptiva*, la cual se encarga de describir y resumir datos; y la estadística *inferencial* que tiene que ver con la obtención de conclusiones a partir de los datos. Ahora bien, a la colección total de elementos de la cual se quiere obtener alguna información se le denomina *población*. Frecuentemente la población es demasiado grande como para que se examinen cada uno de sus miembros, es por ello, que se selecciona y examina a un subgrupo de los elementos. Este subgrupo recibe el nombre de *muestra*. Queda el tema de si la muestra seleccionada es representativa de toda la población. En general, no se puede asumir que una muestra dada sea representativa de toda la población si no se ha seleccionado de forma *aleatoria*, esto quiere decir que la elección se realiza de un modo tal que todas las formas posibles de elegir a los elementos sean igualmente probables. Si la selección de la muestra se hace de alguna forma no aleatoria, seguramente, la muestra estará sesgada hacia ciertos valores o directamente opuestos a otros.

2.2. Descripción de datos

2.2.1. Descripción numérica de datos

Medidas de tendencia central

La medida más común de tendencia central, o localización de los datos, es la media aritmética ordinaria. Debido a que casi siempre consideramos a los datos como la muestra, nos referimos a la media aritmética como la *media de la muestra*. Si las observaciones en una muestra de tamaño n son x_1, x_2, \dots, x_n entonces la media de la muestra es:

$$\begin{aligned}\bar{x} &= \frac{x_1 + x_2 + \dots + x_n}{n} \\ &= \left(\sum_{i=1}^n x_i \right) / n\end{aligned}\tag{2.1}$$

Al valor promedio de todas las observaciones en una población se le denomina la *media de la población*, y se denota por medio de la letra griega μ (mu). Cuando hay un número finito de observaciones (por ejemplo, N) en la población, entonces la media de la población es:

$$\mu = \left(\sum_{i=1}^N x_i \right) / N\tag{2.2}$$

Otra medida de tendencia central es la *mediana*, o punto en el cual la muestra se divide en dos mitades iguales. Sean x_1, x_2, \dots, x_n los elementos de una muestra arreglada en orden creciente de magnitud; esto es, x_1 denota la observación más pequeña, x_2 denota la segunda observación más pequeña, \dots , y x_n denota la observación más grande. Entonces la mediana se define matemáticamente como:

$$\tilde{x} = \begin{cases} (x_{n/2} + x_{n/2+1})/2, & n \text{ par} \\ x_{(n+1)/2}, & n \text{ impar} \end{cases}\tag{2.3}$$

La mediana tiene la ventaja de que no es afectada de manera considerable por los *valores extremos*, esto es, valores mucho más grandes o mucho más pequeños que los otros.

La *moda* es la observación que ocurre con mayor frecuencia en la muestra.

Si los datos son *simétricos*, entonces coinciden la media y la mediana. Si además, los datos sólo tienen una moda (diremos que los datos son unimodales), entonces coinciden la media, la mediana y la moda. Si los datos están *sesgados* (asimétricos, con una larga cola en un lado), la media, la mediana y la moda no

coincidirán. Suele encontrarse que la moda $<$ mediana $<$ media si la distribución es asimétrica hacia la derecha, en tanto que la moda $>$ mediana $>$ media si la distribución es asimétrica hacia la izquierda.

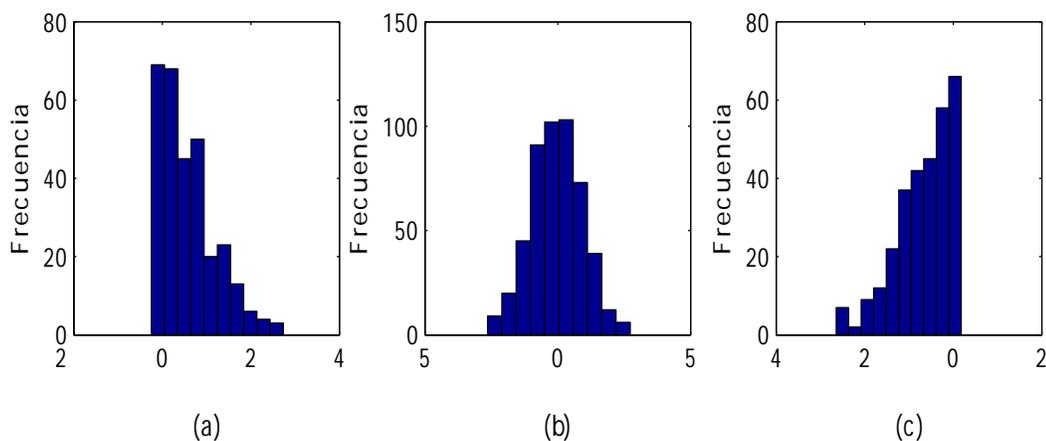


Figura 2.1: Ejemplo de datos sesgados: (a) hacia la derecha, (b) simétricos, (c) hacia la izquierda. Aparece en [13]

Cuartiles y otros percentiles

Además de la mediana, que divide en mitades un conjunto de datos, podemos considerar otros puntos de división. Cuando un conjunto ordenado de datos se divide en cuatro partes, los puntos de división resultantes se llaman *cuartiles de la muestra*. El primer cuartil, Q_1 , es un valor con un cuarto, o 25% , de las observaciones por debajo de su valor. Los cuartiles son el 25o., 50o., 75o. percentiles:

- Primer cuartil $Q_1=25\text{o. percentil}$
- Segundo cuartil $Q_2=50\text{o. percentil}$
- Tercer cuartil $Q_3=75\text{o. percentil}$

En términos generales se define el 100o.p percentil de la muestra como sigue: El 100o.p percentil de la muestra es el valor en el que al menos 100 p% de las observaciones son iguales o menores a este valor y al menos 100(1-p) % son iguales o mayores a este valor.

Medidas de dispersión

La medida de dispersión más importante es la *varianza de la muestra*. Si x_1, x_2, \dots, x_n es una muestra de n observaciones, entonces la varianza de la muestra es:

$$s^2 = \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right) / (n - 1) \quad (2.4)$$

Una fórmula computacional más eficiente (hablando de tiempo de CPU) para la varianza de la muestra es la siguiente:

$$s^2 = \left(\left(\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 / n \right) / (n - 1) \right) \quad (2.5)$$

Si s^2 es pequeña, hay una relativamente pequeña variabilidad en los datos, pero si s^2 es grande, la variabilidad es relativamente grande. Las unidades de medida para la varianza de la muestra son el cuadrado de las unidades originales de la variable, por tanto no es fácil interpretarla. El problema de la dimensionalidad se resuelve trabajando con la raíz cuadrada (positiva) de la varianza, s , denominada *desviación estándar de la muestra*. Esto brinda una medida de la dispersión expresada en las mismas unidades que la variable original.

De igual forma que para la muestra s^2 , hay una medida de variabilidad en la población llamada *varianza de la población*. Para denotarla se usará la letra griega σ^2 . La raíz cuadrada positiva de σ^2 , σ , denotará la *desviación estándar de la población*. Cuando la población es finita y consiste en N valores podemos definir la varianza de la población como:

$$\sigma^2 = \left(\sum_{i=1}^N (x_i - \mu)^2 \right) / N \quad (2.6)$$

Otra medida de dispersión útil es el *intervalo de la muestra* o también llamada *rango*:

$$R = \max(x_i) - \min(x_i) \quad (2.7)$$

El intervalo de la muestra se calcula con mucha facilidad, pero tiene la inconveniencia de que se ignora la información que existe entre las observaciones más pequeña y más grande. Para tamaños de muestra pequeños, digamos $n \leq 10$, esta pérdida de información no es demasiado seria, en algunas situaciones. El intervalo tiene una amplia aplicación en el control de calidad estadístico.

En ocasiones, se desea expresar la variación como una fracción de la media.

Una medida de la variación relativa denominada *coeficiente de variación de la muestra* se define como:

$$CV = \frac{s}{\bar{x}} \quad (2.8)$$

El coeficiente de variación es útil cuando se compara la variabilidad de dos o más conjuntos de datos que difieren de modo considerable en la magnitud de las observaciones.

Otra medida de dispersión es la llamada *rango intercuartil* la cual representa el monto de variación en la mitad intermedia de los datos y se describe mediante:

$$\text{Rango intercuartil} = \text{tercer cuartil} - \text{primer cuartil} = Q_3 - Q_1 \quad (2.9)$$

Coeficiente de correlación muestral

Dados s_x y s_y , la desviación estándar muestral de los valores x y la desviación estándar muestral de los valores y , respectivamente; el *coeficiente de correlación muestral*, denominado r , de un conjunto de parejas de datos (x_i, y_i) , $i = 1, \dots, n$ viene dado por:

$$\begin{aligned} r &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \end{aligned} \quad (2.10)$$

Cuando $r > 0$ se dice que el conjunto de parejas de datos están correlacionadas positivamente, y cuando $r < 0$ se dice que están correlacionadas negativamente.

Las propiedades del coeficiente de correlación muestral son:

1. $-1 \leq r \leq 1$
2. Si para constantes a y b , con $b > 0$, $y_i = a + bx_i$, $i = 1, \dots, n$, entonces $r=1$.
3. Si para constantes a y b , con $b < 0$, $y_i = a + bx_i$, $i = 1, \dots, n$, entonces $r=-1$.
4. Si r es el coeficiente de correlación muestral del conjunto de parejas de datos x_i, y_i , $i = 1, \dots, n$, entonces también es el coeficiente de correlación muestral de la parejas de datos $a + bx_i, c + dy_i$, $i = 1, \dots, n$, siempre y cuando b y d sean ambos positivos o negativos.

La propiedad 1 dice que el coeficiente de correlación muestral r está siempre entre -1 y +1. La propiedad 2 señala que r será igual a +1 cuando exista una relación lineal entre las parejas de datos de forma tal que a valores grandes de x correspondan valores grandes de y . La propiedad 3 asegura que r será igual a

-1 cuando exista una relación lineal entre las parejas de datos de forma tal que a valores pequeños de x correspondan valores grandes de y . La propiedad 4 implica que r no depende de las dimensiones en las cuales se midieron los datos.

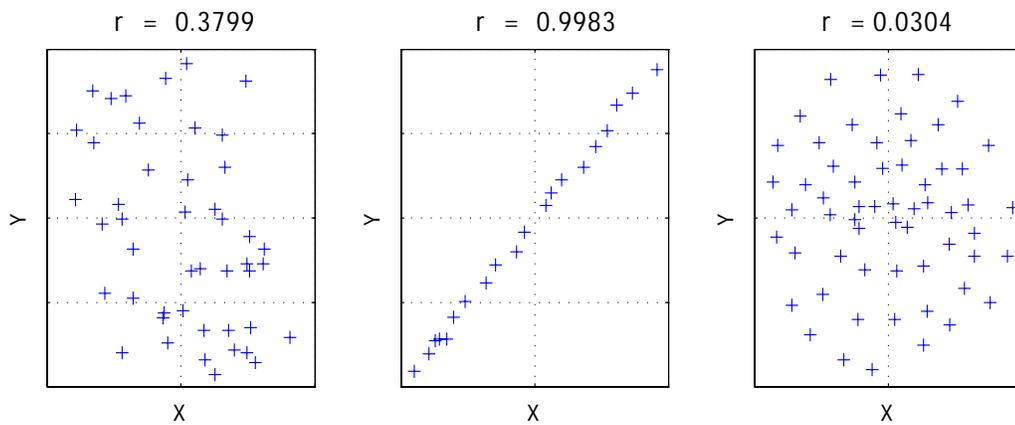


Figura 2.2: Ejemplos de diferentes coeficientes de correlación muestral. (a) $r < 0$, (b) $r \approx 1$, (c) $r \approx 0$. Aparece en [13]

2.2.2. Presentación gráfica de datos

La distribución de frecuencia, frecuencia relativa, frecuencia relativa acumulativa y el histograma

Una *distribución de frecuencias* es un resumen más compacto de datos que las observaciones originales. Para construir una distribución de frecuencia, se debe dividir la gama de datos en *intervalos de clase*. Generalmente, los intervalos de clase son de igual ancho, para incrementar la información visual en la distribución de frecuencias. El número de intervalos de clase que se utiliza depende del número de observaciones y de la cantidad de discriminación o dispersión en los datos. Se encuentra, en general, que entre 5 y 20 intervalos es satisfactorio en muchos casos, y que el número de intervalos de clase debe aumentar con n (el número de observaciones en un conjunto de datos). La elección del número de intervalos de clase aproximadamente igual a la raíz cuadrada del número de observaciones a menudo funciona bien en la práctica. La *distribución de frecuencia relativa* se determina dividiendo la frecuencia observada en cada intervalo de clase por el número total de observaciones. Las frecuencias relativas expresadas en una base acumulativa representa la *distribución de frecuencias relativas acumulativas*.

El histograma

El *histograma* es una representación gráfica de la distribución de frecuencias. Para dibujar un histograma, se usa el eje horizontal para representar la escala medida, y se dibujan las fronteras de los intervalos de clase, el eje vertical representa la escala de frecuencia (o frecuencia relativa). En la práctica, los intervalos de clase son de igual ancho, las alturas de los rectángulos dibujados en el histograma son proporcionales a las frecuencias. El histograma brinda una interpretación visual de la forma de la distribución de las mediciones, así como información acerca de la dispersión de los datos. Es importante acotar que la distribución de frecuencias y el histograma pueden ser relativamente sensibles a la elección del número de intervalos de clase y su ancho. En conjuntos de datos más pequeños, los histogramas pueden exhibir fluctuaciones un poco perturbadoras si el número y ancho de los intervalos de clase cambia. Las distribuciones de frecuencia y los histogramas son más estables para conjuntos de datos más grandes, por ejemplo de tamaño 50, 100 ó mayores.

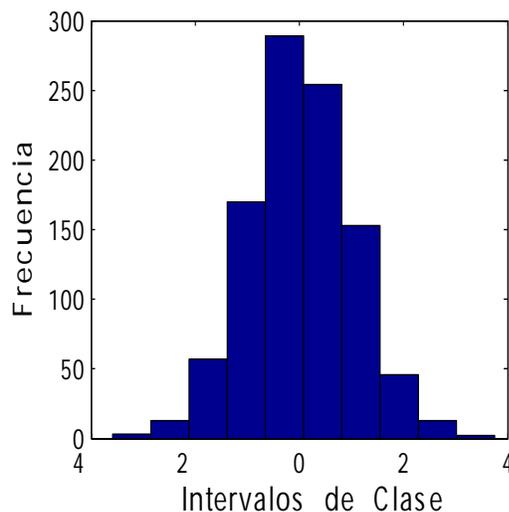


Figura 2.3: Ejemplo de histograma

El diagrama de caja

Un *diagrama de caja* exhibe los tres cuartiles, el mínimo y el máximo de los datos en una caja rectangular, alineada en forma horizontal o vertical. La caja encierra el intervalo intercuartil con la línea izquierda (o inferior) en el primer cuartil Q_1 y la línea derecha (o superior) en el tercer cuartil Q_3 . Se dibuja una línea a través de la caja en el segundo cuartil (que es el quincuagésimo percentil o la mediana) $Q_2 = \tilde{x}$. Una línea en cualquier extremo se extiende hasta los valores extremos. Estas líneas pueden extenderse sólo hasta los percentiles 10o. y 90o., o el 5o. y el 95o. en grandes conjuntos de datos. Los diagramas de caja son especialmente eficaces para la descripción gráfica de comparaciones entre conjuntos de observaciones.

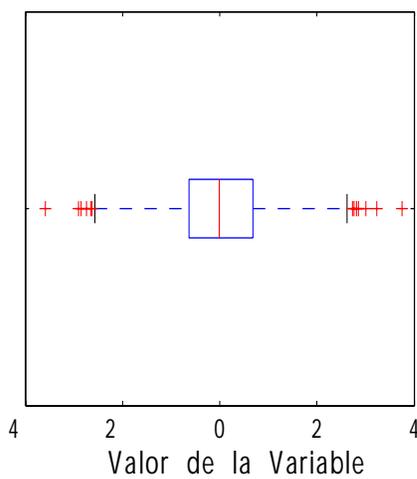


Figura 2.4: Ejemplo de Diagrama de caja

2.3. Probabilidades

2.3.1. Teoría de conjuntos

Antes de presentar los conceptos básicos de la probabilidad, se deben conocer algunos conceptos de la teoría de conjuntos. Se conoce como *conjunto* a un agregado o colección de objetos. A los conjuntos se le designa con letras mayúsculas, A, B, C , etc. A los componentes de un conjunto A se les llama *elementos* de A . En general, para decir que un elemento x es un elemento de A , se escribe $x \in A$; y si x no es un elemento de A , se escribe $x \notin A$. Con frecuencia los conjuntos se denotan así: $A = \{x_1, x_2, \dots, x_n\}$.

El conjunto de todos los objetos en consideración se llama *conjunto universal* y se denota por la letra U . Otro conjunto especial es el *conjunto nulo* o conjunto vacío puesto que no contiene ningún valor, se denota por \emptyset .

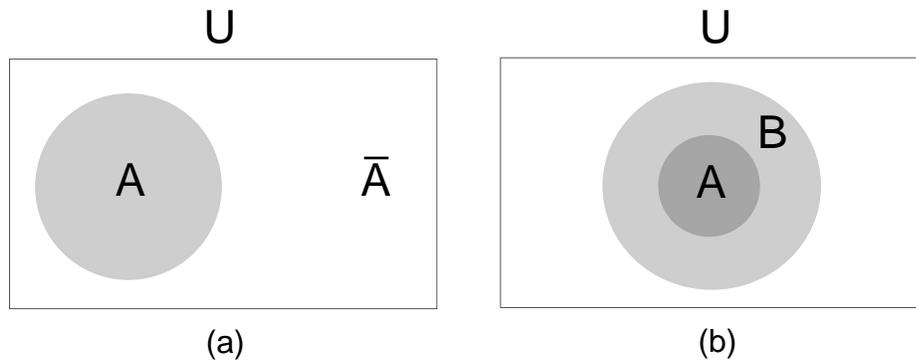


Figura 2.5: (a) $A \subset U$, (b) $A \subset B$. Aparece en [15]

Si se consideran dos conjuntos, por ejemplo A y B , se dice que A es un *subconjunto* de B , lo cual se denota como $A \subset B$, si y solo si cada elemento de A es un elemento de B . Se dice que los conjuntos A y B son iguales si cumplen con: $A \subset B$ y $B \subset A$. Ahora se consideraran algunas operaciones con conjuntos. Sean A y B subconjuntos cualesquiera del conjunto universal U . Entonces:

1. El *complemento* de A (con respecto a U) es el conjunto formado por los elementos de U que no pertenecen a A . A este conjunto complementario se denota por \bar{A} . Se escribe como: $\bar{A} = \{x; x \in U/x \notin A\}$
2. La intersección de A y B es el conjunto de elementos que pertenecen tanto a A como a B , denotándose como: $A \cap B$. Esto se escribe como: $A \cap B = \{x; x \in A \wedge x \in B\}$

3. La unión de A y B es el conjunto de elementos que pertenecen al menos a uno de los dos conjuntos A y B . Se puede expresar como: $A \cup B = \{x; x \in A \vee x \in B\}$

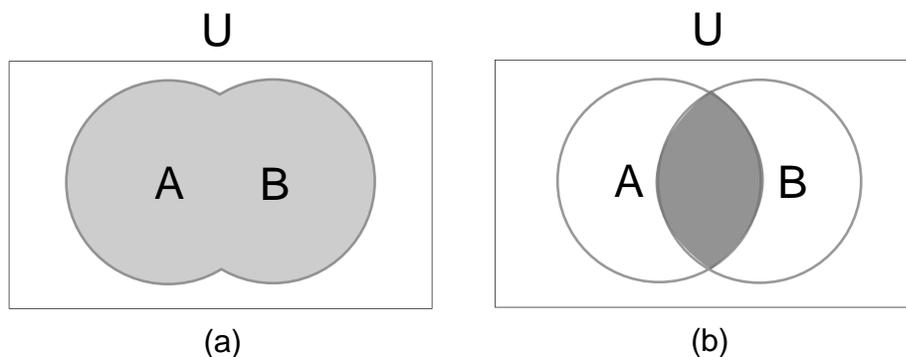


Figura 2.6: (a) $A \cup B$, (b) $A \cap B$. Aparece en [15]

A continuación se presentan varias leyes importantes de operaciones de conjuntos, algunas de las cuales ya se han definido:

- Leyes de identidad: $A \cup \emptyset = A$; $A \cap U = A$; $A \cup U = U$; $A \cap \emptyset = \emptyset$.
- Leyes de Morgan: $\overline{A \cup B} = \bar{A} \cap \bar{B}$; $\overline{A \cap B} = \bar{A} \cup \bar{B}$.
- Leyes asociativas: $A \cup (B \cup C) = (A \cup B) \cup C$; $A \cap (B \cap C) = (A \cap B) \cap C$
- Leyes distributivas: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$; $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

2.3.2. Experimentos y espacios muestrales

En la vida real existen situaciones en la que un investigador realiza un experimento y observa un resultado, el cual no puede predecirse con certeza. A estos experimentos se les llama *experimentos aleatorios*. Estos experimentos presentan ciertas características:

- Aunque no podemos predecir los resultados con certeza, si podemos predecir el conjunto de resultados posibles.
- Aunque el experimento se podría repetir en condiciones invariables, los resultados ocurren de una manera fortuita, desde un punto de vista conceptual.

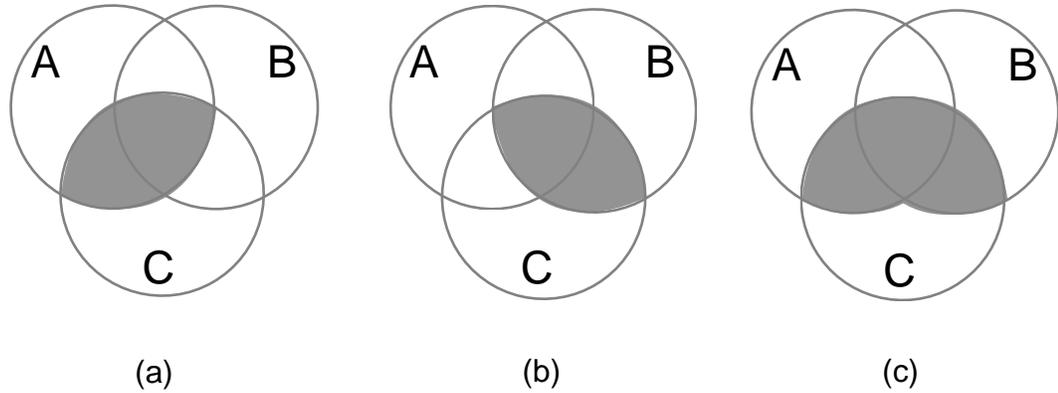


Figura 2.7: (a) $A \cap C$, (b) $B \cap C$, (c) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$. Aparece en [15]

También se consideran *experimentos idealizados*, esto se explica mejor con un ejemplo: cuando se lanza una moneda, se puede descartar la posibilidad de que no salgan ninguna de las dos caras sino que caiga en el borde de la moneda. Esta suposición se realiza más por conveniencia que por necesidad. A el conjunto de resultados posibles se le llama *espacio muestral*. Los símbolos \mathcal{E} y \mathcal{S} se utilizarán para representar al experimento aleatorio y el espacio muestral, respectivamente. Se puede diferenciar entre dos tipos de espacios muestrales: el *espacio muestral discreto* en el cual hay un número finito de resultados, y el *espacio muestral continuo* que tiene resultados incontables, como por ejemplo que el resultado es un intervalo y no un sólo número.

2.3.3. Eventos

A todo subconjunto E de un espacio muestral se denomina *evento*. Es decir, un evento es un conjunto de resultados posibles de un experimento. Si el resultado de un evento está contenido en E , entonces se dice que E ha ocurrido. El espacio muestral se considera el conjunto universal, por tanto E es un subconjunto de \mathcal{S} . También se puede notar que tanto E como \emptyset son subconjuntos de \mathcal{S} . Dado que un evento es un conjunto se le puede aplicar todas las leyes, propiedades y operaciones de conjuntos, por tanto se puede decir que, sean A y B dos eventos de un espacio muestral \mathcal{S} , entonces:

- $A \subset \mathcal{S}$ y $\bar{A} \subset \mathcal{S}$.
- Dos eventos A y B son mutuamente excluyentes si: $A \cap B = \emptyset$.

2.3.4. Probabilidad

Dado un evento E de un experimento con un espacio muestral \mathcal{S} , entonces se supone que existe un número, denotado por $P(E)$, que cumple con los siguientes axiomas:

1. $0 \leq P(E) \leq 1$
2. $P(\mathcal{S})=1$
3. Dado toda secuencia de eventos mutuamente excluyentes E_1, E_2, \dots , entonces: $P(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i)$, $n = 1, 2, \dots, \infty$.

A $P(E)$ se le denomina la probabilidad del evento E . El axioma 1 indica que la probabilidad de que el resultado de un experimento esté contenido en E es algún número entre 0 y 1. El axioma 2 establece que, con la probabilidad de 1, el resultado será un elemento contenido en el espacio muestral \mathcal{S} . El axioma 3 señala que para cualquier conjunto de eventos excluyentes, la probabilidad de que por lo menos uno de estos eventos ocurra es igual a la suma de sus respectivas probabilidades.

Para un gran número de experimentos es común suponer que cada punto en el espacio muestral tiene la misma posibilidad de ocurrir. Por ejemplo para un espacio muestral dado por $\mathcal{S} = \{x_1, x_2, \dots, x_n\}$ es natural suponer que:

$$P(\{x_1\}) = P(\{x_2\}) = \dots = P(\{x_n\}) = p$$

Recordando los axiomas 2 y 3, se tiene que:

$$1 = P(\mathcal{S}) = P(\{x_1\}) + \dots + P(\{x_n\}) = np$$

lo que muestra que:

$$P(\{x_i\}) = p = 1/n \tag{2.11}$$

Aplicando el axioma 3, se tiene que para todo evento E :

$$P(E) = \frac{\text{Números de puntos en } E}{n} \tag{2.12}$$

Es decir, suponiendo que todos los resultados de un experimento tienen la misma posibilidad de ocurrir, entonces la probabilidad de cada evento E es igual a la proporción de puntos en el espacio muestral que están contenidos en E .

2.4. Variable aleatoria y esperanza

2.4.1. Variable aleatoria

Cuando se realiza un experimento aleatorio, muy a menudo, no es relevante todos los detalles de los resultados del experimento, sino sólo algunas cantidades numéricas determinadas por los resultados. A estas cantidades se les conoce como *variables aleatorias*, y como están determinadas por el resultado de un experimento dado, se les puede asignar probabilidades a sus valores posibles. Se pueden distinguir dos tipos de variable aleatoria: las *variables aleatorias discretas* son aquellas que toman un número finito de valores posibles, las *variables aleatorias continuas* aquellas que toman valores continuos, es decir intervalos. La *función de distribución acumulada* F (también denotada como CDF, por sus siglas en inglés Cumulative Distribution Function), de una variable aleatoria X está definida para todo número real x mediante:

$$F(x) = P(X \leq x) \quad (2.13)$$

Esto se interpreta como: $F(x)$ es la probabilidad de que la variable aleatoria X tome un valor menor o igual a x . Las funciones de distribución acumuladas tienen las siguientes propiedades:

1. $0 \leq F(x) \leq 1, -\infty < x < \infty$
2. $\lim_{x \rightarrow \infty} F(x) = 1; \lim_{x \rightarrow -\infty} F(x) = 0$
3. La función es no decreciente, es decir, dados $x_1 < x_2$ entonces: $F(x_1) < F(x_2)$
4. La función es continua desde la derecha. Para todo x y $\delta > 0$, $\lim_{\delta \rightarrow 0} [F(x + \delta) - F(x)] = 0$

Dada una variable aleatoria discreta X , la *función de masa de probabilidad* $p(a)$ de X está definida como:

$$p(a) = P(X = a) \quad (2.14)$$

Si X asume uno de los valores x_1, x_2, \dots , entonces:

$$p(x_i) > 0, i=1,2,\dots \quad (2.15)$$

$$p(x) = 0, \text{ para todos los otros valores de } x \quad (2.16)$$

Como X debe tomar uno de los valores x_i , se tiene:

$$\sum_{i=1}^{\infty} p(x_i) = 1 \quad (2.17)$$

La función de distribución acumulada F expresada en términos de $p(x)$ es:

$$F(a) = \sum_{\forall x \leq a} p(x) \quad (2.18)$$

Sea X una variable aleatoria discreta que toma valores x_1, x_2, \dots , y cumple con $x_1 < x_2 < \dots$, entonces su función de distribución F es una *función escalonada*. Esto es, el valor que toma F es constante en los intervalos $(x_i - 1, x_i)$ y después sube un escalón de tamaño $p(x_i)$ en x_i .

Considerando ahora el caso en que la variable aleatoria es continua: sea X una variable aleatoria continua entonces, existe una función no negativa, $f(x)$, definida para todo x real $\in (-\infty, \infty)$, que cumple con la siguiente propiedad para todo conjunto B de números reales:

$$P\{X \in B\} = \int_B f(x) dx \quad (2.19)$$

La función $f(x)$ se denomina *función de densidad de probabilidad* de la variable aleatoria X . Como X debe asumir algún valor, $f(x)$ debe satisfacer:

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x) dx \quad (2.20)$$

Es importante notar que cuando B es $[a, a]$, es decir un punto, se tiene:

$$P\{X = a\} = \int_a^a f(x) dx = 0 \quad (2.21)$$

Esto indica que la probabilidad de que una variable aleatoria continua tome un valor *particular* cualquiera es cero.

La relación entre la función de distribución acumulada F y la función de densidad de probabilidad f se expresa como:

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^a f(x) dx \quad (2.22)$$

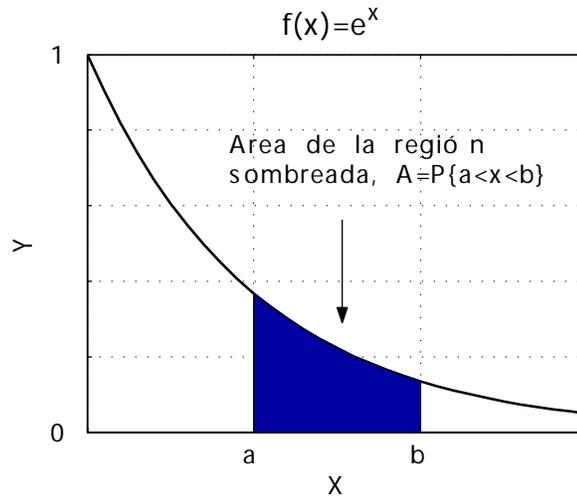


Figura 2.8: Función de densidad de probabilidad, $f(x) = \{e^{-x}, x \geq 0\}$

2.4.2. Variables aleatorias distribuidas conjuntamente

Generalmente, en un experimento dado, se requiere no únicamente la función de distribución de probabilidad de una variable aleatoria individual, sino la relación que existe entre dos o más variables aleatorias. Para determinar esta relación se define la función de distribución de probabilidad acumulada conjunta de X y Y como:

$$F(x, y) = P\{X \leq x, Y \leq y\} \quad (2.23)$$

Un conocimiento de la función de distribución de probabilidad conjunta podría permitir calcular la probabilidad de cualquier relación entre los valores de X y Y .

Para el caso en que X y Y son variables aleatorias discretas, que pueden tomar los valores x_1, x_2, \dots y y_1, y_2, \dots ; se define la *función de masa de probabilidad conjunta* de X y Y mediante:

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\} \quad (2.24)$$

A partir de la función de masa de probabilidad conjunta se pueden determinar las funciones de masas individuales:

$$P\{X = x_i\} = \sum_j p(x_i, y_j) \quad (2.25)$$

$$P\{Y = y_j\} = \sum_i p(x_i, y_j) \quad (2.26)$$

Se dice que X y Y tienen un *conjunto continuo* si existe una función $f(x, y)$ definida para cualquier x, y real con la propiedad de que para todo conjunto C de pares de números reales se tiene:

$$P\{(X, Y) \in C\} = \iint_{(x,y) \in C} f(x, y) dx dy \quad (2.27)$$

A la función $f(x, y)$ se le llama *función de densidad de probabilidad conjunta* de X y Y . La función de densidad de probabilidad de X está definida por:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (2.28)$$

Y la función de densidad de probabilidad de Y por:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad (2.29)$$

2.4.3. Variables aleatorias independientes

Sean X y Y variables aleatorias, se dicen que son *independientes* si para cualesquiera dos conjuntos A y B de números reales:

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\} \quad (2.30)$$

Es decir, X y Y son independientes si para todo A y B , los eventos $E_A = \{X \in A\}$ y $F_B = \{Y \in B\}$ son independientes.

Si X y Y son variables aleatorias discretas, la condición de independencia de la ecuación anterior es equivalente a:

$$p(x, y) = p_x(x)p_y(y), \text{ para toda } x, y \quad (2.31)$$

donde p_x y p_y son las funciones de masa de probabilidad de X y Y .

Si X y Y son variables aleatorias continuas, la condición de independencia es equivalente a:

$$f(x, y) = f_x(x)f_y(y), \text{ para toda } x, y \quad (2.32)$$

En otras palabras, X y Y son independientes si el conocimiento del valor de alguna no modifica la distribución de la otra.

2.4.4. Esperanza matemática

El concepto de esperanza de una variable aleatoria es uno de los más importantes en teoría de la probabilidad. Sea X una variable aleatoria discreta cuyos valores posibles son $x_1 < x_2 < \dots$, la *esperanza* o valor esperado de X , denotado por $E[X]$, se define como:

$$E[X] = \sum_i x_i P\{X = x_i\} \quad (2.33)$$

Se puede observar que el valor esperado de X es una media ponderada de los posibles valores que puede tomar X , cada uno de estos valores ponderado por la probabilidad de que X asuma ese valor. También se puede definir la esperanza para una variable aleatoria continua. Sea X una variable aleatoria continua con función de densidad f . Para un valor pequeño dx

$$f(x)dx \approx P\{x < X < x+dx\} \quad (2.34)$$

el promedio ponderado de todos los valores posibles de X , con los pesos dados a x igual a la probabilidad de que X esté cerca de x , es la integral sobre todas las x de $xf(x)$. Por tanto se puede definir el valor esperado de X mediante:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx \quad (2.35)$$

En este punto es necesario hacer dos observaciones: la primera es que el concepto de esperanza es análogo al concepto físico de centro de gravedad de una distribución de masa, y la segunda que $E[X]$ tiene las mismas unidades que X .

2.4.5. Esperanza de una función de una variable aleatoria

- Si X es una variable aleatoria discreta con función de masa de probabilidad $p(x)$, entonces para cualquier función g con valores reales,

$$E[g(X)] = \sum_x g(x)p(x) \quad (2.36)$$

- Si X es una variable aleatoria continua con función de densidad de probabilidad $f(x)$, entonces para cualquier función g con valores reales,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \quad (2.37)$$

De aquí se puede deducir que:

$$E[aX+b] = aE[X] + b$$

A la esperanza de una variable aleatoria X también se le denomina la *media* o el *primer momento* de X , y a la cantidad $E[X^n]$, $n \geq 1$, el *enésimo momento* de X . Esto se puede resumir como:

$$\begin{aligned} E[X^n] &= \sum_x x^n p(x), \text{ si } X \text{ es discreta} \\ &= \int_{-\infty}^{\infty} x^n f(x)dx, \text{ si } X \text{ es continua} \end{aligned} \quad (2.38)$$

2.4.6. Esperanza de sumas de variables aleatorias

Sean X y Y variables aleatorias, y g una función de dos variables, entonces:

$$\begin{aligned} E[g(X, Y)] &= \sum_y \sum_x g(x, y) p(x, y), \text{ si } X \text{ y } Y \text{ son discretas} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy, \text{ si } X \text{ y } Y \text{ son continuas} \end{aligned} \quad (2.39)$$

De esta definición se puede deducir que:

$$E[X+Y] = E[X] + E[Y] \quad (2.40)$$

2.4.7. Varianza

Sea X una variable aleatoria con media μ , entonces la *varianza* de X , que se denota mediante $\text{Var}(X)$, se escribe como:

$$\text{Var}(X) = E[(X - \mu)^2] \quad (2.41)$$

Desarrollando el término del cuadrado se llega a una forma más sencilla de calcular $\text{Var}(X)$:

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= E[(X^2 - 2\mu X + \mu^2)] \\ &= E[X^2] - E[2\mu X] + E[\mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ \text{Var}(X) &= E[X^2] - \mu^2 = E[X^2] - (E[X])^2 \end{aligned} \quad (2.42)$$

Se le llama *desviación estándar* de X a la cantidad $\sqrt{\text{Var}(X)}$. La desviación estándar viene en las mismas unidades que la media.

2.4.8. Covarianza y correlación

La *covarianza* de dos variables aleatorias X y Y , que se escribe $\text{Cov}(X, Y)$ se define mediante:

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

donde μ_x y μ_y son las medias de X y Y , respectivamente.

Si se expande el término derecho de la ecuación anterior se obtiene una expresión más útil para la covarianza:

$$\begin{aligned}
Cov(X, Y) &= E[(XY - \mu_x Y - \mu_y X + \mu_x \mu_y)] \\
&= E[XY] - \mu_x E[Y] - \mu_y E[X] + \mu_x \mu_y \\
&= E[XY] - \mu_x \mu_y - \mu_y \mu_x + \mu_x \mu_y \\
&= E[XY] - E[X]E[Y]
\end{aligned} \tag{2.43}$$

De lo anterior podemos deducir las siguientes propiedades:

- $Cov(X, Y) = Cov(Y, X)$
- $Cov(X, X) = Var(X)$
- $Cov(aX, Y) = aCov(X, Y)$

La fuerza de la relación entre X y Y , viene expresada por la *correlación* entre X y Y , un valor sin dimensión que se obtiene dividiendo la covarianza entre el producto de las desviaciones estándar de X y Y . Por tanto la correlación es igual a:

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \tag{2.44}$$

2.4.9. Funciones generadoras de momentos

La *función generadora del momento* $\phi(t)$ de una variable aleatoria X está definida para todos los valores de t mediante:

$$\begin{aligned}
\phi(t) &= E[e^{tX}] \\
&= \sum_x e^{tx} p(x), \text{ si } X \text{ es discreta} \\
&= \int_{-\infty}^{\infty} e^{tx} f(x) dx \text{ si } X \text{ es continua}
\end{aligned} \tag{2.45}$$

A $\phi(t)$ se le dice la función generadora de momento debido a que todos los momentos de X se pueden obtener derivando sucesivamente $\phi(t)$. Por ejemplo,

$$\begin{aligned}
\phi'(t) &= \frac{d}{dt} E[e^{tX}] \\
&= E\left[\frac{d}{dt}(e^{tX})\right] \\
&= E[Xe^{tX}]
\end{aligned} \tag{2.46}$$

Por lo tanto,

$$\phi'(0) = E[X] \quad (2.47)$$

De esta forma, la n -ésima derivada de $\phi(t)$ evaluada en $t = 0$ es igual a $E[X^n]$, es decir:

$$\phi^n(0) = E[X^n], n \geq 1 \quad (2.48)$$

Una de las principales propiedades de la función generadora de momentos es que la función generadora de momento de la suma de variables aleatorias independientes es el producto de funciones generadoras de momentos individuales. Esto se puede entender mejor como: si X y Y son independientes y tienen las funciones generadoras de momento $\phi_x(t)$ y $\phi_y(t)$, respectivamente; entonces la función generadora de momento, $\phi_{x+y}(t)$ de $X + Y$ viene dada por:

$$\phi_{x+y}(t) = E[e^{t(X+Y)}] = \phi_x(t)\phi_y(t) \quad (2.49)$$

2.5. Distribución normal

Se dice que una variable aleatoria X tiene una *distribución normal* con media μ ($-\infty < \mu < \infty$) y varianza $\sigma^2 > 0$ si tiene la función de densidad:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2}, \quad -\infty < x < \infty \quad (2.50)$$

La distribución normal se emplea de manera tan amplia que a menudo se recurre a la notación abreviada $X \sim N(\mu, \sigma^2)$ para indicar que la variable aleatoria X se distribuye normalmente con media μ y varianza σ^2 . La densidad normal $f(x)$ es una curva en forma de campana que es simétrica con respecto al eje vertical y que tiene su máximo valor de $1/\sigma\sqrt{(2 * \pi)} \approx 0,399/\sigma$ en $x = \mu$.

2.5.1. Propiedades de la distribución normal

La distribución normal tiene varias propiedades importantes:

1. $\int_{-\infty}^{\infty} f(x); dx = 1$; requisito en todas las funciones de densidad.
2. $f(x) \geq 0$, para todo x ; requisito en todas las funciones de densidad.
3. $\lim_{x \rightarrow \infty} f(x) = 0$ y $\lim_{x \rightarrow -\infty} f(x) = 0$
4. $f[(x + \mu)] = f[-(x - \mu)]$. La densidad es simétrica alrededor de μ .
5. El valor máximo de f ocurre en $x = \mu$.
6. Los puntos de inflexión de f están en $x = \mu \pm \sigma$.

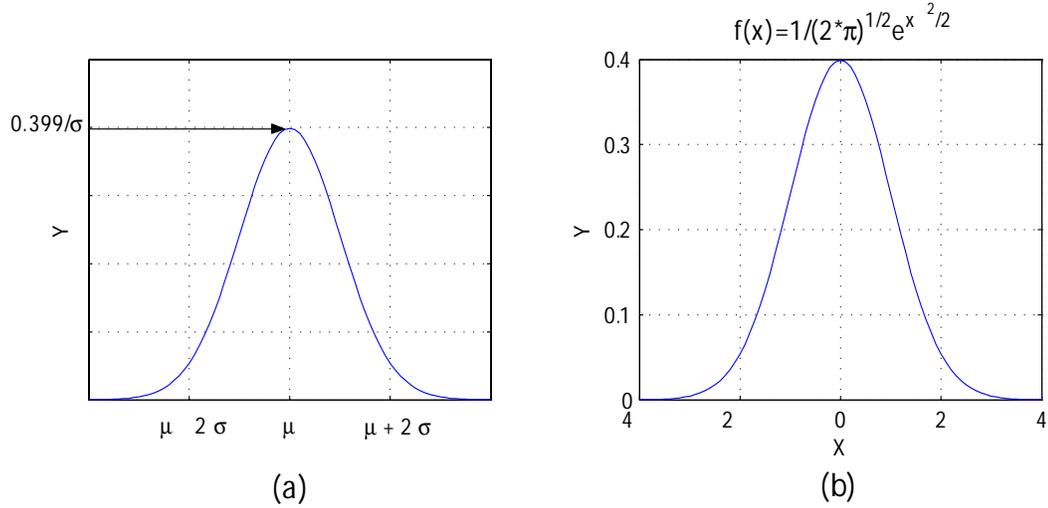


Figura 2.9: Función de densidad normal con: (a) μ y σ arbitrarias, (b) $\mu = 0$ y $\sigma = 1$

2.5.2. Media y varianza de la distribución normal

La función generadora de momento de una variable aleatoria normal con parámetros μ y σ^2 se obtiene de la siguiente manera:

$$\begin{aligned}
 \phi(t) &= E[e^{tX}] \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-(x-\mu)^2/2\sigma^2} dx, \text{ haciendo } y = \frac{x-\mu}{\sigma} \\
 &= \frac{1}{\sqrt{2\pi}} e^{\mu t} \int_{-\infty}^{\infty} e^{t\sigma y} e^{-y^2/2} dy \\
 &= \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\left[\frac{y^2 - 2t\sigma y}{2}\right]\right\} dy \\
 &= \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(y-t\sigma)^2}{2} + \frac{t^2\sigma^2}{2}\right\} dy \\
 &= \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y-t\sigma)^2/2} dy \\
 &= \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}
 \end{aligned} \tag{2.51}$$

debido a que la integral de

$$\frac{1}{\sqrt{2\pi}} e^{-(y-t\sigma)^2/2}$$

es igual a 1 por ser la densidad de una variable aleatoria normal (con parámetros $t\sigma$ y 1).

Si se deriva $\phi(t)$ se obtiene:

$$\begin{aligned}\phi'(t) &= (\mu + t\sigma^2) \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\} \\ \phi''(t) &= \sigma^2 \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\} + \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\} (\mu + t\sigma^2)^2\end{aligned}$$

Evaluando para $t = 0$ se obtiene:

$$\begin{aligned}E[X] &= \phi'(0) = \mu \\ E[X^2] &= \phi''(0) = \sigma^2 + \mu^2\end{aligned}$$

así se obtiene:

$$E[X] = \mu \tag{2.52}$$

$$Var(X) = E[X^2] - (E[X])^2 = \sigma^2 \tag{2.53}$$

De manera que μ y σ^2 representan la media y la varianza de la distribución, respectivamente. También es de hacer notar que si X es normal con media μ y varianza σ^2 , entonces $Y = \alpha X + \beta$ es normal con media $\alpha\mu + \beta$ y varianza $\alpha^2\sigma^2$.

2.5.3. Distribución normal estándar

Dado $X \sim N(\mu, \sigma^2)$, entonces

$$Z = \frac{X - \mu}{\sigma} \tag{2.54}$$

es una variable aleatoria con media 0 y varianza 1; esto es $Z \sim N(0, 1)$, entonces se dice que Z tiene una distribución normal estándar. La distribución correspondiente es Φ , donde

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \quad -\infty < x < \infty \tag{2.55}$$

Debido a la simetría de la función de densidad de probabilidad normal estándar con respecto a cero puede decirse que:

$$\begin{aligned}\Phi(-x) &= P\{Z < -x\} \\ &= P\{Z > x\} \\ &= 1 - \Phi(x)\end{aligned} \tag{2.56}$$

Para $\alpha \in (0, 1)$, sea z_α de manera que:

$$P\{Z > z_\alpha\} = 1 - \Phi(z_\alpha) = \alpha \quad (2.57)$$

Esto quiere decir que la probabilidad de que una variable aleatoria normal estándar sea mayor a z_α es igual a α .

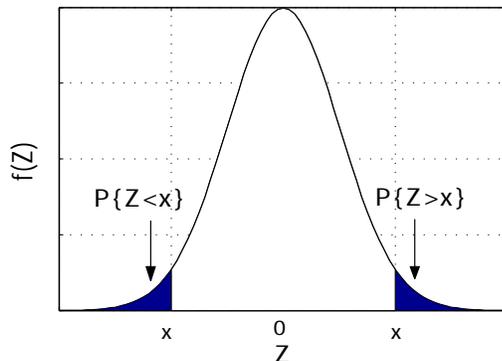


Figura 2.10: Función de densidad normal estándar

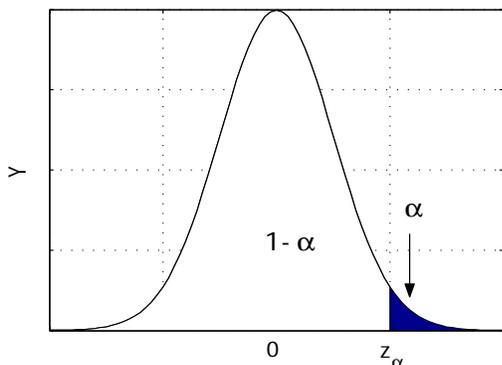


Figura 2.11: Función de densidad normal estándar, $P\{Z > Z_\alpha\} = \alpha$

2.6. Distribución Gamma

Dada una variable aleatoria con parámetros (α, λ) , $\lambda > 0, \alpha > 0$, tiene una distribución gamma si su función de densidad es:

$$f(x) = \begin{cases} (\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}) / (\Gamma(\alpha)) & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2.58)$$

Donde:

$$\begin{aligned}\Gamma(\alpha) &= \int_0^{\infty} \lambda e^{-\lambda x} (\lambda x)^{\alpha-1} dx \\ &= \int_0^{\infty} e^{-y} y^{\alpha-1} dy, \text{ ya que } y = \lambda x\end{aligned}\quad (2.59)$$

Integrando por partes con $u = y^{\alpha-1}$, $dv = e^{-y}$, $v = -e^{-y}$, se aplica $\int u dv = uv - \int v du$, con $\alpha > 1$:

$$\begin{aligned}\int_0^{\infty} e^{-y} y^{\alpha-1} dy &= -e^{-y} y^{\alpha-1} \Big|_{y=0}^{y=\infty} + \int_0^{\infty} e^{-y} (\alpha-1) y^{\alpha-2} dy \\ &= (\alpha-1) \int_0^{\infty} e^{-y} y^{\alpha-2} dy\end{aligned}\quad (2.60)$$

Esto se expresa mejor como:

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)\quad (2.61)$$

Deduciendo para un $\alpha = n$, se tiene que:

$$\begin{aligned}\Gamma(n) &= (n-1)\Gamma(n-1) \\ &= (n-1)(n-2)\Gamma(n-2) \\ &= (n-1)(n-2)(n-3)\Gamma(n-3) \\ &= \vdots \\ &= (n-1)! \Gamma(1)\end{aligned}$$

dado que $\Gamma(1) = \int_0^{\infty} e^{-y} dy = 1$ entonces se deduce que:

$$\Gamma(n) = (n-1)!\quad (2.62)$$

A esta función se le llama función *gamma*. La función generadora de momento de una variable aleatoria gamma X con parámetros (α, λ) es:

$$\phi(t) = E[e^{tX}] = \left(\frac{\lambda}{\lambda-t} \right)^{\alpha}\quad (2.63)$$

Diferenciando $\phi(t)$, se tiene que:

$$\begin{aligned}\phi'(t) &= \frac{\alpha \lambda^{\alpha}}{(\lambda-t)^{\alpha+1}} \\ \phi''(t) &= \frac{\alpha(\alpha+1)\lambda^{\alpha}}{(\lambda-t)^{\alpha+2}}\end{aligned}$$

De donde:

$$E[X] = \phi'(0) = \frac{\alpha}{\lambda} \quad (2.64)$$

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 = \phi''(0) - \left(\frac{\alpha}{\lambda}\right)^2 \\ \text{Var}(X) &= \frac{\alpha}{\lambda^2} \end{aligned} \quad (2.65)$$

2.7. Distribuciones que se derivan de la normal

2.7.1. La distribución chi cuadrada (χ^2)

Sean Z_1, Z_2, \dots, Z_n variables aleatorias normales estándar independientes, entonces se dice que si X es:

$$X = Z_1^2, Z_2^2, \dots, Z_n^2 \quad (2.66)$$

entonces tiene una *distribución chi cuadrada con n grados de libertad*. También se puede escribir como $X \sim \mathcal{X}_n^2$. Si X es una variable aleatoria chi cuadrada con n grados de libertad, entonces para toda $\alpha \in (0, 1)$, el valor $\mathcal{X}_{\alpha, n}^2$ está definida así:

$$P\{X \geq \mathcal{X}_{\alpha, n}^2\} = \alpha \quad (2.67)$$

La función generadora de momento de una variable aleatoria chi cuadrada con n grados de libertad, para $n=1$ es:

$$\begin{aligned} E[e^{tX}] &= E[e^{tZ}], \text{ con } Z \sim N(0, 1) \\ &= \int_{-\infty}^{\infty} e^{tx^2} f_Z(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2(1-2t)/2} dx \\ &= (1-2t)^{-1/2} \frac{1}{\bar{\sigma}\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2\bar{\sigma}^2} dx, \text{ con } \bar{\sigma}^2 = (1-2t)^{-1} \\ &= (1-2t)^{-1/2} \end{aligned}$$

ya que la integral de la densidad normal $(0, \bar{\sigma}^2)$ es:

$$\frac{1}{\bar{\sigma}\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2\bar{\sigma}^2} dx = 1$$

Para el caso general de n grados de libertad se tiene:

$$E[e^{tX}] = (1 - 2t)^{-n/2} \quad (2.68)$$

Se puede notar que $(1 - 2t)^{-n/2}$ es la función generadora de momento de una variable aleatoria gamma con parámetros $(n/2, 1/2)$. De esto se deduce que la densidad de X para una distribución chi cuadrada con n grados de libertad está dada por:

$$f(x) = \frac{1/2 e^{-x/2} (x/2)^{(n/2-1)}}{\Gamma(n/2)}, x > 0 \quad (2.69)$$

Ya que la distribución chi cuadrada con n grados de libertad es igual a la distribución gamma con parámetros $\alpha = n/2$ y $\lambda = 1/2$, entonces la media y la varianza de una variable aleatoria X con esta distribución es:

$$E[X] = n \quad (2.70)$$

$$Var(X) = 2n \quad (2.71)$$

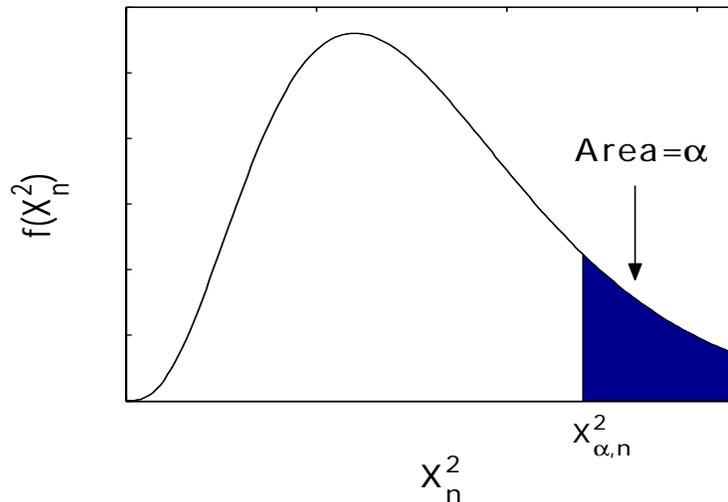


Figura 2.12: Función de densidad chi cuadrada con 8 grados de libertad

2.7.2. La distribución t

Sean Z y \mathcal{X}_n^2 variables aleatorias independientes, Z con distribución normal estándar y \mathcal{X}_n^2 con distribución chi cuadrada con n grados de libertad, entonces la variable aleatoria T_n definida por:

$$T_n = \frac{Z}{\sqrt{\mathcal{X}_n^2/n}} \quad (2.72)$$

tiene *distribución t con n grados de libertad*. La media y la varianza de T_n vienen dadas por:

$$\begin{aligned} E[T_n] &= 0, n > 1 \\ \text{Var}(T_n) &= \frac{n}{n-2}, n > 2 \end{aligned} \quad (2.73)$$

Así la varianza de T_n decrece hacia 1, al igual que la varianza de una variable aleatoria normal estándar, conforme n tiende a ∞ . Para α , con $0 < \alpha < 1$, sea $t_{\alpha,n}$ tal que:

$$P\{T_n \geq t_{\alpha,n}\} = \alpha \quad (2.74)$$

Debido a la simetría respecto a cero de la función de densidad se puede decir que:

$$-t_{\alpha,n} = t_{1-\alpha,n}$$

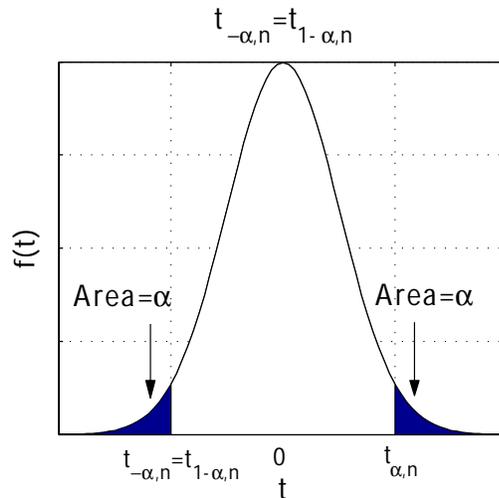


Figura 2.13: Distribución t, $-t_{\alpha,n} = t_{1-\alpha,n}$

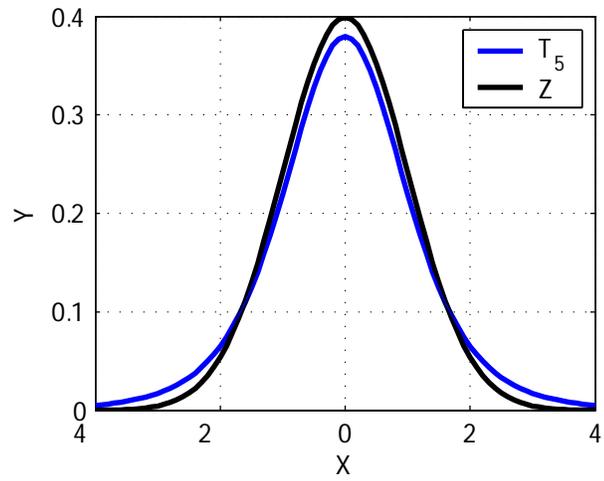


Figura 2.14: Comparación de la función de densidad normal estándar con la densidad T_5

2.7.3. La distribución F

Sean \mathcal{X}_n^2 y \mathcal{X}_m^2 variables aleatorias chi cuadrada independientes con n y m grados de libertad, respectivamente, entonces se dice que $F_{n,m}$ es una variable aleatoria definida por:

$$F_{n,m} = \frac{\mathcal{X}_n^2/n}{\mathcal{X}_m^2/m} \quad (2.75)$$

tiene una *distribución F* con n y m grados de libertad. Para $\alpha \in (0, 1)$, sea $F_{\alpha,n,m}$, tal que:

$$P\{F_{n,m} > F_{\alpha,n,m}\} = \alpha \quad (2.76)$$

También es importante notar que:

$$1/F_{\alpha,n,m} = F_{1-\alpha,n,m}$$

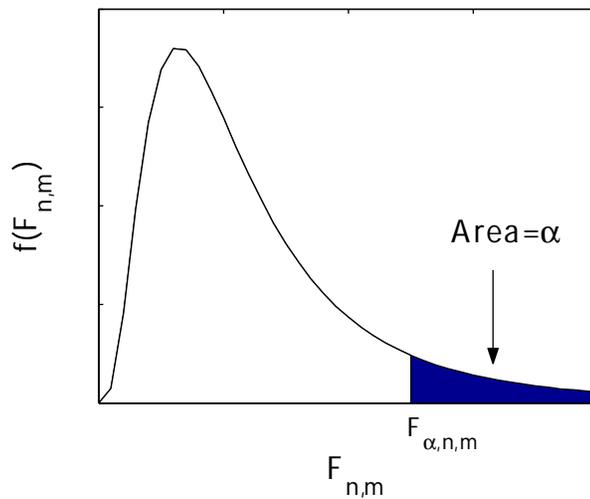


Figura 2.15: Distribución F

2.8. Distribuciones de estadísticos muestrales

Es muy importante tener el concepto de *estadístico* el cual se refiere a una variable aleatoria cuyos valores se determinan a partir de los datos muestrales. Dos de los más importantes estadísticos son la media muestral y la varianza muestral.

2.8.1. La media muestral

Sea una población dada cuyos valores provienen de una variable aleatoria con media μ y varianza σ^2 . A estas cantidades de μ y σ^2 se les llama *media poblacional* y *varianza poblacional*, respectivamente. Si X_1, X_2, \dots, X_n son una muestra de valores de esta población entonces la media muestral se define como:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (2.77)$$

debido a que la media muestral \bar{X} proviene de los valores de la variable aleatoria en la muestra, se puede decir que \bar{X} es también una variable aleatoria. El valor esperado y la varianza son:

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] \\ &= \frac{1}{n}(E[X_1] + \dots + E[X_n]) \\ &= \mu \end{aligned} \quad (2.78)$$

y

$$\begin{aligned} Var(X) &= Var\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) \\ &= \frac{1}{n^2}[Var(X_1) + \dots + Var(X_n)] \\ &= \frac{n\sigma^2}{n^2} \\ &= \frac{\sigma^2}{n} \end{aligned} \quad (2.79)$$

2.8.2. El teorema del límite central

Si X_1, X_2, \dots, X_n son una sucesión de variables aleatorias independientes e idénticamente distribuidas con media μ y varianza σ^2 , entonces para un valor de

n grande la distribución de $X_1 + X_2 + \dots + X_n$ es aproximadamente normal con media $n\mu$ y varianza $n\sigma^2$. Se puede decir que:

$$\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \quad (2.80)$$

es aproximadamente una variable aleatoria normal estándar, con n grande:

$$P\left\{\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} < x\right\} \approx P\{Z < x\} \quad (2.81)$$

donde Z es una variable aleatoria normal estándar.

2.8.3. Distribución aproximada de la media muestral

Si X_1, X_2, \dots, X_n son una muestra de una población con media μ y varianza σ^2 , el teorema del límite central sirve para aproximar la distribución de la media muestral. Como la media muestral tiene valor esperado μ y desviación estándar σ/\sqrt{n} , entonces se dice que:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (2.82)$$

tiene aproximadamente una distribución normal estándar. Ahora existe el problema de que tan grande debe ser el tamaño n de la muestra para que la aproximación sea válida. En general se puede asumir una aproximación normal siempre y cuando el tamaño n de la muestra sea de por lo menos 30.

2.8.4. La varianza muestral

Sean X_1, X_2, \dots, X_n una muestra de una población con media μ y varianza σ^2 , además con media muestral \bar{X} , entonces al estadístico S^2 definido por:

$$S^2 = \left(\sum_{i=1}^n (X_i - \bar{X})^2\right) / (n - 1) \quad (2.83)$$

se le llama *varianza muestral*, y a $S = \sqrt{S^2}$ se le denomina *desviación estándar muestral*. Recordando que

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

de manera que $\bar{x} = \sum_{i=1}^n x_i/n$. Sustituyendo en la ecuación de S^2 se tiene:

$$(n-1)S^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$$

recordando que para una variable aleatoria Q , $E[Q^2] = Var(Q) + (E[Q])^2$, y tomando la esperanza a ambos lados de la ecuación anterior:

$$\begin{aligned} (n-1)E[S^2] &= E\left[\sum_{i=1}^n X_i^2\right] - nE[\bar{X}]^2 \\ &= nE[X_1^2] - nE[\bar{X}]^2 \\ &= nVar(X_1) + n(E[X_1])^2 - nVar(\bar{X}) - n(E[\bar{X}])^2 \\ &= n\sigma^2 + n\mu^2 - n(\sigma^2/n) - n\mu^2 \\ &= (n-1)\sigma^2 \end{aligned}$$

por tanto,

$$E[S^2] = \sigma^2 \tag{2.84}$$

Esto quiere decir que el valor esperado de la varianza muestral S^2 es igual a la varianza poblacional.

En resumen, sea X_1, X_2, \dots, X_n una muestra de una población normal con media μ y varianza σ^2 , entonces como la suma de variables aleatorias normales independientes tiene una distribución normal, entonces \bar{X} es normal con media

$$E[\bar{X}] = \mu \tag{2.85}$$

y varianza

$$Var(X) = \frac{\sigma^2}{n} \tag{2.86}$$

Entonces, \bar{X} , el promedio de la muestra, es normal con una media igual a la media poblacional y con una varianza reducida por un factor de $1/n$. Por tanto se puede decir que:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{2.87}$$

es una variable aleatoria normal estándar. También es de hacer notar que si X_1, X_2, \dots, X_n es una muestra de una población normal con media μ y varianza σ^2 , entonces \bar{X} y S^2 son variables aleatorias independientes, siendo \bar{X} normal con media μ y varianza σ^2/n y siendo $(n-1)S^2/\sigma^2$ chi cuadrada con $n-1$ grados de libertad. También es importante decir que: Sea X_1, X_2, \dots, X_n una muestra de

una población normal con media μ . Si \bar{X} es la media muestral y S la desviación muestral, entonces:

$$\sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{n-1} \quad (2.88)$$

Es decir, $\sqrt{n}(\bar{X} - \mu)/S$ tiene una distribución t con $n-1$ grados de libertad.

2.9. Estimación de parámetros e intervalos de confianza

2.9.1. Estimación de máxima verosimilitud

Primero es necesario definir algunos conceptos, a todo estadístico que se emplee para estimar el valor de un parámetro desconocido se conoce como un *estimador* de θ , y al valor obtenido se le llama *estimado*. El *estimador de máxima verosimilitud* es aquel valor de θ que maximiza $f(x_1, x_2, \dots, x_n|\theta)$ donde x_1, x_2, \dots, x_n son los valores observados. La función $f(x_1, x_2, \dots, x_n|\theta)$ se le denomina *función de verosimilitud* de θ . Para obtener el valor maximizante de θ , se utiliza el hecho de que $f(x_1, x_2, \dots, x_n|\theta)$ y $\log[f(x_1, x_2, \dots, x_n|\theta)]$ tiene sus máximos en el mismo valor de θ . De esta manera se obtiene $\hat{\theta}$ maximizando $\log[f(x_1, x_2, \dots, x_n|\theta)]$.

Sean X_1, X_2, \dots, X_n variables aleatorias normales, independientes, cada una con media desconocida μ y desviación estándar desconocida σ . La densidad conjunta está dada por:

$$\begin{aligned} f(x_1, x_2, \dots, x_n|\mu, \sigma) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2}\right] \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^{n/2} \frac{1}{\sigma^n} \exp\left[-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right] \end{aligned} \quad (2.89)$$

y el logaritmo de verosimilitud por:

$$\log f(x_1, x_2, \dots, x_n|\mu, \sigma) = -\frac{n}{2} \log(2\pi) - n \log(\sigma) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \quad (2.90)$$

Derivando parcialmente con respecto a μ y σ e igualando a cero se obtienen:

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\hat{\sigma} = \left(\frac{\sum_{i=1}^n (x_i - \hat{\mu})^2}{n} \right)^{1/2}$$

Entonces los estimadores de máxima verosimilitud de μ y σ vienen dados, respectivamente, por:

$$\bar{X} \quad \text{y} \quad \left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \right)^{1/2} \quad (2.91)$$

Es de hacer notar que el estimador de máxima verosimilitud de la desviación estándar σ difiere de la desviación estándar muestral en que en el denominador del estimador de máxima verosimilitud se encuentra \sqrt{n} en vez de $\sqrt{n-1}$.

2.9.2. Intervalos de confianza

Sean X_1, X_2, \dots, X_n una muestra de una población normal con media desconocida μ y desviación estándar conocida σ . En ocasiones es preferible, más que un estimado puntual de μ , resulta mucho más valioso tener un intervalo del que se tenga cierto grado de confianza de que μ esté dentro de él.

Intervalo de confianza para una media normal cuando se conoce la varianza

Recordando que el estimador puntual \bar{X} es normal con media μ y desviación estándar σ^2/n , entonces

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n} \frac{(\bar{X} - \mu)}{\sigma} \quad (2.92)$$

tiene una distribución normal estándar. Entonces, recordando z_α es tal que $P\{Z > z_\alpha\} = \alpha$, siendo Z una variable aleatoria normal. Esto implica que para toda α se tiene:

$$P\{-z_{\alpha/2} < Z < z_{\alpha/2}\} = 1 - \alpha \quad (2.93)$$

Por lo tanto,

$$\begin{aligned} P\{-z_{\alpha/2} < \sqrt{n} \frac{(\bar{X} - \mu)}{\sigma} < z_{\alpha/2}\} &= 1 - \alpha \\ P\{-z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\} &= 1 - \alpha \\ P\{-z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu - \bar{X} < z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\} &= 1 - \alpha \\ P\{\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\} &= 1 - \alpha \end{aligned} \quad (2.94)$$

Entonces un intervalo de confianza bilateral del $100(1-\alpha)\%$ de confianza para μ es:

$$\left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \quad (2.95)$$

donde \bar{x} es la media muestral obtenida.

También se pueden definir intervalos de confianza unilateral, recordando que $P\{Z > z_\alpha\} = \alpha$ y $P\{Z < -z_\alpha\} = \alpha$. Así se obtienen:

$$\left(\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty \right) \quad (2.96)$$

$$\left(-\infty, \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}} \right) \quad (2.97)$$

siendo intervalos de confianza de $100(1-\alpha)\%$ unilateral superior y de confianza de $100(1-\alpha)\%$ unilateral inferior para μ , respectivamente.

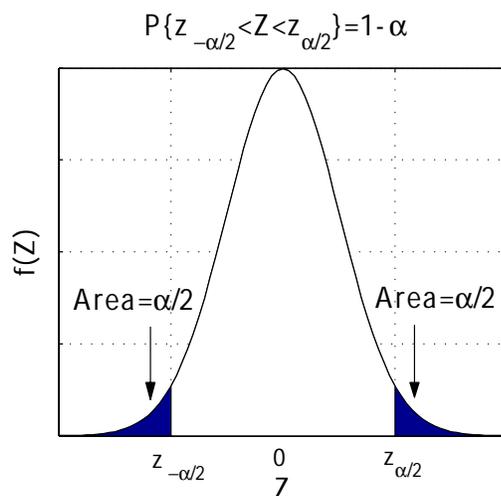


Figura 2.16: Intervalos de confianza, σ^2 conocida

Intervalo de confianza para una media normal cuando no se conoce la varianza

Recordando que $\sqrt{n} \frac{(\bar{X} - \mu)}{S}$ es una variable aleatoria t con $n-1$ grados de libertad, entonces para toda $\alpha \in (0, 1/2)$

$$P\{-t_{\alpha/2, n-1} < \sqrt{n} \frac{(\bar{X} - \mu)}{S} < t_{\alpha/2, n-1}\} = 1 - \alpha$$

$$P\left\{\bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}\right\} = 1 - \alpha$$

como $\bar{X} = \bar{x}$ y $S=s$, podemos decir que con $100(1 - \alpha)\%$ de confianza

$$\mu \in \left(\bar{x} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}\right) \quad (2.98)$$

Los intervalos de confianza son:

$$\mu \in \left(\bar{x} - t_{\alpha, n-1} \frac{s}{\sqrt{n}}, \infty\right) \quad (2.99)$$

$$\mu \in \left(-\infty, \bar{x} + t_{\alpha, n-1} \frac{s}{\sqrt{n}}\right) \quad (2.100)$$

de $100(1 - \alpha)\%$ unilateral superior y de $100(1 - \alpha)\%$ unilateral inferior, respectivamente.

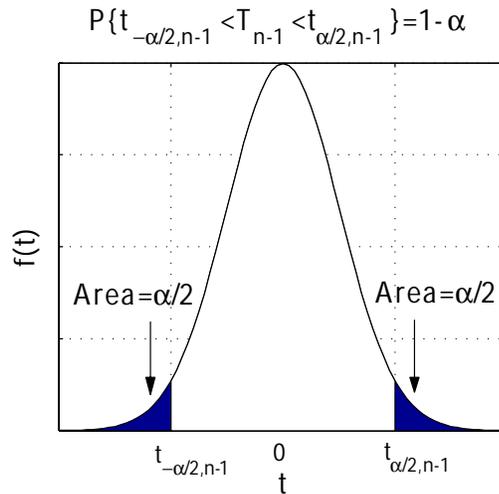


Figura 2.17: Intervalo de confianza, σ^2 desconocida

Intervalo de confianza para la varianza de una distribución normal

Sean X_1, X_2, \dots, X_n una muestra de una población normal con media desconocida μ y desviación estándar desconocida σ y considerando que

$$(n - 1)S^2/\sigma^2 \sim \chi_{n-1}^2$$

Entonces,

$$P \left\{ \chi_{1-\alpha/2, n-1}^2 \leq (n-1)S^2/\sigma^2 \leq \chi_{\alpha/2, n-1}^2 \right\} = 1 - \alpha$$

$$P \left\{ \frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2} \right\} = 1 - \alpha$$

Suponiendo que $S^2 = s^2$, entonces un intervalo de confianza de $100(1 - \alpha) \%$ para σ^2 es

$$\left(\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} \right) \quad (2.101)$$

Los intervalos de confianza son:

$$\left(0, \frac{(n-1)s^2}{\chi_{\alpha, n-1}^2} \right) \quad (2.102)$$

$$\left(\frac{(n-1)s^2}{\chi_{1-\alpha, n-1}^2}, \infty \right) \quad (2.103)$$

de $100(1 - \alpha) \%$ unilateral inferior y de $100(1 - \alpha) \%$ unilateral superior, respectivamente.

2.10. Pruebas de hipótesis

Una *hipótesis estadística* es una afirmación acerca de un conjunto de parámetros de la distribución poblacional en base a observaciones hechas a una muestra aleatoria obtenida de la población, y se le denomina hipótesis por que no se sabe si es verdadera o no. Hay que hacer notar que al aceptar una hipótesis no se está diciendo que sea verdadera sino más bien que los resultados obtenidos parecen ser consistente con la hipótesis.

2.10.1. Nivel de significancia

Sea una población con una distribución F_θ , con θ desconocida, y se desea comprobar una hipótesis acerca de θ . A estas hipótesis se le llaman *hipótesis nulas* y se denotan por H_θ . Para decidir si se acepta la hipótesis nula H_o , se observará una muestra poblacional de tamaño n (X_1, \dots, X_n). También se definirá una región C en el espacio n -dimensional, la hipótesis se rechaza si la muestra aleatoria, X_1, \dots, X_n se encuentra en C , y se acepta si no está en C . A esta región se le denomina *región crítica*. Esto se puede expresar de la siguiente manera:

$$\begin{array}{ll} \text{se acepta} & H_o \text{ si } (X_1, X_2, \dots, X_n) \notin C \\ \text{se rechaza} & H_o \text{ si } (X_1, X_2, \dots, X_n) \in C \end{array} \quad (2.104)$$

La decisión de aceptar o rechazar la hipótesis nula H_o se basa en un estadístico calculado a partir de los datos en la muestra poblacional. Como se está tomando una decisión a partir de información de una muestra aleatoria, esta decisión podría estar sujeta a errores. Entonces se pueden definir dos tipos de errores: si la hipótesis nula se rechaza cuando en realidad es verdadera se dice que se ha cometido *error del tipo I*. Si la hipótesis nula se acepta cuando es falsa, entonces el error que se comete es del *tipo II*. Para evitar que se cometa errores del tipo I se asigna un valor α y se exige que la prueba tenga la propiedad de que H_o sea siempre verdadera, es decir, la probabilidad de que H_o sea rechazada no sea mayor a α . La probabilidad α del error tipo I se le llama *nivel de significancia* de la prueba.

2.10.2. Pruebas de hipótesis sobre la media de una población normal

Cuando se conoce la varianza

Sean X_1, \dots, X_n una muestra de tamaño n con media desconocida μ y varianza conocida σ^2 , además sea μ_o una constante dada, se desea probar la hipótesis nula

$$H_o : \mu = \mu_o \quad (2.105)$$

contra la hipótesis alternativa H_1

$$H_1 : \mu \neq \mu_o \quad (2.106)$$

Debido a que \bar{X} es un estimado puntual de μ entonces sería lógico aceptar H_o si \bar{X} no está muy lejos de μ_o . La región crítica quedaría como

$$C = \{X_1, \dots, X_n : |\bar{X} - \mu_o| > c\} \quad (2.107)$$

para algún valor de c adecuado. También se quiere que la prueba tenga un nivel de significancia α , entonces c también debe satisfacer

$$P_{\mu_o}\{|\bar{X} - \mu_o| > c\} = \alpha \quad (2.108)$$

donde P_{μ_o} indica la probabilidad de que $\mu = \mu_o$. Si $\mu = \mu_o$, \bar{X} tendrá distribución normal con media μ_o y varianza σ^2/n , entonces podemos definir Z como

$$Z \equiv \frac{\bar{X} - \mu_o}{\sigma/\sqrt{n}} \quad (2.109)$$

que tiene una distribución normal estándar. Por lo tanto,

$$\begin{aligned} P\left\{|Z| > \frac{c\sqrt{n}}{\sigma}\right\} &= \alpha \\ 2P\left\{Z > \frac{c\sqrt{n}}{\sigma}\right\} &= \alpha \end{aligned}$$

y recordando que $P\{Z > z_{\alpha/2}\} = \alpha/2$, se tiene que

$$\begin{aligned}\frac{c\sqrt{n}}{\sigma} &= z_{\alpha/2} \\ c &= \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\end{aligned}$$

Por tanto, la prueba para un nivel de significancia α va a rechazar la hipótesis nula H_o si $|\bar{X} - \mu_o| > z_{\alpha/2}\sigma/\sqrt{n}$ y se acepta si no se cumple. Esto se puede expresar como

$$\text{se acepta } H_o \text{ si } \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_o| > z_{\alpha/2} \quad (2.110)$$

$$\text{se rechaza } H_o \text{ si } \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_o| \leq z_{\alpha/2} \quad (2.111)$$

También se puede entender de la siguiente manera: para todo valor observado del estadístico de prueba $\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_o|$ (denotándolo por v), la prueba dice que se rechaza la hipótesis nula si la probabilidad de que el estadístico de prueba alcance un valor tan grande como v siendo H_o verdadera, es menor o igual al nivel de significancia α . Para determinar si se acepta o no la hipótesis nula, primero se calcula el estadístico de prueba, y luego la probabilidad de que la normal unitaria sea mayor (en valor absoluto) a esa cantidad. A esta probabilidad se le denomina *valor p*, y representa el nivel de significación más pequeño que conllevaría al rechazo. En general, si el valor p es menor o igual a α , se rechaza H_o , y si el valor p es mayor a α se aceptaría H_o .

De una manera similar a la ya expuesta aquí se puede deducir las *pruebas unilaterales*, las cuales se definen como aquellas pruebas en que la hipótesis alternativa está dada por:

$$H_o : \mu = \mu_o \text{ contra } H_1 : \mu > \mu_o \text{ o } H_1 : \mu < \mu_o$$

a diferencia de una prueba bilateral que en que la hipótesis alternativa es

$$H_1 : \mu \neq \mu_o$$

De esta forma, para la prueba de hipótesis

$$H_o : \mu = \mu_o \text{ contra } H_1 : \mu > \mu_o$$

se rechaza H_o si $\bar{X} - \mu_o > z_{\alpha}\sigma/\sqrt{n}$, y de lo contrario se acepta. Expresándolo en la forma usual

$$\begin{aligned}\text{se acepta } H_o &\text{ si } \frac{\sqrt{n}}{\sigma}[\bar{X} - \mu_o] \leq z_{\alpha} \\ \text{se rechaza } H_o &\text{ si } \frac{\sqrt{n}}{\sigma}[\bar{X} - \mu_o] > z_{\alpha}\end{aligned} \quad (2.112)$$

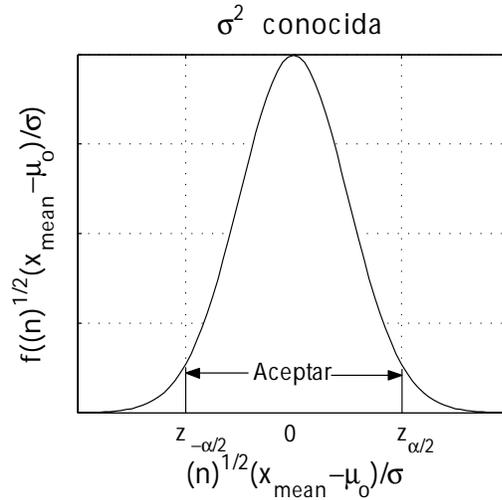


Figura 2.18: Prueba de hipótesis con σ^2 conocida

Cuando no se conoce la varianza

Como no se conoce la varianza se estima mediante

$$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n - 1)$$

y recordando que,

$$T = \frac{\bar{X} - \mu_o}{S/\sqrt{n}}$$

tiene, cuando $\mu = \mu_o$, una distribución t con $n-1$ grados de libertad. De aquí, entonces:

$$P\{-t_{\alpha/2, n-1} \leq \frac{\bar{X} - \mu_o}{S/\sqrt{n}} \leq t_{\alpha/2, n-1}\} = 1 - \alpha$$

donde $t_{\alpha/2, n-1}$ representa el valor del percentil superior $100 \alpha/2$ de la distribución t con $n-1$ grados de libertad. Entonces la prueba de nivel de significancia α queda como:

$$H_o : \mu = \mu_o \text{ contra } H_1 : \mu \neq \mu_o$$

$$\text{se acepta } H_o \text{ si } \left| \frac{\sqrt{n}}{S} [\bar{X} - \mu_o] \right| \leq t_{\alpha/2, n-1} \quad (2.113)$$

$$\text{se rechaza } H_o \text{ si } \left| \frac{\sqrt{n}}{S} [\bar{X} - \mu_o] \right| > t_{\alpha/2, n-1} \quad (2.114)$$

El *valor p* de la prueba es la probabilidad de que $|T|$ sea mayor a $|t|$ cuando H_o es verdadera. Por lo tanto, H_o se rechaza para todo nivel de significancia mayor que el *valor p* y se acepta para los niveles de significancia menores.

Las pruebas unilaterales para esta prueba son:

$$H_o : \mu = \mu_o \text{ o } H_o : \mu \leq \mu_o \text{ contra } H_1 : \mu > \mu_o$$

$$\text{se acepta } H_o \text{ si } \frac{\sqrt{n}}{S}[\bar{X} - \mu_o] \leq t_{\alpha, n-1} \quad (2.115)$$

$$\text{se rechaza } H_o \text{ si } \frac{\sqrt{n}}{S}[\bar{X} - \mu_o] > t_{\alpha, n-1} \quad (2.116)$$

El valor p de la prueba es la probabilidad de que una variable aleatoria t con $n-1$ grados de libertad sea al menos tan grande como $\frac{\sqrt{n}}{S}[\bar{X} - \mu_o]$.

$$H_o : \mu = \mu_o \text{ o } H_o : \mu \geq \mu_o \text{ contra } H_1 : \mu < \mu_o$$

$$\text{se acepta } H_o \text{ si } \frac{\sqrt{n}}{S}[\bar{X} - \mu_o] \geq -t_{\alpha, n-1} \quad (2.117)$$

$$\text{se rechaza } H_o \text{ si } \frac{\sqrt{n}}{S}[\bar{X} - \mu_o] < -t_{\alpha, n-1} \quad (2.118)$$

El valor p de la prueba es la probabilidad de que una variable aleatoria t con $n-1$ grados de libertad sea menor o igual que $\frac{\sqrt{n}}{S}[\bar{X} - \mu_o]$.

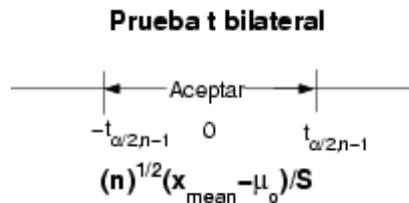


Figura 2.19: Distribución t , $-t_{\alpha, n} = t_{1-\alpha, n}$

2.10.3. Pruebas de hipótesis sobre la varianza de una población normal

Sea X_1, X_2, \dots, X_n una muestra de una población normal con media desconocida μ y varianza desconocida σ^2 , y se desea probar la hipótesis de que

$$H_o : \sigma^2 = \sigma_o^2 \text{ contra } H_1 : \sigma^2 \neq \sigma_o^2$$

para un valor dado de σ_o^2 .

Recordando que $(n-1)S^2/\sigma^2$ tiene una distribución chi cuadrada con $n-1$ grados de libertad. Entonces si H_o es verdadera,

$$\frac{(n-1)S^2}{\sigma_o^2} \sim \chi_{n-1}^2$$

y entonces

$$P\{\chi_{1-\alpha/2, n-1}^2 \leq \frac{(n-1)S^2}{\sigma_o^2} \leq \chi_{\alpha/2, n-1}^2\} = 1 - \alpha$$

por lo tanto, para un nivel de significancia de α la prueba es,

$$\text{se acepta } H_o \text{ si } \chi_{1-\alpha/2, n-1}^2 \leq \frac{(n-1)S^2}{\sigma_o^2} \leq \chi_{\alpha/2, n-1}^2 \quad (2.119)$$

$$\text{se rechaza } H_o \text{ de otra manera} \quad (2.120)$$

El valor de p para los datos viene dado por:

$$\text{valor p} = 2\min\{P\{\chi_{n-1}^2 < (n-1)S^2/\sigma_o^2\}, 1 - P\{\chi_{n-1}^2 < (n-1)S^2/\sigma_o^2\}\}$$

Prueba de igualdad de varianzas de dos poblaciones normales

Sean X_1, X_2, \dots, X_n y Y_1, Y_2, \dots, Y_n muestras de una población normal con media desconocida μ_x, μ_y ; y varianza desconocida σ_x^2, σ_y^2 , y se desea probar la hipótesis de que

$$H_o : \sigma_x^2 = \sigma_y^2 \text{ contra } H_1 : \sigma_x^2 \neq \sigma_y^2$$

Recordando que $(n-1)S_x^2/\sigma_x^2, (m-1)S_y^2/\sigma_y^2$ tienen una distribución chi cuadrada con $n-1$ y $m-1$ grados de libertad, respectivamente. Entonces si H_o es verdadera,

$$S_x^2/S_y^2 \sim F_{n-1, m-1}$$

y entonces

$$P\{F_{1-\alpha/2, n-1, m-1} \leq S_x^2/S_y^2 \leq F_{\alpha/2, n-1, m-1}\} = 1 - \alpha$$

entonces, la prueba para un nivel de significancia de α es,

$$\text{se acepta } H_o \text{ si } F_{1-\alpha/2, n-1, m-1} \leq S_x^2/S_y^2 \leq F_{\alpha/2, n-1, m-1} \quad (2.121)$$

$$\text{se rechaza } H_o \text{ de otra manera} \quad (2.122)$$

El valor de p para los datos viene dado por:

$$\text{valor p} = 2\min\{P\{F_{n-1, m-1} < S_x^2/S_y^2\}, 1 - P\{F_{n-1, m-1} < S_x^2/S_y^2\}\}$$

Entonces, H_o se rechaza siempre que el nivel de significancia α al menos sea tan grande como el valor p.

Capítulo 3

Herramientas estadísticas a utilizar

3.1. Regresión lineal simple y correlación

El *análisis de regresión* es una técnica estadística para modelar e investigar la relación entre dos o más variables. En general, supóngase que hay una sola variable o respuesta y independiente, que se relaciona con k variables independientes o regresivas, digamos x_1, x_2, \dots, x_k . La variable de respuesta y , es una variable aleatoria, en tanto que las variables regresivas x_1, x_2, \dots, x_k se miden con error despreciable. Las x_j se llaman variables matemáticas y con frecuencia son controladas por el experimentador. El análisis de regresión también puede utilizarse en situaciones en las que y, x_1, x_2, \dots, x_k son variables aleatorias distribuidas conjuntamente, tal como en el caso cuando los datos se recaban como mediciones diferentes en una unidad experimental común. La relación entre estas variables se caracteriza por medio de un modelo matemático llamado *ecuación de regresión*. De modo más preciso, hablamos de la regresión de y en x_1, x_2, \dots, x_k . Este modelo de regresión se ajusta a un conjunto de datos. En algunas situaciones, el experimentador conocerá la forma exacta de la relación función verdadera entre y y x_1, x_2, \dots, x_k , por ejemplo, $y = \phi(x_1, x_2, \dots, x_k)$. Sin embargo, en la mayor parte de los casos, la verdadera relación funcional se desconoce, y el experimentador elegirá una función apropiada para aproximar ϕ . El siguiente desarrollo mantiene la esencia de las deducciones realizadas en la referencia [15], por lo que se recomienda consultarla para indagar más sobre el tema.

3.1.1. Regresión lineal simple

Deseamos determinar la relación entre una sola variable regresiva x y una variable de respuesta y . La variable regresiva x se supone como una variable ma-

temática continua, controlable por el experimentador. Supóngase que la verdadera relación entre y y x es una línea recta, y que la observación y en cada nivel de x es una variable aleatoria. Luego, el valor esperado de y para cada valor de x es:

$$E(y|x) = \beta_o + \beta_1 x \quad (3.1)$$

donde la ordenada de origen β_o y la pendiente β_1 son constantes desconocidas. Suponemos que cada observación, y , puede describirse mediante el modelo

$$y = \beta_o + \beta_1 x + \epsilon \quad (3.2)$$

donde ϵ es un error aleatorio con media cero y varianza σ^2 . Los ϵ se suponen también que son variables aleatorias no correlacionadas. El modelo de regresión de la ecuación 3.2, que involucra sólo una variable regresiva x a menudo recibe el nombre de *modelo de regresión lineal simple*. Supóngase que tenemos n pares de observaciones, por ejemplo $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$. Estos datos pueden emplearse para estimar los parámetros desconocidos β_o y β_1 en la ecuación 3.2. El procedimiento de optimización será el método de mínimos cuadrados. Esto es, se estimarán β_o y β_1 de manera que la suma de cuadrados de las desviaciones entre las observaciones y la línea de regresión sean mínimas. Empleando luego la ecuación 3.2, podemos escribir:

$$y_i = \beta_o + \beta_1 x_i + \epsilon, \quad i=1,2,\dots,n \quad (3.3)$$

y la suma de cuadrados de las desviaciones de las observaciones respecto a la línea de regresión verdadera es:

$$L = \sum_{i=1}^n \epsilon^2 = \sum_{i=1}^n (y_i - \beta_o - \beta_1 x_i)^2; \quad i=1,2,\dots,n \quad (3.4)$$

Los estimadores de mínimos cuadrados de β_o y β_1 , digamos $\hat{\beta}_o$ y $\hat{\beta}_1$, deben satisfacer:

$$\begin{aligned} \left. \frac{\partial L}{\partial \beta_o} \right|_{\hat{\beta}_o, \hat{\beta}_1} &= -2 \sum_{i=1}^n (y_i - \hat{\beta}_o - \hat{\beta}_1 x_i) = 0 \\ \left. \frac{\partial L}{\partial \beta_1} \right|_{\hat{\beta}_o, \hat{\beta}_1} &= -2 \sum_{i=1}^n (y_i - \hat{\beta}_o - \hat{\beta}_1 x_i) x_i = 0 \end{aligned} \quad (3.5)$$

La simplificación de estas dos ecuaciones produce:

$$n\hat{\beta}_o + \hat{\beta}_1 \sum_{i=1}^n x_i$$

$$\hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i \quad (3.6)$$

Las ecuaciones 3.6, se denominan *ecuaciones normales de mínimos cuadrados*. La solución para la ecuación normal es:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (3.7)$$

$$\hat{\beta}_1 = \left(\sum_{i=1}^n y_i x_i - \left(\sum_{i=1}^n y_i \right) \left(\sum_{i=1}^n x_i \right) / n \right) / \left(\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 / n \right) \quad (3.8)$$

donde $\bar{y} = (1/n) \sum_{i=1}^n y_i$ y $\bar{x} = (1/n) \sum_{i=1}^n x_i$. Por tanto, las ecuaciones 3.7 y 3.8 son los estimadores por mínimos cuadrados de la ordenada al origen y la pendiente, respectivamente. el modelo de regresión lineal simple ajustado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (3.9)$$

Respecto a la notación, es conveniente dar símbolos especiales al numerador y al denominador de la ecuación 3.8. Esto es, sea:

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 / n \quad (3.10)$$

$$S_{xy} = \sum_{i=1}^n y_i (x_i - \bar{x}) = \sum_{i=1}^n y_i x_i - \left(\sum_{i=1}^n y_i \right) \left(\sum_{i=1}^n x_i \right) / n \quad (3.11)$$

A S_{xx} se le llama la *suma corregida de cuadrados* de x y a S_{xy} la *suma corregida de productos cruzados* de x y y . Los lados del extremo derecho de las ecuaciones 3.10 y 3.11 son las fórmulas de calcularlas usualmente. Al emplear esta nueva notación, el estimador de mínimos cuadrados de la pendiente es:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad (3.12)$$

Suele ser necesario obtener una estimación de σ^2 . La diferencia entre la observación y_i y el correspondiente valor predicho \hat{y}_i , digamos $e_i = y_i - \hat{y}_i$, se denomina un *residuo*. La suma de los cuadrados de los residuos, o la *suma de cuadrados del error*, sería:

$$\begin{aligned} SSE &= \sum_{i=1}^n e_i^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \end{aligned} \quad (3.13)$$

Una fórmula de cálculo más conveniente para SS_E puede encontrarse sustituyendo el modelo ajustado $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ en la ecuación anterior y simplificando. El resultado es:

$$SS_E = \sum_{i=1}^n (y_i^2 - n\bar{y}^2 - \hat{\beta}_1 S_{xy}) \quad (3.14)$$

y si dejamos que $\sum_{i=1}^n (y_i^2 - n\bar{y}^2) = \sum_{i=1}^n (y_i - \bar{y})^2 \equiv S_{yy}$, entonces se puede escribir SS_E como:

$$SS_E = S_{yy} - \hat{\beta}_1 S_{xy} \quad (3.15)$$

Otro término que es importante definir es la suma total de cuadrados corregida para y :

$$S_{yy} \equiv \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.16)$$

Las dos componentes de S_{yy} miden, respectivamente, el tamaño de la variabilidad en la y_i , explicada por la línea de regresión, y la variación residual dejada sin explicar por la línea de regresión. A $SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ se le denomina la suma de cuadrados del *error* y a $SS_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ la *suma de regresión* de cuadrados. Por consiguiente, la ecuación anterior puede escribirse como:

$$S_{yy} = SS_R + SS_E \quad (3.17)$$

Al comparar la ecuación anterior con la ecuación 3.15, se puede observar que la suma de regresión de cuadrados SS_R es:

$$SS_R = \hat{\beta}_1 S_{xy} \quad (3.18)$$

Hay que hacer notar que $\hat{\beta}_0$ y $\hat{\beta}_1$ están distribuidos de la siguiente manera,

$$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2 \sum_i x_i^2}{nS_{xx}}\right)$$

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2/S_{xx})$$

y tomando en cuenta que,

$$SS_E = \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

entonces,

$$\frac{SS_E}{\sigma^2} \sim \chi_{n-2}^2$$

y que SS_E es independiente de los estimadores de mínimos cuadrados.

Debe tenerse cuidado al seleccionar las variables con las cuales se construyen los modelos de regresión y al determinar la forma de la función de aproximación. Es

muy posible desarrollar relaciones estadísticas entre variables que no tienen ninguna relación en un sentido práctico. Una fuerte asociación observada entre variables no necesariamente implica que existe una relación *causal* entre esas variables. Los experimentos diseñados son la única manera para determinar relaciones causales. Las relaciones de regresión son válidas sólo para valores de la variable independiente dentro del intervalo de los datos originales. La relación lineal que hemos supuesto tentativamente puede ser válida sobre el intervalo original de x , pero puede ser improbable que eso persista conforme encontremos valores de x más allá de ese intervalo. En otras palabras, cuando nos movemos fuera del intervalo de valores de x de los datos, menos certeza se tiene respecto a la validez del modelo supuesto. Los modelos de regresión no son necesariamente válidos para propósitos de extrapolación. por último, en ocasiones parece ser que el modelo $y = \beta x + \epsilon$ es apropiado. La omisión de la ordenada al origen de este modelo implica, que $y = 0$ cuando $x = 0$. Ésta es una suposición muy fuerte que a menudo no es justificada.

3.1.2. Coeficiente de determinación

La cantidad:

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}} \quad (3.19)$$

se denomina el *coeficiente de determinación*, y se emplea a menudo para juzgar la suficiencia de un modelo de regresión. Queda claro que $0 < R^2 < 1$. Con frecuencia nos referimos informalmente a R^2 como el grado de variabilidad en los datos explicados por el modelo de regresión. La estadística R^2 debe utilizarse con precaución, puesto que siempre es posible hacer a R^2 unitaria agregando simplemente suficientes términos al modelo. Por ejemplo, podemos obtener un ajuste “perfecto” a los n puntos dato con un polinomio de grado $n-1$. Además, R^2 siempre aumentará si añadimos una variable al modelo, pero esto no necesariamente significa que el nuevo modelo sea superior al antiguo.

3.1.3. Coeficiente de correlación

Es posible extraer deducciones en torno al coeficiente de correlación r para el modelo $\bar{y} = \beta_0 + \beta_1 \bar{x}$. El estimador r es el *coeficiente de correlación de la muestra*

$$\begin{aligned} r &= \left(\sum_{i=1}^n y_i(x_i - \bar{x}) \right) / \left(\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2 \right)^{1/2} \\ &= \frac{S_{xy}}{(S_{xx}S_{yy})^{1/2}} \end{aligned} \quad (3.20)$$

Nótese que:

$$\hat{\beta}_1 = \left(\frac{S_{yy}}{S_{xx}} \right)^{1/2} r \quad (3.21)$$

de modo que la pendiente $\hat{\beta}_1$ es justamente el coeficiente de correlación r de la muestra multiplicado por un factor de escala que es la raíz cuadrada de la dispersión de los valores y divididos por la dispersión de los valores de x . De tal modo $\hat{\beta}_1$ y r están estrechamente relacionadas, aunque ellas brindan información un poco diferente. El coeficiente de correlación r de la muestra mide la asociación lineal entre y y x , en tanto que $\hat{\beta}_1$ mide el cambio predicho en la media para un cambio unitario en x . En el caso de una variable matemática x , r no tiene significado porque la magnitud de r depende de la elección del espaciamiento para x . También podemos escribir, de la ecuación anterior:

$$\begin{aligned} R^2 \equiv r^2 &= \hat{\beta}_1^2 \frac{S_{xx}}{S_{yy}} \\ &= \frac{\hat{\beta}_1 S_{xy}}{S_{yy}} \\ &= SS_R / S_{yy} \end{aligned}$$

que podemos reconocer como el coeficiente de determinación. Esto es, el coeficiente de determinación R^2 es justamente el cuadrado del coeficiente de correlación entre y y x .

3.1.4. Inferencias sobre los estimadores de mínimos cuadrados

Inferencias sobre β_1

En el modelo de regresión lineal simple $y = \beta_0 + \beta_1 x + \epsilon$ se va a considerar la hipótesis de que $\beta_1 = 0$, es decir se va a comprobar de que no existe regresión en la variable de entrada. En la manera usual de expresarlo,

$$H_0 : \beta_1 = 0 \text{ contra } H_1 : \beta_1 \neq 0$$

y recordando que,

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2 / S_{xx})$$

entonces se puede proponer lo siguiente:

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / S_{xx}}} = \frac{\sqrt{S_{xx}} \hat{\beta}_1 - \beta_1}{\sigma} \sim N(0, 1) \quad (3.22)$$

que es independiente de $SS_E/\sigma^2 \sim \chi_{n-2}^2$ y recordando la definición de una variable aleatoria t , se tiene,

$$\frac{\sqrt{S_{xx}}(\hat{\beta}_1 - \beta_1)/\sigma}{\sqrt{\frac{SS_E}{\sigma^2(n-2)}}} = \sqrt{\frac{(n-2)S_{xx}}{SS_E}}(\hat{\beta}_1 - \beta_1) \approx t_{n-2} \quad (3.23)$$

es decir, que tiene una distribución t con $n-2$ grados de libertad. Si ocurriese el caso de que $\beta_1 = 0$ (es decir, que H_o sea verdadera) entonces

$$\sqrt{\frac{(n-2)S_{xx}}{SS_E}}\hat{\beta}_1 \sim t_{n-2} \quad (3.24)$$

Por tanto se puede proponer la siguiente prueba para H_o para un nivel de significancia α ,

$$\text{se rechaza } H_o \text{ si } \sqrt{\frac{(n-2)S_{xx}}{SS_E}}|\hat{\beta}_1| > t_{\alpha/2, n-2} \quad (3.25)$$

$$\text{se acepta } H_o \text{ de otra manera} \quad (3.26)$$

El valor p viene dado por:

$$\text{valor } p = P \left\{ |T_{n-2}| > \sqrt{\frac{(n-2)S_{xx}}{SS_E}}|\hat{\beta}_1| \right\} \quad (3.27)$$

$$= 2P \left\{ T_{n-2} > \sqrt{\frac{(n-2)S_{xx}}{SS_E}}|\hat{\beta}_1| \right\} \quad (3.28)$$

También es posible obtener un intervalo de confianza para $\hat{\beta}_1$ para cualquier α , $0 < \alpha < 1$,

$$P \left\{ -t_{\alpha/2, n-2} < \sqrt{\frac{(n-2)S_{xx}}{SS_E}}(\hat{\beta}_1 - \beta_1) < t_{\alpha/2, n-2} \right\} = 1 - \alpha$$

$$P \left\{ \hat{\beta}_1 - \sqrt{\frac{SS_E}{(n-2)S_{xx}}}t_{\alpha/2, n-2} < \beta_1 < \hat{\beta}_1 + \sqrt{\frac{SS_E}{(n-2)S_{xx}}}t_{\alpha/2, n-2} \right\} = 1 - \alpha$$

lo que da un estimador de intervalo de confianza de $100(1 - \alpha)\%$ para β_0

$$\left(\hat{\beta}_1 - \sqrt{\frac{SS_E}{(n-2)S_{xx}}}t_{\alpha/2, n-2}, \hat{\beta}_1 + \sqrt{\frac{SS_E}{(n-2)S_{xx}}}t_{\alpha/2, n-2} \right) \quad (3.29)$$

Inferencias sobre β_0

El mismo procedimiento utilizado para hallar la prueba de hipótesis y el intervalo de confianza para β_1 se utiliza para obtener

$$\sqrt{\frac{n(n-2)S_{xx}}{\sum_i x_i^2 S S_E}}(\hat{\beta}_0 - \beta_0) \sim t_{n-2} \quad (3.30)$$

por tanto el intervalo de confianza de $100(1 - \alpha)\%$ para β_0 es

$$\left(\hat{\beta}_0 - \sqrt{\frac{\sum_i x_i^2 S S_E}{n(n-2)S_{xx}}} t_{\alpha/2, n-2}, \hat{\beta}_0 + \sqrt{\frac{\sum_i x_i^2 S S_E}{n(n-2)S_{xx}}} t_{\alpha/2, n-2} \right) \quad (3.31)$$

3.2. Análisis de Varianza

La técnica de *análisis de varianza* permite hacer inferencias acerca de un conjunto de parámetros relacionados con medias poblacionales. Suponga que se disponen de muestras de tamaño n que provienen de m poblaciones diferentes, entonces el análisis de varianza permitirá probar la hipótesis de que las m medias poblacionales son iguales. Cuando la media de una variable aleatoria depende de un sólo factor, generalmente la muestra de la que proviene la variable, entonces se realizará un *análisis de varianza de un factor*. Una suposición importante que hace esta técnica es que los datos están distribuidos normalmente y con la misma varianza σ^2 (desconocida). Antes de comenzar una exposición formal del método es preciso demostrar una igualdad que se necesitará posteriormente.

Sean X_1, X_2, \dots, X_n variables aleatorias independientes y normales, pudiendo tener diferentes medias pero una varianza común σ^2 , y además $\mu_i = E[X_i], i = 1, 2, \dots, n$. Recordando que

$$Z_i = (X_i - \mu_i)/\sigma, \quad i=1, \dots, n \quad (3.32)$$

tiene una distribución normal estándar y también que

$$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n (X_i - \mu_i)^2 / \sigma^2$$

es una variable aleatoria chi cuadrada con n grados de libertad. Suponga que para cada uno de los valores de $\mu_i, i = 1, \dots, n$, es posible definirlos como una función lineal de un conjunto fijos de k parámetros desconocidos; también suponga que es posible conseguir estimadores de estos k parámetros, quedando así determinados estimadores para las medias μ_i . Designando por $\hat{\mu}_i$ al estimador resultante de $\mu_i, i = 1, \dots, n$, se puede decir entonces que

$$\sum_{i=1}^n (X_i - \hat{\mu}_i)^2 / \sigma^2 \quad (3.33)$$

tendrá una distribución chi cuadrada con $n-k$ grados de libertad. Es decir, se pierde 1 grado de libertad por cada parámetro que se sustituye por su estimador.

El método consiste en obtener dos estimadores para σ^2 , el primero independiente de la veracidad de la hipótesis nula, y el segundo válido sólo cuando la hipótesis nula es verdadera. Si resulta que H_o no es verdadera entonces el segundo estimador tiende a ser mayor que σ^2 . Luego se realiza la prueba, la cual consiste en calcular valores para estos dos estimadores y rechazar la hipótesis nula si la relación del segundo estimador entre el primero es lo suficientemente grande. Para información adicional sobre esta herramienta estadística por favor refiérase a [13, 14, 17].

3.2.1. Análisis de varianza de un factor

Sean m muestras independientes de tamaño n , donde los elementos de la muestra i ($X_{i1}, X_{i2}, \dots, X_{in}$), son variables independientes con media conocida μ_i y varianza desconocida σ^2 . Entonces se quiere probar que

$$H_o : \mu_1 = \mu_2 = \dots = \mu_m$$

contra la hipótesis alternativa

$$H_1 : \text{existen al menos dos medias diferentes}$$

Esto también se puede interpretar como: suponga que se tienen m tratamientos diferentes, donde el resultado de aplicar el tratamiento i a una muestra dada es un variable aleatoria normal con media μ_i y varianza σ^2 . Después de aplicar cada uno de los tratamientos a muestras distintas de n objetos y analizar los resultados, se quiere probar la hipótesis de que todos los tratamientos poseen el mismo efecto.

En total hay nm variables aleatorias independientes normales X_{ij} cuya suma de los cuadrados de sus formas normal estándar resultan en variables aleatorias chi cuadrada con nm grados de libertad. Expresándolo en la notación usual

$$\sum_{i=1}^m \sum_{j=1}^n (X_{ij} - E[X_{ij}])^2 / \sigma^2 = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \mu_i)^2 / \sigma^2 \sim X_{nm}^2 \quad (3.34)$$

Los estimadores de los m parámetros desconocidos, μ_1, \dots, μ_m , vienen dados por

$$X_i = \sum_{j=1}^n X_{ij} / n$$

donde X_i representa el promedio de todos los elementos en la muestra i , es decir, la media muestral de la población i -ésima, por tanto, X_i es el estimador de la

media muestral μ_i . Sustituyendo este estimador por la media μ_i en la ecuación 3.34 se tiene que

$$\sum_{i=1}^m \sum_{j=1}^n (X_{ij} - X_{i.})^2 / \sigma^2 \sim \chi_{nm-m}^2 \quad (3.35)$$

esto quiere decir que tendrá una distribución chi cuadrada con $nm-m$ grados de libertad. Designando al numerador por SS_W ,

$$SS_W = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - X_{i.})^2$$

a este estadístico se le denomina *suma de cuadrados dentro de la muestra*. Si se sustituye SS_W dentro de la ecuación 3.35 queda

$$SS_W / \sigma^2$$

Recordando que el valor esperado de una variable aleatoria chi cuadrada es su número de grados de libertad, entonces

$$E[SS_W / \sigma^2] = nm - m$$

o su forma equivalente,

$$E[SS_W / (nm - m)] = \sigma^2$$

Así se obtiene el primer estimador de σ^2 , $SS_W / (nm - m)$, el cual es independiente de la veracidad o falsedad de la hipótesis nula. El segundo estimador de σ^2 es válido sólo cuando la hipótesis nula, H_o , es verdadera. Si H_o es verdadera entonces todas las medias poblacionales μ_i son iguales, es decir $\mu_i = \mu, \forall i$, y las m medias muestrales $X_{1.}, \dots, X_{m.}$ tendrán la misma distribución con media μ y varianza σ^2 . Por tanto si H_o es verdadera la suma de los cuadrados de las m variables estándar será una variable aleatoria chi cuadrada con m grados de libertad. Es decir,

$$n \sum_{i=1}^m (X_{i.} - \mu)^2 / \sigma^2 \sim \chi_m^2 \quad (3.36)$$

Como todas las medias poblacionales son iguales a μ , entonces el estimador de μ es el promedio de los nm valores de datos,

$$X_{..} = \frac{\sum_{i=1}^m \sum_{j=1}^n X_{ij}}{nm} = \frac{\sum_{i=1}^m X_{i.}}{m} \quad (3.37)$$

y sustituyendo $X_{..}$ por μ en la ecuación 3.36, y si H_o es verdadera entonces

$$n \sum_{i=1}^m (X_{i.} - X_{..})^2 / \sigma^2$$

es una variable aleatoria chi cuadrada con $m-1$ grados de libertad.

Definiendo SS_b como la *suma de cuadrados entre muestras*,

$$SS_b = n \sum_{i=1}^m (X_{i.} - X_{..})^2$$

por lo tanto si H_o es verdadera entonces, SS_b/σ^2 tiene distribución chi cuadrada con $m-1$ grados de libertad, es decir,

$$E[SS_b/\sigma^2] = m - 1$$

o su forma equivalente,

$$E[SS_b/(m - 1)] = \sigma^2$$

lo que representa un estimador de σ^2 sólo si H_o es verdadera.

Es importante mostrar la siguiente igualdad ya que permite calcular SS_W teniendo el valor de SS_b ,

$$SS_W = \sum_{i=1}^m \sum_{j=1}^n X_{ij}^2 - nmX_{..}^2 - SS_b$$

Dado que $SS_b/(m - 1)$ tiende a ser mayor si H_o no es verdadera, y tomando la razón entre los dos estimadores de σ^2 se tiene que,

$$TS = \frac{SS_b/(m - 1)}{SS_W/(nm - m)}$$

es el estadístico de prueba, y se rechaza H_o si el estadístico de prueba TS es suficientemente grande, ahora es necesario determinar cuan grande debe ser TS para justificar el rechazo de H_o . Si H_o es verdadera SS_b y SS_W son independientes y TS tiene una distribución $F_{m-1, nm-m}$ es decir, es una variable aleatoria F con $m-1$ grados de libertad en el numerador y $nm-m$ grados de libertad en el denominador. Si $F_{m-1, nm-m, \alpha}$ es el $100(1 - \alpha)$ percentil de esta distribución, es decir,

$$P\{F_{m-1, nm-m} > F_{m-1, nm-m, \alpha}\} = \alpha$$

entonces la prueba de nivel de significancia α para H_o es

$$\text{se rechaza } H_o \text{ si } \frac{SS_b/(m - 1)}{SS_W/(nm - m)} > F_{m-1, nm-m, \alpha} \quad (3.38)$$

$$\text{se acepta } H_o \text{ de otra manera} \quad (3.39)$$

El valor p viene dado por:

$$\text{valor p} = P\{F_{m-1, nm-m} \geq TS\} \quad (3.40)$$

Ahora se presenta una tabla resumen con los resultados obtenidos, **Tabla ANOVA de un factor**

Fuente de Variación	Suma de Cuadrados	Grados de Libertad
Entre las muestras	$SS_b = n \sum_{i=1}^m (X_i - X_{..})^2$	n-1
Dentro de las muestras	$SS_W = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - X_i)^2$	nm-m

Prueba de nivel de significancia α	Valor p
Rechazar si $TS \geq F_{m-1, nm-m, \alpha}$	$P\{F_{m-1, nm-m} \geq TS\}$

3.2.2. Análisis de varianza de dos factores

Estimadores de los parámetros

Suponga que se tiene un conjunto de datos y que se pueden distribuir de la siguiente manera, arreglados en renglones y columnas,

$$\begin{pmatrix} X_{11} & X_{12} & \dots & X_{1j} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2j} & \dots & X_{2n} \\ X_{i1} & X_{i2} & \dots & X_{ij} & \dots & X_{in} \\ X_{m1} & X_{m2} & \dots & X_{mj} & \dots & X_{mn} \end{pmatrix}$$

de aquí se puede decir que hay m niveles para el primer factor y n niveles para el segundo factor, siendo X_{ij} el valor obtenido con el primer factor en el nivel i y el segundo factor en el nivel j . A el primer factor se le denomina *factor renglón*, y al segundo factor se le llama *factor columna*. Hay que recordar que se supone que todos los X_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$, son variables aleatorias independientes normales con varianza común σ^2 . También se supondrá que el valor medio de los datos depende, de manera aditiva de los factores renglón y columna. Sea $E[X_{ij}] = \mu_{ij}$, entonces el modelo aditivo para un par de constantes $a_i, i = 1, \dots, m$ y $b_j, j = 1, \dots, n$, queda,

$$\mu_{ij} = a_i + b_j$$

y definiendo los siguientes términos,

$$\mu_{i.} = \sum_{j=1}^n \mu_{ij}/n, \quad \mu_{.j} = \sum_{i=1}^m \mu_{ij}/m, \quad \mu_{..} = \sum_{i=1}^m \sum_{j=1}^n \mu_{ij}/nm$$

además de,

$$a_i = \sum_{j=1}^n \mu_{ij}/n, \quad b_j = \sum_{i=1}^m \mu_{ij}/m$$

se debe observar que,

$$\mu_{i.} = \sum_{j=1}^n (a_i + b_j)/n = a_i + b.$$

también que,

$$\mu_{.j} = a_{.} + b_j, \quad \mu_{..} = a_{.} + b_{.}$$

y denotando los términos anteriores como,

$$\begin{aligned} \mu &= \mu_{..} = a_{.} + b_{.} \\ \alpha_i &= \mu_{i.} - \mu = a_i - a_{.} \\ \beta_j &= \mu_{.j} - \mu = b_j - b_{.} \end{aligned}$$

por tanto el modelo queda descrito como,

$$\mu_{ij} = E[X_{ij}] = \mu + \alpha_i + \beta_j \quad (3.41)$$

donde,

$$\sum_{i=1}^m \alpha_i = \sum_{j=1}^n \beta_j = 0 \quad (3.42)$$

el valor de μ recibe el nombre de la *gran media*, α_i , la *desviación de la gran media debida al renglón i* , y β_j , la *desviación de la gran media debida a la columna j* . También se pueden definir los siguientes términos,

$$\begin{aligned} X_{i.} &= \sum_{j=1}^n X_{ij}/n = \text{promedio de los valores en el renglón } i \\ X_{.j} &= \sum_{i=1}^m X_{ij}/m = \text{promedio de los valores en la columna } j \\ X_{..} &= \sum_{i=1}^m \sum_{j=1}^n X_{ij}/n = \text{promedio de todos los valores} \end{aligned}$$

y recordando la ecuación 3.41 se obtiene que,

$$\begin{aligned} E[X_{i.}] &= \mu + \alpha_i, \text{ ya que } \sum_{j=1}^n \beta_j = 0 \\ E[X_{.j}] &= \mu + \beta_j, \text{ ya que } \sum_{i=1}^m \alpha_i = 0 \\ E[X_{..}] &= \mu \end{aligned}$$

o de forma equivalente,

$$\begin{aligned} E[X_{..}] &= \mu \\ E[X_{i.} - X_{..}] &= \alpha_i \\ E[X_{.j} - X_{..}] &= \beta_j \end{aligned}$$

es decir, los estimadores de μ, α_i, β_j , denotados por $\hat{\mu}, \hat{\alpha}_i, \hat{\beta}_j$, son,

$$\begin{aligned}\hat{\mu} &= X_{..} \\ \hat{\alpha}_i &= X_{i.} - X_{..} \\ \hat{\beta}_j &= X_{.j} - X_{..}\end{aligned}$$

Habiendo definido los estimadores de μ, α_i, β_j , se continua con la prueba de hipótesis,

$$H_o : \text{todas las } \alpha_i = 0$$

contra la hipótesis alternativa

$$H_1 : \text{al menos una } \alpha_i \text{ es diferente de } 0$$

en otras palabras la hipótesis nula dice que no existe efecto de renglón. También se definirá la hipótesis para el efecto columna,

$$H_o : \text{todas las } \beta_j = 0$$

contra la hipótesis alternativa

$$H_1 : \text{al menos una } \beta_j \text{ es diferente de } 0$$

Ahora se comienza con el procedimiento descrito anteriormente de determinar dos estimadores para σ^2 . Para el primer estimador se debe recordar que,

$$\sum_{i=1}^m \sum_{j=1}^n (X_{ij} - E[X_{ij}])^2 / \sigma^2 = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \mu - \alpha_i - \beta_j)^2 / \sigma^2$$

tiene distribución chi cuadrada con nm grados de libertad. Si se sustituyen μ, α_i, β_j por sus estimadores $\hat{\mu}, \hat{\alpha}_i, \hat{\beta}_j$, resultará que la ecuación anterior seguirá siendo chi cuadrada pero con un grado de libertad menos por cada parámetro estimado. Ahora bien, recordando que $\sum_{i=1}^m \alpha_i = \sum_{j=1}^n \beta_j = 0$, se observa que la suma de todas las α_i deben sumar cero, y una vez estimado $m-1$ de las α_i la última también quedara estimada. El mismo razonamiento se aplica a las β_j , estimando sólo $n-1$ de ellas. Por último se debe determinar μ lo que resulta en un grado de libertad menos. En total, los grados de libertad que se pierden son $1 + (m-1) + (n-1) = (n+m-1)$. Por tanto,

$$\sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2 / \sigma^2 \tag{3.43}$$

tendrá distribución chi cuadrada con $nm - (n+m-1) = (n-1)(m-1)$ grados de libertad. Observando que,

$$\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = X_{i.} + X_{.j} - X_{..}$$

y sustituyendo en la ecuación 3.43 queda,

$$\sum_{i=1}^m \sum_{j=1}^n (X_{ij} - X_{i.} - X_{.j} - X_{..})^2 / \sigma^2$$

la cual es una variable aleatoria chi cuadrada con $(n-1)(m-1)$ grados de libertad. Definiendo el estadístico SS_e como,

$$SS_e = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - X_{i.} - X_{.j} - X_{..})^2$$

a este estadístico se le llama *suma de cuadrados del error*. Dado que el valor esperado de una variable aleatoria chi cuadrada es su número de grados de libertad, y sustituyendo SS_e en la ecuación 3.43 se tiene que,

$$E[SS_e / \sigma^2] = (n-1)(m-1)$$

o,

$$E[SS_e / ((n-1)(m-1))] = \sigma^2$$

por tanto $SS_e / ((n-1)(m-1))$ es un estimador de σ^2 el cual es independiente de si la hipótesis nula H_o es verdadera o falsa. Para determinar el segundo estimador de σ^2 el cual es sólo válido cuando la hipótesis nula es verdadera se propone la hipótesis de,

$$H_o : \text{todas las } \alpha_i = 0$$

contra la hipótesis alternativa

$$H_1 : \text{al menos una } \alpha_i \text{ es diferente de } 0$$

Por lo tanto si H_o es verdadera, todas las α_i son cero, y entonces,

$$E[X_{i.}] = \mu + \alpha_i = \mu$$

como $X_{i.}$ representa el promedio de n variables aleatorias con varianza común σ^2 , se tiene que,

$$Var(X_{i.}) = \sigma^2 / n$$

y si H_o es verdadera entonces,

$$\sum_{i=1}^m (X_{i.} - E[X_{i.}])^2 / Var(X_{i.}) = n \sum_{i=1}^m (X_{i.} - \mu)^2 / \sigma^2$$

es una variable aleatoria chi cuadrada con m grados de libertad. Además si se sustituye μ por su estimador entonces se perderá un grado de libertad por lo que,

$$n \sum_{i=1}^m (X_{i.} - X_{..})^2 / \sigma^2$$

tendrá distribución chi cuadrada con $m-1$ grados de libertad, si H_o es verdadera.

Al estadístico definido por,

$$SS_r = n \sum_{i=1}^m (X_i - X_{..})^2$$

se le llama *suma de cuadrados de renglón*. De donde se puede ver que,

$$E[SS_r/\sigma^2] = m - 1$$

o

$$E[SS_r/(m - 1)] = \sigma^2$$

es decir, es un estimador de σ^2 sólo cuando H_o es verdadera, si no lo es, este valor tenderá a ser mayor que σ^2 .

Para probar la hipótesis nula de que no existe efecto de renglón, se determina la relación entre los dos estimadores a través del siguiente estadístico,

$$TS = \frac{SS_r/(m - 1)}{SS_e/(n - 1)(m - 1)}$$

hay que recordar que estos dos estimadores son independientes, si H_o es verdadera. La prueba para un nivel de significancia α es,

$$\text{se rechaza } H_o \text{ si } TS \geq F_{m-1, (n-1)(m-1), \alpha} \quad (3.44)$$

$$\text{se acepta } H_o \text{ de otra manera} \quad (3.45)$$

El valor p para esta prueba está definido por,

$$\text{valor p} = P\{F_{m-1, (n-1)(m-1)} \geq TS\} \quad (3.46)$$

De una manera similar se puede probar la hipótesis nula de que todas las β_j son cero. A continuación se presenta una tabla resumen para los resultados,

Tabla ANOVA de dos factores

	Suma de cuadrados	Grados de libertad
Renglón	$SS_r = n \sum_{i=1}^m (X_i - X_{..})^2$	m-1
Columna	$SS_c = m \sum_{j=1}^n (X_j - X_{..})^2$	n-1
Error	$SS_e = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - X_i - X_j - X_{..})^2$	(n-1)(m-1)

Hipótesis nula	Estadístico de prueba	Prueba de nivel de significancia α	Valor p
$\forall \alpha_i = 0$	$\frac{SS_r/(m-1)}{SS_e/(n-1)(m-1)}$	Rechazar si $TS \geq F_{m-1, (n-1)(m-1), \alpha}$	$P\{F_{m-1, (n-1)(m-1)} \geq TS\}$
$\forall \beta_j = 0$	$\frac{SS_c/(n-1)}{SS_e/(n-1)(m-1)}$	Rechazar si $TS \geq F_{n-1, (n-1)(m-1), \alpha}$	$P\{F_{n-1, (n-1)(m-1)} \geq TS\}$

3.3. Diseño de experimentos

Existe una gran diferencia entre los datos obtenidos de un experimento y las conclusiones que se deduzcan a partir de éstos. En general, el experimentador hace suposiciones acerca del sistema o proceso que genera los datos, para obtener información del fenómeno que se estudia. En base a las suposiciones y a la física que rige el proceso se desarrolla un modelo matemático del sistema. Por lo general estos modelos involucran constantes desconocidas. Es aquí donde la experimentación juega un papel fundamental, la de adquirir los datos que permitan estimar las constantes del modelo. En este punto se debe mencionar ciertas características de los datos que causan problemas a la hora de modelar los fenómenos estadísticamente:

- Suponga que se observa un cambio en las variables de operación de un sistema seguido por un cambio en las salidas del sistema. Esto no necesariamente significa que el cambio en el sistema causó el cambio en las salidas.
- Una suposición común en el modelado estadístico es que las observaciones son independientes entre sí. Los sistemas normales no funcionan de esta manera.
- Controlar un sistema en operación con frecuencia significa cambiar las variables de operación del sistema en grupos. Pero si dos variables cambian juntas es imposible separar sus efectos matemáticamente.

El diseño experimental se enfoca en estos problemas. La inmensa ventaja de un experimento diseñado es que se manipula activamente el sistema que se está estudiando. El diseño experimental es una técnica bien conocida para maximizar la información obtenida de un conjunto de experimentos. Los diseños experimentales funcionan midiendo los efectos que diferentes entradas tienen en un proceso. Esto se hace identificando un conjunto prospecto de variables de entrada, variándolas sobre una serie de experimentos, recolectando los datos y analizando los resultados. Una entrada se puede variar sobre un rango de valores, por ejemplo dos extremos (diseños de 2 niveles) o dos extremos y el caso base (diseño de tres niveles). Diseños de más de tres niveles raramente se usan. Para información más detallada acerca del diseño de experimentos, además de algunos ejemplos prácticos, consultar [18, 19].

3.3.1. Tipos de diseños

En general, los diseños experimentales más comunes son los *full factorial* y los *diseños de superficie de respuesta*. La selección de uno de estos diseños dependerá de los objetivos que se quieren lograr con el experimento. Los diseños

del tipo full factorial están orientados a investigar sobre una serie de factores, los que tienen más influencia sobre la salida del proceso. Los diseños de superficie de respuesta tienen como objetivo encontrar qué combinación de las variables de entrada producen la respuesta óptima deseada; esto involucra por ejemplo, maximizar la producción o minimizar las pérdidas, reducir la variabilidad del proceso, hacer el proceso más robusto, etc.

Diseños full factorial

Suponga que se quiere determinar si la variabilidad de un proceso mecánico se debe a la diferencia en el torno que corta las partes o los operadores quienes manejan los tornos. Si el mismo operador siempre maneja un torno dado entonces no se puede concluir si es la maquina o el operador quienes causan la variación en la salida. Sin embargo, si a cada operador se le pone a manejar cada torno se puede separar los efectos. Este es el enfoque del diseño full factorial, en el cual todos los niveles de un factor (entrada) dado se combinan con todos los niveles de los diferentes factores en el experimento. La principal desventaja de este tipo de diseño es que el número de combinaciones se incrementa exponencialmente con el número de variables que se quiere estudiar [20]. A continuación se explica el diseño full factorial de dos niveles, y posteriormente en la sección correspondiente a los *diseños de superficie de respuesta* se explicarán los diseños full factorial de tres niveles.

Diseños full factorial de dos niveles Uno de los diseños experimentales más comunes es aquel en el que los factores de entrada toman sólo dos valores cada uno. Estos niveles se denominan *bajo* y *alto* o -1 y +1, respectivamente. A los diseños con todas las posibles combinaciones alto/bajo de los factores de entrada se le denomina *diseño full factorial de dos niveles*. Por tanto se puede ver que si existen k factores, cada uno de los cuales tiene sólo dos niveles, entonces un diseño full factorial de dos niveles tendrá 2^k corridas o experimentos.

Número de factores	Número de corridas
2	4
3	8
4	16
5	32
6	64
7	128

Cuadro 3.1: Número de corridas para un diseño full factorial 2^k

Considere el diseño full factorial de dos niveles para tres factores, es decir, el

diseño 2^3 . Esto implica ocho corridas (sin tomar en cuenta *repeticiones* o *puntos centrales*). Una representación gráfica de este diseño se muestra en la figura 3.1. Las flechas indican la dirección en la que aumentan los valores de los factores. Los números del 1 al 8 en las esquinas del cubo de diseño indican el *orden estándar* de las corridas (sin aleatoriedad). El diseño también se presenta en forma tabular en el cuadro 3.2.

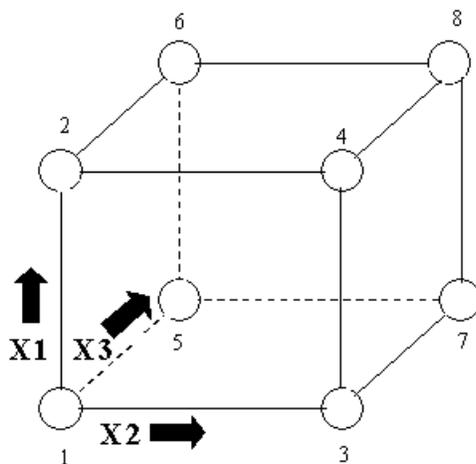


Figura 3.1: Diseño full factorial de dos niveles 2^3 , con factores X1, X2, y X3. Aparece en [18]

Nº de corrida	X1	X2	X3
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1

Cuadro 3.2: Diseño full factorial 2^3 en orden estándar

La primera columna del cuadro 3.2 indica el orden estándar de las corridas. Por ejemplo, en la corrida 1 todos los factores se encuentran en su valor bajo.

Regla para generar un diseño full factorial 2^k en orden estándar: La primera columna (X1) comienza con -1 y se va alternando el signo en todas las 2^k

corridas. La segunda columna (X2) comienza con -1 repetido dos veces, entonces cambia de signo, repetido dos veces hasta completar las 2^k corridas. La tercera columna (X3) con -1 repetido cuatro veces, entonces alterna a cuatro repeticiones de +1, así hasta completar el número total de corridas. En general, la columna (Xi) comienza con 2^{i-1} repeticiones de -1 seguida de 2^{i-1} repeticiones de +1, y así.

Ahora considere el siguiente ejemplo para un diseño 2^3 . Se necesita realizar un experimento con tres factores, estos son, presión (factor X1), velocidad (factor X2), y fuerza (factor X3), cada uno con dos valores posibles *alto* y *bajo*, en una máquina de producción para determinar cual tiene mayor influencia sobre la uniformidad del producto. Se corrieron dos repeticiones en cada combinación. Esto es, para un diseño full factorial 2^3 con dos repeticiones se necesita $2^3 * 2 = 8 * 2 = 16$ corridas.

I	X1	X2	X1*X2	X3	X1*X3	X2*X3	X1*X2*X3	Rep1	Rep2
+1	-1	-1	+1	-1	+1	+1	-1	-3	-1
+1	+1	-1	-1	-1	-1	+1	+1	0	-1
+1	-1	+1	-1	-1	+1	-1	+1	-1	0
+1	+1	+1	+1	-1	-1	-1	-1	+2	+3
+1	-1	-1	+1	+1	-1	-1	+1	-1	0
+1	+1	-1	-1	+1	+1	-1	-1	+2	+1
+1	-1	+1	-1	+1	-1	+1	-1	+1	+1
+1	+1	+1	+1	+1	+1	+1	+1	+6	+5

Cuadro 3.3: Matriz de análisis para un diseño full factorial 2^3

La parte de la matriz con valores de -1 y +1 se llama *matriz del modelo*, o *matriz de análisis* (ver cuadro 3.3). La tabla formada por las columnas X1, X2, y X3 se le llama *tabla de diseño* o *matriz de diseño*. Cuando todos los valores se codifican de manera que el valor alto es +1 y el valor bajo es -1, la matriz de diseño para cualquier experimento full factorial tiene columnas que son *ortogonales* por pares y todas las columnas suman cero. La propiedad de ortogonalidad es muy importante ya que elimina la correlación entre los estimados de los efectos principales y las interacciones.

Ejemplo de diseño full factorial¹: En el siguiente ejemplo se ilustra los conceptos de repetición, aleatoriedad, y puntos centrales. Suponga que deseamos mejorar la producción de una operación de pulido. Las tres entradas que se consideran importantes a la operación son: velocidad (X1), alimentación (X2), y profundidad (X3). Se quiere evaluar la importancia relativa de cada uno de esos factores en la producción (Y). La velocidad, alimentación y profundidad pueden ser variadas continuamente a lo largo de sus escalas respectivas, de una configuración baja a

¹Tomado de [18]

una alta. La producción se observa que varía suavemente cuando se hacen cambios progresivos a las entradas. Esto nos conduce a creer que la superficie de respuesta definitiva para Y será suave.

	baja(-1)	estándar (0)	alta (+1)	Unidades
Velocidad	16	20	24	rpm
Alimentación	0.001	0.003	0.005	cm/sec
Profundidad	0.01	0.015	0.02	cm/sec

Cuadro 3.4: Configuraciones alta (+1), baja (-1), y estándar (0) para una operación de pulido

Combinaciones de factores: Se quiere intentar varias combinaciones de esta configuración de manera que se pueda establecer la mejor forma de operar la pulidora. Hay ocho diferentes maneras de combinar las configuraciones alta y baja de velocidad, alimentación, y profundidad. Estas ocho se muestran en las esquinas del diagrama de la figura 3.1.

Se debe recordar que para k factores, pudiendo correr cada uno a dos niveles, habrá 2^k combinaciones de niveles. En este caso $k = 3$, entonces $2^3 = 8$ corridas. Si se deja correr la cantidad completa de todas las posibles combinaciones de los factores entonces podemos estimar todos los efectos principales y de interacción. Específicamente, hay tres efectos principales, tres efectos de interacción de dos factores, y una interacción de tres factores, los cuales aparecen todos en el siguiente modelo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{123} X_1 X_2 X_3 \quad (3.47)$$

Entonces, un diseño full factorial permite estimar los ocho coeficientes $\{\beta_0, \dots, \beta_{123}\}$.

El orden estándar: El número de las esquinas de la caja en la figura 3.1 se refiere a la forma estándar de escribir la configuración de un experimento llamado orden estándar. El orden estándar se puede observar en la representación tabular del diseño 2^3 . Es de hacer notar que las configuraciones de los factores han sido codificada, reemplazando los valores bajos por -1, y los valores altos por +1, como resultado se obtiene el cuadro 3.2.

Repetición: Correr el diseño entero varias veces hace el análisis de los datos más fácil, debido a que para cada corrida (es decir, para cada esquina de la caja de diseño) se obtiene un valor promedio para la respuesta, además de una idea de la dispersión (variabilidad, consistencia) de la respuesta en esa configuración. Una de las suposiciones usual de análisis es que la dispersión de la respuesta es uniforme a través del espacio experimental. El término técnico es la *homogeneidad de la varianza*. La repetición permite chequear esta suposición y posiblemente encontrar las combinaciones de configuración que dan producciones inconsistentes, lo que nos

permite evitar aquella área en el espacio del factor. Ahora se muestra una tabla de diseño para un diseño full factorial de dos niveles con tres factores, repetida dos veces (cuadro 3.5).

	Velocidad, X1	Alimentación, X2	Profundidad, X3
1	16, -1	.001, -1	.01, -1
2	24, +1	.001, -1	.01, -1
3	16, -1	.005, +1	.01, -1
4	24, +1	.005, +1	.01, -1
5	16, -1	.001, -1	.02, +1
6	24, +1	.001, -1	.02, +1
7	16, -1	.005, +1	.02, +1
8	24, +1	.005, +1	.02, +1
9	16, -1	.001, -1	.01, -1
10	24, +1	.001, -1	.01, -1
11	16, -1	.005, +1	.01, -1
12	24, +1	.005, +1	.01, -1
13	16, -1	.001, -1	.02, +1
14	24, +1	.001, -1	.02, +1
15	16, -1	.005, +1	.02, +1
16	24, +1	.005, +1	.02, +1

Cuadro 3.5: Diseño full factorial 2^3 repetido dos veces y presentado en orden estándar

Aleatoriedad: Si se corre el diseño en el orden mostrado tendría dos deficiencias, estas son:

- Sin aleatoriedad
- Sin puntos centrales

Mientras más libertad se da a la aleatoriedad de las corridas experimentales, más seguridad se tendrán contra factores externos que posiblemente afecten los resultados, lo que nos llevaría a desperdiciar nuestro tiempo y esfuerzo experimental. En nuestro ejemplo, considere la columna correspondiente a la profundidad, las configuraciones de la profundidad en orden estándar siguen un patrón de cuatro bajos; cuatro altos, etc. Ahora suponga que cuatro configuraciones se corran en el día y cuatro en la noche, y que la temperatura del ambiente en la tienda de pulidoras afecta la producción (algo desconocido para el experimentador). Si se corriera el experimento en esta forma concluiríamos que la profundidad influyó la producción, cuando en realidad la temperatura del ambiente fue la influencia significativa. De aquí se concluye algo importante: *aleatorizar las corridas experimentales tanto como sea posible*. A continuación se muestra la matriz de diseño para mostrar el efecto de la aleatorización.

Orden Aleatorio	Orden Estándar	X1	X2	X3
1	5	-1	-1	+1
2	15	-1	+1	+1
3	9	-1	-1	-1
4	7	-1	+1	+1
5	3	-1	+1	-1
6	12	+1	+1	-1
7	6	+1	-1	+1
8	4	+1	+1	-1
9	2	+1	-1	-1
10	13	-1	-1	+1
11	8	+1	+1	+1
12	16	+1	+1	+1
13	1	-1	-1	-1
14	14	+1	-1	+1
15	11	-1	+1	-1
16	10	+1	-1	-1

Cuadro 3.6: Diseño full factorial 2^3 repetido dos veces y presentado en orden aleatorio

Puntos centrales: Los puntos centrales nos permite determinar si durante las corridas, el proceso muestra alguna tendencia o ha cambiado drásticamente en la

salida del sistema. Así, nuestro ejemplo quedaría mejor si se añaden tres puntos centrales, al comienzo, mitad y final del experimento. La matriz de diseño final se muestra en el cuadro 3.7.

Orden Aleatorio	Orden Estándar	X1	X2	X3
1		0	0	0
2	5	-1	-1	+1
3	15	-1	+1	+1
4	9	-1	-1	-1
5	7	-1	+1	+1
6	3	-1	+1	-1
7	12	+1	+1	-1
8	6	+1	-1	+1
9		0	0	0
10	4	+1	+1	-1
11	2	+1	-1	-1
12	13	-1	-1	+1
13	8	+1	+1	+1
14	16	+1	+1	+1
15	1	-1	-1	-1
16	14	+1	-1	+1
17	11	-1	+1	-1
18	10	+1	-1	-1
19		0	0	0

Cuadro 3.7: Diseño full factorial 2^3 repetido dos veces y presentado en orden aleatorio con puntos centrales

Partición de diseños full factorial: A veces es necesario eliminar la influencia de factores externos cuando corremos el experimento, esto se logra a través de la *partición* o división del diseño. Por ejemplo, suponga que anticipamos que cambios predecibles ocurrirán mientras se corre el experimento. Esto podría suceder cuando se cambia a un nuevo lote de materiales a mitad del experimento. Se conoce que el cambio de los materiales tiene cierta influencia, y queremos eliminar su influencia en el subsiguiente análisis de los datos. En este caso, se necesita dividir el experimento en dos mitades (o bloques), uno con el primer lote de materiales, y el otro con el nuevo lote. La división tiene que balancear el efecto del cambio de materiales en una forma tal que elimine su influencia en el análisis, esto se logra mediante la *partición* (o *blocking*). Ejemplo: un diseño 2^3 tiene que ser dividido en dos grupos de 4 corridas cada uno. Considere la caja de diseño de la siguiente figura para el diseño 2^3 full factorial. La *partición* se puede lograr si al primer bloque se le asigna las esquinas oscuras y al segundo las esquinas claras (ver figura 3.2).

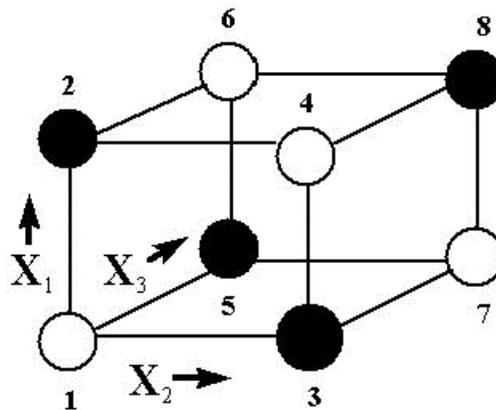


Figura 3.2: Esquema de partición para un diseño full factorial de dos niveles 2^3 , usando esquinas alternadas. Tomado de [18]

Esto funciona debido a que se está asignando la estimación del efecto de bloque al término de interacción de tres factores, y por la propiedad especial de los diseños de dos niveles llamada *ortogonalidad*. Estos es, los efectos de bloques se confunden con los efectos de interacción. La ortogonalidad garantiza que siempre podamos estimar el efecto de un factor o interacción libremente de alguna influencia debida a algún otro factor o interacción. La ortogonalidad es una propiedad muy deseada en los diseños experimentales y es por esta razón que los factoriales de dos niveles son tan populares y exitosos. En el cuadro 3.8 se presenta la tabla de diseño con

la columna correspondiente a la interacción de tres factores.

Velocidad, X1	Alimentación, X2	Profundidad, X3	X1*X2*X3	Bloque
-1	-1	-1	-1	I
+1	-1	-1	+1	II
-1	+1	-1	+1	II
+1	+1	-1	-1	I
-1	-1	+1	+1	II
+1	-1	+1	-1	I
-1	+1	+1	-1	I
+1	+1	+1	+1	II

Cuadro 3.8: Diseño full factorial 2^3 en dos bloques

Las filas que tienen -1 en la columna de interacción de tres factores se asignan al bloque I (las filas 1,4,6,7), mientras las otras filas se asignan al bloque II (filas 2,3,5,8). Note que el bloque I corresponde a las esquinas claras de la caja de diseño, y el bloque II a las esquinas oscuras. Como regla general para hacer la partición es usar una o una combinación de las columnas de ordenes superiores para construir los bloques. Esto proporciona una manera formal de dividir diseños complejos. El precio que se debe pagar por hacer la partición a través del uso de las columnas de ordenes superiores es que ya no se puede distinguir las interacciones de ordenes superiores de los efectos de bloque, en otras palabras los efectos se confundieron. De hecho, el efecto de bloque es ahora la suma del efecto de bloque y el efecto de interacción de orden superior. Esto funcionará bien mientras se cumpla la suposición de que los términos de interacción de ordenes superiores son despreciables, lo que generalmente ocurre. En cuanto a la aleatoriedad y puntos centrales dentro del bloque se realiza en la manera usual como si el bloque fuera un experimento aparte, lo que realmente es.

Diseños de superficie de respuesta

Algunas veces modelos lineales simples y de interacción no son adecuados. Por ejemplo, suponga que las salidas son defectos o producción, y el objetivo es minimizar los defectos y maximizar la producción. Si estos puntos óptimos están en el interior de una región en la cual el experimento será llevado a cabo, se necesita un modelo matemático que pueda representar la curvatura de manera que tenga un óptimo local. El más simple de este tipo de modelos tiene forma cuadrática:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 \quad (3.48)$$

que posee términos lineales para todos los factores, términos cuadráticos para todos los factores y productos de pares de factores. Los diseños para ajustar este tipo de modelos se conocen como *diseños de superficie de respuesta*. Uno de tales diseños es el diseño full factorial el cual tiene tres valores para cada entrada. Aunque este diseño no es muy satisfactorio debido a que realiza muchas más corridas de las que son necesarias para ajustar el modelo. Los dos diseños más comúnmente usados son los *diseños central composite* y los *diseños Box-Behnken*. En estos diseños las entradas toman o tres o cinco valores distintos (niveles), pero no todas las combinaciones de estos factores aparecen en el diseño.

Diseños full factorial de tres niveles Los diseños de tres niveles se escriben como un diseño factorial 3^k . Esto significa que se consideran k factores, cada uno con tres niveles. A estos se les denomina, generalmente, nivel bajo, intermedio, y alto. Numéricamente estos niveles se expresan como 0, 1 y 2. La razón para no considerar los números -1, 0 y +1, es que podría confundirse con los diseños de dos niveles en los cuales hay puntos centrales. Los diseños full factorial de tres niveles se propusieron para modelar posibles curvaturas en la función de respuesta y para tomar en cuenta el caso nominal de tres factores. Un tercer nivel para un factor continuo facilita la investigación de una relación cuadrática entre la respuesta y cada uno de los factores. Desafortunadamente los diseños de tres niveles son deficientes en cuanto al número de corridas, por tanto en términos de costo y esfuerzo. Es por ello que los diseños de dos niveles con puntos centrales es mucho menos costoso mientras todavía y en muy buena forma puede detectar la presencia o ausencia de curvatura. El diseño 3^2 es el más simple de los diseño de tres niveles. Tiene dos factores, cada uno a tres niveles. Las combinaciones para este tipo de diseño se puede mostrar gráficamente como en la figura 3.3. Por ejemplo la notación “20” significa que el factor A está al nivel (2) mientras el factor B al nivel (0).

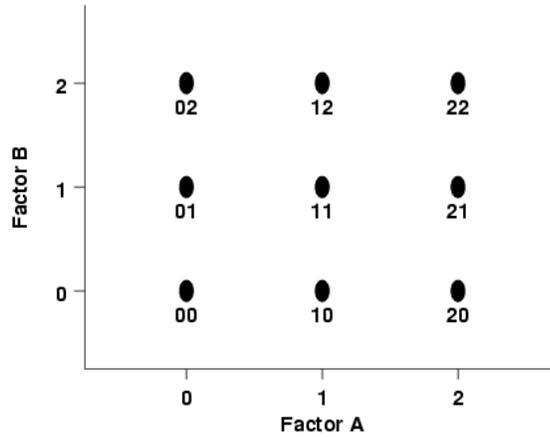


Figura 3.3: Diseño full factorial de tres niveles 3^2 . Tomado de [18]

El diseño 3^3 :

Este diseño consiste de tres factores, cada uno con tres niveles. El modelo para el experimento viene dado por:

$$Y_{ijk} = \mu + A_i + B_j + C_k + AB_{ij} + BC_{jk} + ABC_{ijk} + \epsilon_{ijk} \quad (3.49)$$

donde cada factor se incluye como factor nominal más que como una variable continua. En tales casos, los efectos principales tienen 2 grados de libertad, las interacciones de dos factores tienen $2^2 = 4$ grados de libertad y las interacciones de k factores tienen 2^k grados de libertad. El modelo contiene $2+2+2+4+4+4+8 = 26$ grados de libertad. Note que, si no hay repetición, el ajuste es exacto y no hay término de error en el modelo. En el caso de que no exista repeticiones, si se asume que no hay interacciones de tres factores, entonces se pueden usar los ocho grados de libertad restantes para estimar el término del error. Para este modelo $i = 1, 2, 3$ e igualmente para j y k , haciendo en total 27 corridas o experimentos. Las corridas se muestran en el cuadro 3.9. Esto puede ser representado gráficamente como en la figura 3.4. Uno de los mayores defectos de los diseños full factorial de tres niveles es que no poseen *rotación*.

Factor B	Factor C	Factor A 0	Factor A 1	Factor A 2
0	0	000	100	200
0	1	001	101	201
0	2	002	102	202
1	0	010	110	210
1	1	011	111	211
1	2	012	112	212
2	0	020	120	220
2	1	021	121	221
2	2	022	122	222

Cuadro 3.9: Diseño full factorial 3^3

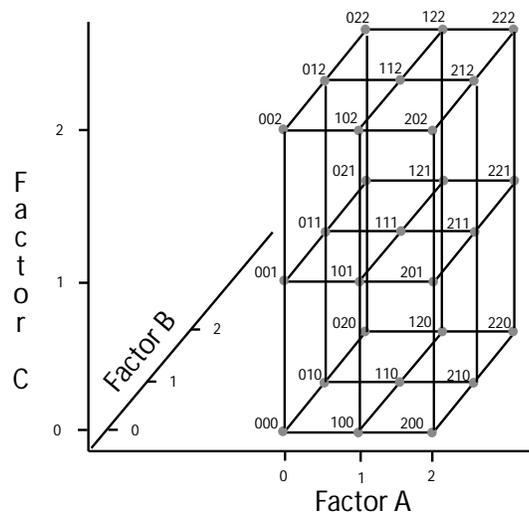


Figura 3.4: Diseño full factorial de tres niveles 3^3 . Tomado de [18]

Rotación de los diseños:

En un diseño rotativo la varianza de los valores predichos de y , es una función de la distancia de un punto desde el centro del diseño y no una función de la posición relativa del punto con respecto al centro. Por lo tanto, antes de comenzar un estudio no se tiene idea (o muy poca) acerca de la región que contiene la respuesta óptima. En los diseños rotativos los contornos asociados con la varianza de los valores predichos son círculos concéntricos (ver figura 3.5). En los gráficos se puede observar que la función de información depende sólo de la distancia de un punto desde el centro del espacio. La función de información viene definida por: $1/V(\hat{y})$, V denotando la varianza (la cual es una función de la matriz de diseño) del valor predicho \hat{y} . Para un diseño 3^2 el gráfico tridimensional y el gráfico de contorno viene dado por la figura 3.6. Como se puede observar la función de información del diseño no es sólo función de la distancia de un punto desde el centro del diseño, sino también de la dirección.

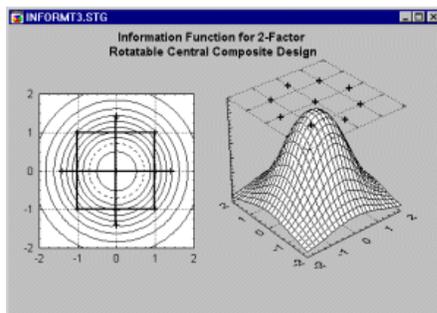


Figura 3.5: Función de información para un diseño cuadrático rotativo. Aparece en [19]

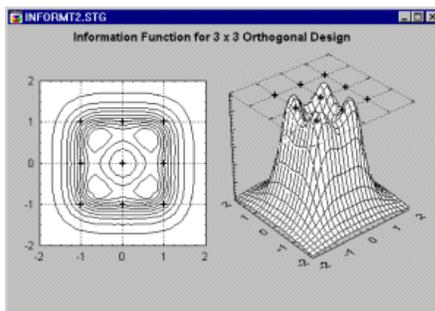


Figura 3.6: Función de información para un diseño full factorial 3^2 . Aparece en [19]

Diseños central composite Los diseños central composite son diseños de superficie de respuesta que pueden ajustar un modelo cuadrático completo. Los diseños central composite consisten de un diseño factorial más los *puntos estrellas* que permiten la estimación de posibles curvaturas. Si la distancia del centro del diseño a los puntos factoriales (aquellos puntos definidos por el diseño factorial) es ± 1 , entonces la distancia desde el centro del diseño a los puntos estrellas es $\pm \alpha$, con $|\alpha| > 1$. Al valor de α generalmente se le denomina *distancia axial*, y depende del número de factores, y otras propiedades que se desea que tenga el diseño.

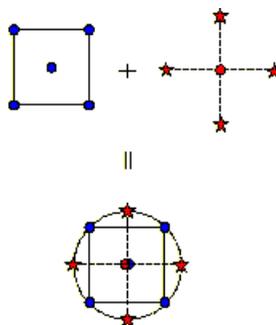


Figura 3.7: Generando un diseño central composite de dos factores. Tomado de [18]

En la figura 3.7 se puede observar como se genera un diseño central composite a partir de un diseño factorial de dos factores adicionándole los puntos estrellas. Es de hacer notar que los diseños central composite siempre tienen dos veces más puntos estrellas como factores haya en el diseño. Debido a como se definieron los puntos estrellas, éstos representan los nuevos valores extremos.

A continuación se presenta los diferentes tipos de diseños central composite a través de un diseño para tres factores. Para ilustrar un diseño central composite, imagine que tiene varios factores que pueden variar entre valores altos y bajos. Por conveniencia, suponga que cada uno de estos factores varía de -1 a $+1$. Un diseño central composite para tres factores consiste de las esquinas de un cubo que resulta del producto del intervalo $(-1,+1)$, puntos estrellas a lo largo de los ejes y que se encuentran fuera del volumen del cubo, y puntos centrales en el origen. Los diseños central composite son de tres tipos. Diseños circunscritos (CCC) son los descritos anteriormente. Diseños inscritos (CCI) son como los descritos arriba sólo que se escala de manera que los puntos estrellas tomen los valores de -1 y $+1$, y las esquinas del cubo queden dentro del volumen del cubo definido por el producto del intervalo $(-1,+1)$. Los diseños de caras (CCF) o *faceted* tiene los puntos

estrellas en las caras del cubo. Los diseños de caras tiene tres niveles por factor, en contraste con los otros tipos que tiene cinco niveles por factor.

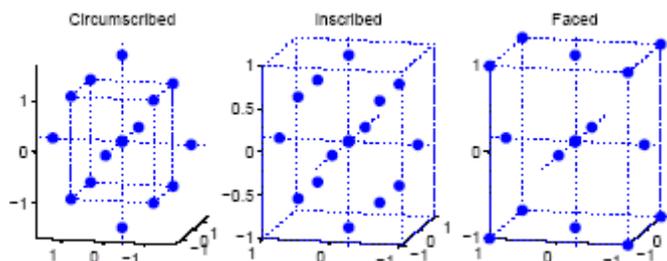


Figura 3.8: Tipos de diseños central composite de tres factores. Tomado de [20]

Características de los tipos de diseños central composite

Diseño CCC: Esto fueron los diseños originales de central composite, los puntos estrellas se encuentran a una distancia α desde el centro del diseño, dependiendo del número de factores y las propiedades deseadas del diseño. Los puntos estrellas representan los nuevos valores límites o extremos, esto implica que cada factor ahora posee cinco niveles. Los diseños poseen simetría circular, esférica, o hiper-esférica dependiendo del número de factores en el diseño, es decir, son rotativos. También son los diseños que abarcan una mayor superficie de respuesta o espacio del proceso (ver figura 3.9). Debido a que este diseño genera nuevos valores extremos, el investigador debe observar si estos nuevos límites son posibles para el proceso que se estudia.

Diseño CCI: En este diseño las configuraciones dadas vienen a ser los puntos estrellas, lo que implica que verdaderamente representan los valores límites, y a partir de estos se calculan los puntos factoriales dentro de los valores límites. En otras palabras, es un diseño que se obtiene a partir de un diseño CCC, dividiendo cada nivel del factor del diseño CCC por el valor de α para el diseño. Este diseño también posee cinco niveles por cada factor, además de que son rotativos, sin embargo son los diseños que poseen la menor superficie de respuesta (ver figura 3.9).

Diseño CCF: En este diseño los puntos estrellas se encuentran en el centro de cada cara del diseño factorial, de manera que $\alpha = \pm 1$. Al contrario de los otros dos diseños este sólo posee tres niveles por cada factor. Como desventaja de este tipo de diseño se encuentra que no son rotativos.

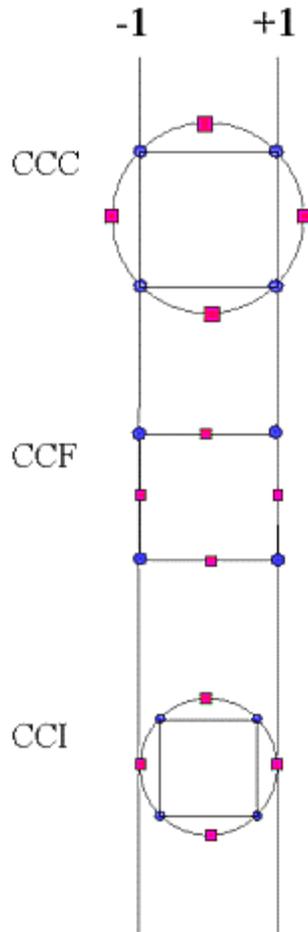


Figura 3.9: Comparando los tres tipos de diseño central composite de dos factores.
Tomado de [18]

Diseños Box-Behnken A diferencia de la mayoría de los diseños central composite, los diseños Box-Behnken usan sólo tres niveles por cada factor, además de que no se obtienen a partir de un diseño factorial. Esto los hace atractivos cuando los factores son cuantitativos pero el conjunto de valores que se puede obtener es pequeño. Los diseños de caras (CCF) también usan sólo tres niveles por cada factor. Sin embargo ellos no son rotativos como lo son los diseños Box-Behnken. Por otra parte, los diseños Box-Behnken se espera que tengan poco poder de predicción en las esquinas del cubo que encierra el diseño, porque a diferencia de los diseños CCF ellos no incluyen puntos en las esquinas del cubo. La figura muestra un diseño de Box-Behnken para tres factores, con el punto resaltado en el origen y posiblemente repetido en varias corridas. Para este diseño todos los puntos, excepto los puntos centrales se encuentran a una distancia de $\sqrt{2}$ del origen. Por supuesto esto no se mantiene para los diseño Box-Behnken que tienen diferentes números de factores.

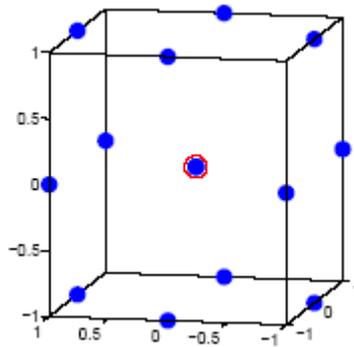


Figura 3.10: Diseño Box-Behnken de tres factores. Tomado de [20]

Capítulo 4

Antecedentes del problema

Diversas técnicas de balance de materiales han sido ampliamente usadas para la estimación de hidrocarburos en sitio mediante la línea recta que mejor ajusta los datos, o ajuste de mínimos cuadrados. Existen muchos casos de campo en los cuales los estimados de hidrocarburos en sitio no eran correctos. Existen muchas publicaciones acerca de las diferentes técnicas de balance de materiales, pero ha habido poca discusión acerca de la incertidumbre asociada a cada una de estas técnicas.

4.1. Investigación de McEwen

McEwen [1] presentó una técnica, para los cálculos de balance de materiales con influjo de agua en la presencia de incertidumbre en las presiones, en la cual el ajuste de mínimos cuadrados produce sólo una constante, el petróleo (o gas) en sitio. La constante de influjo de agua (para flujo radial y lineal) se calculaba independientemente del ajuste, a partir de estimados de la compresibilidad efectiva y la saturación de agua del yacimiento. McEwen menciona que una de las principales suposiciones del análisis de regresión de mínimos cuadrados es que toda la incertidumbre se encuentra en la variable de respuesta, y en base a esto reordena la Ecuación de Balance de Materiales (EBM), encontrando que el término de la expansión de fluidos es el que posee la más grande incertidumbre debido a que contiene la diferencia de dos cantidades aproximadamente del mismo tamaño. Menciona que incluso conociendo los valores de presión con una desviación de 1 por ciento, el error relativo del término de la expansión es muy alto; esta situación disminuye a medida que la presión declina. Asumiendo que el modelo de acuífero es el adecuado y que los parámetros están bien estimados entonces la incertidumbre en el influjo de agua se debe principalmente a la presión. Para aquellos modelos basados en el principio de superposición, la incertidumbre en el

cálculo del flujo adimensional es más grande que en los valores de los datos de presión. Debido a que en el término de flujo existe una sumatoria permite que los errores en las diferencias de presión se compensen entre si, es decir, si una presión es muy alta el Δp por encima de esta presión será muy bajo pero el Δp por debajo será alto por la misma cantidad. Por tanto el término de flujo posee menos incertidumbre que el término de expansión pero más incertidumbre que los datos de presión. Por último los datos de vaciamiento son las cantidades que poseen menos incertidumbre, en la medida que los errores en los datos de presión sean lo suficientemente grandes. Por esta razón McEwen aisló la mayor parte de la incertidumbre en la variable dependiente, es decir, la suma de la expansión de los fluidos y el flujo de agua. La forma en la que reordeno la EBM fue: $y = ax$, donde x representa el vaciamiento, a es el inverso del POES, y y es igual a la suma de la expansión más el flujo. El modelo de acuífero que utilizó vino dado por la teoría de Hurst y Van Everdingen. La técnica que utilizó consistió en graficar la suma del cuadrado de las desviaciones (SSD) de los valores observados y calculados de y , versus el paso de tiempo adimensional Δt_d para el calculo del flujo de agua. Esta gráfica posee un mínimo para el valor de Δt_d que da el mejor estimado del Petróleo Original En Sitio (POES). Los valores de SSD se normalizaron por el cuadrado del máximo valor de y para compensar el efecto en el SSD del cambio en magnitud de y a medida que se cambia Δt_d . Para simular los errores en la presión se tomaron números de una tabla de desviaciones normales aleatorias con promedio 0 y desviación estándar 1, luego se multiplicaron por una constante que representa la desviación estándar del conjunto de presiones, y posteriormente se le sumo algebraicamente a las presiones reales para obtener así los valores de presión “errados”. Las desviaciones estándar que tomó en cuenta fueron de 10 y 20 psi, considerándolas errores modestos para los datos de presión. En resumen, la ecuación resultante de ordenar la EBM posee las siguientes características:

- es una línea recta a través del origen.
- la pendiente es igual al inverso del POES.
- la mayoría de la incertidumbre se aísla en la variable dependiente.

4.2. Investigación de Tehrani

Tehrani [2] analizó varias formas de reordenar la EBM para yacimientos de petróleo con influjo de agua para calcular el petróleo en sitio, demostró que varias de estas formas tienen bajo poder de resolución y están propensas a producir resultados erróneos en la presencia de incertidumbre en las presiones y producción, además del cambio de la incertidumbre de la presión con el tiempo. Comentó que los métodos con la forma: $Z/X = A + BY/X$ (la cual es la forma de los métodos de F/E_o vs E_g/E_o y F/E_t vs W_e/E_t); donde Z representa el vaciamiento, X la expansión de fluidos, y Y el influjo de agua, llevan a cabo un ajuste de mínimos cuadrados ponderando cada punto i con un peso de $(1/Xi)^2$, en otras palabras están dando más peso a los primeros puntos que a los últimos. Tehrani no justifica esto y comenta varias razones:

- la tecnología de medición mejora con el tiempo, por lo tanto, los datos de presión y producción de los primeros años son generalmente menos exactos que aquellos de los últimos años.
- recordando que los términos de expansión de fluidos e influjo de agua son funciones del cambio de presión, se observa que un error pequeño en la presión en los primeros años corresponde a un gran porcentaje del cambio de presión, mientras que el mismo error en la última parte de la historia corresponde a un pequeño porcentaje del cambio de presión. Por tanto los términos de la expansión de fluidos y el influjo de agua son menos confiables para los primeros años.
- en los primeros años de historia existen menos datos disponibles, debido a que existe menos pozos, menos medidas de presión, menos producción, etc. De esta manera la presión del yacimiento obtenida no será tan representativa como la presión del yacimiento en la última parte de la historia de producción.
- cuando se utilizan modelos de influjo de agua que aplican el principio de la superposición, el cambio de presión de cada año tiene influencia sobre todos los siguientes años, mientras que el último cambio de presión sólo tendrá influencia sobre el influjo de agua del último año.

Debido a estas razones concluye que si se debiera pesar algunos puntos los últimos tendrían que pesar más que los primeros. También explica que el dividir la EBM por el término de expansión (u otro cualquier término que no sea una constante), se está reduciendo el número de variables pero se crea una relación entre las nuevas variables a través del término de la expansión de los fluidos. De hecho, demuestra que teniendo tres conjuntos de números aleatorios independientes, los

cuales representan tres variables X , Y , y Z ; y si se dividen por uno de ellos (por ejemplo X), se crean dos nuevos conjuntos (Z/X y Y/X) los cuales muestran una correlación considerable, es decir, se crea una correlación superficial entre estas dos nuevas variables. Por otro lado, si existiese alguna relación entre estas variables (X , Y , y Z), podría ser distorsionada y oscurecida por la división de dos de estas variables (Y , y Z) por la restante (X). El modelo de yacimiento a utilizar fue del tipo de empuje por agua. El método de análisis consistió en reordenar la EBM en diez formas diferentes y utilizarlas para el cálculo del POES y la constante de influjo de agua. Se introdujeron tres tipos de errores:

- errores en la historia de presión, los errores aleatorios en un rango especificado se le sumaron a las presiones “idealizadas” que venían del modelo, las cuales denominó presiones “observadas”.
- errores en el vaciamiento, esto se logro mediante errores aleatorios de 1 por ciento del vaciamiento sumados a los valores de vaciamiento obtenidos del modelo.
- errores que cambian en el tiempo para la historia de presión, es decir, en el primer año se consideró un gran error en el valor de la presión, este error se fue disminuyendo hasta alcanzar el último año, resultando en una magnitud pequeña comparada al error del primer año;

este conjunto de errores se le sumo a los valores de presión de la historia. El siguiente paso fue hacer un conjunto de corridas en las cuales tomó en cuenta los errores de la siguiente manera: un conjunto en el cual sólo tomó en cuenta las presiones “observadas”, otro conjunto con las presiones “observadas” y el vaciamiento con los errores aleatorios, y el último consistía en los tres tipos de error. De cada corrida obtuvo el valor del POES y la constante de influjo de agua. A continuación calculó la historia de presión del yacimiento y la desviación estándar del las presiones calculadas de las presiones “idealizadas” o reales y la desviación estándar de las presiones calculadas de las presiones “observadas”. Al final de los resultados obtenidos de todas las corridas cálculo intervalos de confianza y la desviación estándar del POES para cada forma de la EBM. Las conclusiones a las que llego fueron:

- las ecuaciones con la forma: $Z/X = A + BY/X$, no se deberían utilizar ya que tiene bajo poder de resolución para el cálculo simultáneo de A y B .
- el número de variables en la ecuación original no debería ser reducida de una forma indiscriminada mediante la combinación de varias de sus variables o su eliminación.

- observó que las presiones calculadas estaban más cerca a las presiones reales que a las presiones “observadas”, y que la eliminación de presiones debido a que tienen una gran desviación con respecto a las presiones calculadas no tenía justificación.
- el método que resultó más exacto fue el de minimizar la suma del cuadrado de las desviaciones de las presiones calculadas con respecto a las presiones “observadas”, aunque este método resultó ser el más laborioso y que requería más tiempo de cálculo.
- por último que el proceso de optimización debe basarse en minimizar la varianza de una variable que tiene valores calculados y valores observados.

4.3. Investigación de Fair

Fair [21] realizó un análisis estadístico a la ecuación de balance de materiales modificada para incluir el petróleo que se encuentra en la fase gaseosa, esto con el fin de poder utilizarla en casos de yacimientos con crudos volátiles, de condensado y de gas. La resolución de la EBM se realizó a través de las técnicas de regresión lineal múltiple, aplicándole todas las técnicas estadísticas aplicables al análisis de regresión. Resalta que en las consideraciones estadísticas acerca de los modelos de regresión se encuentran:

- existe una relación causa-efecto entre las variables.
- los valores de los parámetros de control, se ajustan por el experimentador mientras se mide la respuesta. También comenta que en el caso del problema de balance de materiales, esto sería ajustar los términos de expansión de fluidos e influjo de agua, mientras se mide el vaciamiento, o lo que es equivalente, ajustar la presión y medir la producción. A su parecer el proceso es más bien el inverso, es decir, se producen los fluidos, y periódicamente se mide la presión y se calculan los términos de la expansión de fluidos.
- para el caso de balance de materiales los términos en la regresión no son totalmente independientes, aunque su comportamiento modelado así lo sea. Esto se debe a que los fluidos se expanden a medida que la presión se reduce, por tanto, existirá una correlación inherente en los términos de la regresión. Fair recomienda calcular la matriz de covarianza-varianza de los términos de la regresión para verificar este efecto.

Entre las herramientas estadísticas utilizadas para el análisis se encuentran el coeficiente de correlación múltiple, la estadística F, los grados de libertad, intervalos de confianza para los coeficientes estimados, y la prueba de aleatoriedad. Para el autor el coeficiente de correlación múltiple no es siempre útil para evaluar los resultados de la regresión, debido a que siempre se obtendrán valores altos debido a las correlaciones inherentes en las propiedades de los fluidos. También menciona que los intervalos de confianza son una herramienta útil a la hora de evaluar cuan exactos son los estimados. Un punto importante que menciona el autor es que el uso de métodos estadísticos no necesariamente garantiza que se obtendrán resultados confiables, y que no existe alguna herramienta que pueda sustituir el juicio que tenga el ingeniero ante el conocimiento del problema.

4.4. Investigación de Carlson

Carlson [22] presentó una propuesta matemática acerca del análisis de error en los datos de presión aplicado a los cálculos de balance de materiales. El análisis mostró que el error más grande en la determinación del petróleo en sitio es el nivel de presión en relación al punto de burbujeo. El siguiente factor que tiene mas influencia en los resultados es la capa de gas. También mostró que los cálculos eran relativamente insensibles a errores en la tasa gas - petróleo producida. Encontró que debe tenerse cuidado al aplicar la minimización estadística, también que la regresión debe ser hecha en las variables independientes. Los principales comentarios que realizó fueron:

- para el caso estudiado en el cual existe una gran permeabilidad del yacimiento, la exactitud de los datos de presión estaba alrededor de 30 psi debido a las variaciones areales en la presión.
- las mediciones de presión del tipo amerada tienen una exactitud de 0.25 % de la deflexión completa.
- bajo condiciones ideales, los estimados de la capa de gas tienen, en el mejor de los casos, 10 % de exactitud.
- no utilizar regresión lineal en variables agrupadas.
- utilizar a la presión como la variable para comparar valores observados de valores calculados funciona bien, ya que no se está agrupando las variables.
- los datos PVT provenientes del laboratorio son de alta calidad, siendo las dos principales fuentes de error: el usar una muestra no representativa o una prueba de laboratorio mal realizada, la primera ocurriendo con más frecuencia que la segunda.
- al utilizar correlaciones debe conocerse su exactitud, al igual que el porcentaje de error, desviación estándar y el rango de aplicación para los parámetros.
- para el caso analizado se observó una dispersión en los datos de presión con el tiempo, también al observar un gráfico de las presiones promedios anuales algunas veces se observa varios incrementos o máximos en la curva de presión. Establece que el caso ideal sería obtener presiones ponderadas por el volumen poroso, sin embargo es un trabajo muy arduo y resulta más conveniente un ajuste polinomial a los datos, además que se obtiene una curva suave. También comenta que es buena idea establecer rangos de error para los datos.

- los cálculos de balance de materiales pueden indicar capa de gas cuando en realidad no existe, al igual que puede resultar en múltiples soluciones (cuando se intentan estimar varios parámetros a la vez), por tanto, se debe tomar en cuenta los datos volumétricos del yacimiento para comparar y elegir la respuesta más adecuada.

4.5. Investigación de Wang y Hwan

Wang y Hwan [3] no sólo investigaron los efectos de la incertidumbre en las presiones sino que también evaluaron el efecto de los mecanismos de empuje en las estimaciones de hidrocarburos en sitio, especialmente mostraron que aunque los cálculos de hidrocarburos en sitio son bastante imprecisos en yacimientos con fuerte empuje de agua, en presencia de errores en los datos de presión, los estimados de influjo de agua son bastante buenos. De igual forma observaron que los estimados de gas original en sitio (GOES) son más precisos que los estimados del POES para yacimientos con grandes capas de gas. El impacto de errores en los datos de presión en los cálculos de balance de materiales es mínima para un yacimiento del tipo volumétrico, mientras que para los yacimientos con capa inicial de gas o con empuje de agua, el impacto puede ser significativo, dependiendo en el tamaño del acuífero o de la capa de gas. Para yacimientos con capa de gas encontraron que el error en el valor del POES incrementa dramáticamente con el valor de m , además recomiendan no usar el método de la capa de gas para el cálculo simultáneo del POES y m . Para yacimientos con capa inicial de gas e influjo de agua los errores en el POES incrementan dramáticamente incluso para acuíferos pequeños (para una capa de gas pequeña, por ejemplo, $m = 0,1$), también observaron que la tendencia de los POES calculados era hacia valores por debajo del valor real; recomiendan que para minimizar la incertidumbre en los cálculos era necesario encontrar alguna de las incógnitas de alguna fuente diferente a la EBM o limitar los rangos de estas incógnitas.

4.6. Investigación de Walsh

Walsh [4] presento un análisis de la incertidumbre en la presión para varias gráficos de balance de materiales, los gráficos fueron clasificados de acuerdo a su tolerancia al error y sensibilidad. Los gráficos que consideró fueron (1) F vs E_o , (2) F/E_o vs E_g/E_o , (3) F/E_o vs $(E_o + mB_{oi}E_g/B_{gi})$ y (4) regresión planar de mínimos cuadrados. Para el análisis de cada uno de los gráficos también tomó en cuenta el efecto de la capa de gas. Los yacimientos a analizar fueron del tipo volumétrico e inicialmente saturado con capas de gas de 0, 0.25, y 0.5; todos conteniendo el mismo POES y con las mismas propiedades PVT. La incertidumbre se introdujo asumiendo que el error en las medidas de presión era aleatorio y distribuido normalmente. La magnitud del error vino dada por la desviación estándar y el error estándar de cada distribución. Para asegurar el muestreo representativo se tomaron en cuenta diez diferentes realizaciones de error normalizadas. Los errores de la presión estuvieron en el rango de 0-200 psi. La presión “actual” se determinó por sumar el producto del error normalizado y la desviación estándar a las presiones “reales”. Las conclusiones a la que llegó Walsh fueron: el método de graficar (1) F vs E_o para yacimientos volumétricos inicialmente subsaturados, exhiben buena tolerancia a la incertidumbre en la presión. El método de graficar (2) F/E_o vs E_g/E_o para yacimientos volumétricos inicialmente saturados exhiben poca tolerancia a la incertidumbre en la presión, y fue denominado hipersensible. El método de graficar (3) F vs $(E_o + mB_{oi}E_g/B_{gi})$ y (4) de regresión planar para yacimientos volumétricos inicialmente saturados exhiben mejor tolerancia que el método de F/E_o vs E_g/E_o , sin embargo, todavía presentan una tolerancia pobre ante la incertidumbre en la presión. También es muy importante hacer notar que algunos gráficos resultaban en altos coeficientes de correlación arrojando estimados del POES con un error bastante considerable, por tanto Walsh resaltó que la linealidad de un método de balance de materiales no es un indicador confiable de su aplicabilidad.

4.7. Investigación de Baker, Regier y Sinclair

Baker, Regier y Sinclair [5] analizaron los errores de las propiedades PVT en los cálculos de balance de materiales. Errores aleatorios y sistemáticos se introdujeron en las propiedades de: factores volumétrico de formación para el petróleo, gas, y agua (B_o , B_g , y B_w), tasa de gas-petróleo en solución (R_s), presión de burbujeo (P_b), y gravedad API. La cantidad de error sistemático introducido fue de +/- (2-10) %. El error aleatorio de +/- (0-2) % se aplicó a todas las propiedades PVT para tomar en cuenta los errores típicos en el laboratorio. Después hicieron los cálculos de balance de materiales usando el PVT erróneo y calcularon el POES, y en los casos apropiados el influjo de agua We . Estos resultados influenciados por errores entonces fueron comparados a un caso base, sin error. Los efectos de los diferentes parámetros sobre los errores en el POES calculado se examinaron, incluyendo los errores introducidos de forma sistemática a sólo una propiedad PVT a la vez, y usando correlaciones PVT en lugar de datos de laboratorio. También se observaron diferencias en los errores de los POES calculados para diferentes mecanismos de empuje. Los errores sistemáticos introducidos consistió en sumar o restar la desviación especificada a todos los valores de la propiedad PVT obtenida de correlaciones o de datos del laboratorio. Los errores aleatorios se introdujeron tomando cada valor de la propiedad PVT y asignándole un número al azar dentro de un intervalo definido por la desviación especificada. Realizaron cálculos de caso base para cada yacimiento usando los datos PVT existentes para determinar el POES y el influjo de agua. Luego los cálculos del caso base se usaron para obtener el porcentaje de error en el POES y We cuando se usaban los PVT erróneos. También comentaron que el error en el POES depende del mecanismo de empuje, cambio de presión total en el yacimiento, exactitud de las mediciones de presión, exactitud del PVT, y exactitud de las mediciones de producción. Por ejemplo, si un yacimiento está influenciado por una gran capa de gas, entonces el término debido a la expansión de la capa de gas dominará la EBM. Por tanto los cálculos de balance de materiales mostraran volúmenes de capa de gas más exactamente que el POES calculado, es decir las barras de error para el GOES serán más grandes que para el POES. También se observará que la EBM será mucho más sensible al término B_g que a los otros parámetros. Concluyeron que los errores en el petróleo en sitio calculado dependía en el cambio de presión total del yacimiento, mientras mayor el cambio de presión en el yacimiento, el petróleo en sitio calculado es menos sensible al error en el PVT. Mas aún, observaron que el error en el POES se amplificaba en relación al error en el PVT, en otras palabras el error en el POES es mayor que el error en el PVT. Mientras que para los casos en que existe un cambio de presión considerable (30 % de la presión inicial) el error en el POES es aproximadamente igual al error en el PVT. La propiedad PVT que tuvo el más grande impacto en las estimaciones del petróleo en sitio fue la tasa

de gas - petróleo en solución (R_s). Específicamente, mientras se incrementaba el R_s disminuía el POES. Para el término del factor volumétrico de formación del petróleo (B_o) la relación fue inversa, es decir, incrementos en el B_o causaban un incremento en el POES. Por último cuando se aumentaron y disminuyeron todos los valores del PVT se obtuvo menor error del POES en comparación con cada uno de los parámetros individuales, esto fue debido a que los efectos de los parámetros se compensaron entre sí. También demostraron que el uso de las correlaciones PVT sin ajustar a los datos de campo causan un gran error en el estimado del POES, al igual que el uso de una correlación inapropiada.

Capítulo 5

Error en la medición de la presión y de los parámetros PVT

5.1. Introducción

Siempre que se mide un fenómeno se hace hincapié en maximizar la exactitud y precisión. La calidad de la medición dependerá de la exactitud y precisión del sistema de adquisición de datos (que viene dado por la resolución del instrumento y otros factores) [23].

Ahora bien, se debe diferenciar entre exactitud y precisión. La *exactitud* de una medición representa cuán cerca está la medición del valor real de la cantidad que se desea medir. Por tanto, ésta representa que tan correcto es el resultado. La *precisión* de una medición describe cuán exacto se determinó el resultado sin hacer referencia a lo que el resultado significa. Por último, otro término que es importante definir se refiere a la *precisión relativa*, la cual indica la incertidumbre en una medición como una fracción del resultado.

Estos conceptos se entienden mejor con el siguiente ejemplo. Suponga que se está midiendo la altura de una mesa con una cinta métrica y dió que la longitud fue de 1.502 metros. Esto representa que la cinta métrica (además de sus ojos) pudo leer distancias de al menos un milímetro. Para la mayoría de los casos esto se considera bastante preciso, con una precisión relativa alrededor de 1/1500. Ahora suponga que midió de nuevo la altura de la mesa y obtuvo 1.510 metros. Después de reflexionar acerca de los resultados obtenidos se dió cuenta de que la técnica de medición para la primera lectura no era la correcta, porque la cinta métrica no estaba completamente vertical. Por tanto la primera medición no fue exacta.

En la industria petrolera es muy importante diferenciar entre estos términos debido a que puede ayudar en la obtención de mediciones de mejor calidad. Por ejemplo, la exactitud de un instrumento se puede mejorar notablemente si se calibra adecuadamente según las indicaciones del fabricante (como recordará esto

permitirá que el resultado se acerque más al valor real de la magnitud que se quiere medir). La precisión está relacionada con el instrumento en sí, es decir, los dispositivos que utiliza éste para realizar la medición (algo que es más difícil controlar por el ingeniero ya que debería conocer la precisión del dispositivo de medición, o si está a disponibilidad, elegir entre los diferentes tipos de dispositivos de medición con sus respectivas precisiones).

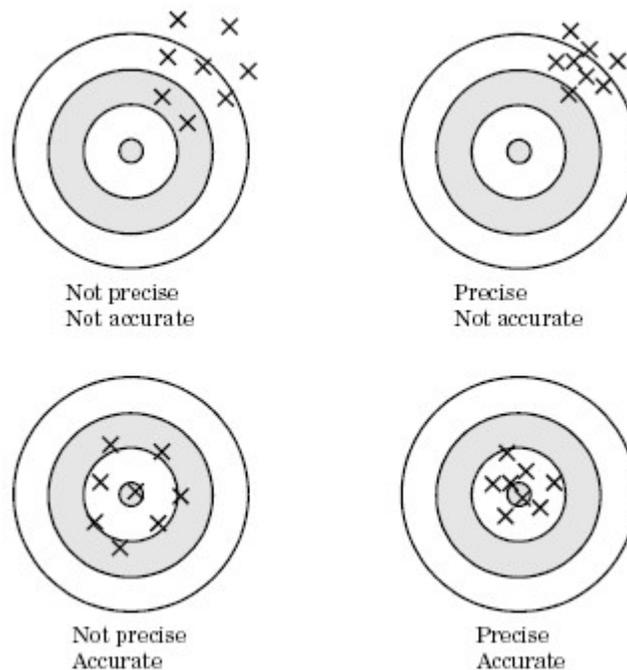


Figura 5.1: Exactitud y precisión. Data Acquisition Toolbox User's Guide. The Mathworks, Inc. 2002. Página 1-29

5.2. Error en la medición de presión

Las herramientas actuales que utilizan medidores de cuarzo poseen resoluciones alrededor de [0.1 psi - 0.01 psi], y la precisión de [0.01 % - 0.05 % del valor de la lectura]. Esto implica que para medir una presión de 10000 psi se tendría un error asociado con la precisión del instrumento de [1 psi - 5 psi] ([24, 25, 26]). Sin embargo, hay que recordar que la historia de declinación de presión posee mediciones hechas en el pasado, y a medida que avanza la tecnología los errores asociados con la medición de presión van disminuyendo. Recordando investigacio-

nes hechas acerca de las incertidumbres en las mediciones de presión se pueden citar varios autores. McEwen(1961) [1] reporta que errores de [10 psi - 20 psi] para una medición de presión de 5000 psi eran bastante modestos para esa época. La incertidumbre de presión que tomó Tehrani(1976) [2] fue de [+/- 5 psi]. Wang y Hwan(1997) [3] tomaron incertidumbres de presión de [+/- 25 psi]. Walsh(1999) [4] menciona que incertidumbres de presión de [100 psi] no son inusuales; en su trabajo tomó hasta [200 psi] de incertidumbre en la presión.

5.3. Error en la medición de la gravedad API del petróleo

La mayoría de los instrumentos disponibles en el laboratorio para medir la gravedad API tienen resoluciones que varían de entre [0.1 °API - 0.2 °API], la precisión también está en ese orden, es decir, [0.1 °API - 0.2 °API]. Ahora bien, en el laboratorio es importante conocer la *reproducibilidad*, es decir la diferencia entre dos resultados independientes, obtenidas por diferentes operadores trabajando en diferentes laboratorios con el mismo material de medición. La reproducibilidad del método de medición estándar para la gravedad API se encuentra alrededor de [0.5 °API] [27].

5.4. Error en la medición de la gravedad específica del gas

De las herramientas consultadas para la medición de la gravedad específica del gas las resoluciones se encontraron estar alrededor de [0.02 % de la lectura realizada], y la precisión de los instrumentos en el siguiente rango [0.1 % - 1 % de la lectura] ([28, 29]).

5.5. Error en la medición de la temperatura del yacimiento

En la referencia [30] se explican los métodos actuales de medición de temperatura del yacimiento, además de un breve recuento histórico y las aplicaciones de esta medición. Explica que aunque las resoluciones están en el orden de [0.0001 °C - 0.001 °C], las precisiones de los equipos varían en el rango [0.02 °C - 0.05 °C]. La exactitud efectiva de los registros de temperatura es de +/- 0.5 °C ([30, 31, 32]).

5.6. Error en la medición de la relación gas-petróleo en solución

La medición del gas en solución en el laboratorio se realiza mediante dos procedimientos. Estos son liberación diferencial y la prueba del separador (también llamada liberación flash). En la liberación diferencial el gas se remueve del petróleo como este sale de solución. Para la prueba del separador todo el gas se encuentra en contacto con todo el petróleo hasta alcanzar el equilibrio entre las dos fases. Estos dos procesos se llevan a cabo en una celda PVT. Por tanto el error en la medición esta asociado con los dispositivos que se utilicen para medir los volúmenes en la celda ([33, 34, 10]). Bu y Damsleth [35] mencionan que la incertidumbre relativa experimental se encuentra en el rango [1 % - 2 %] de la lectura realizada.

Capítulo 6

El Programa EAM (Error Analysis Module)

6.1. Inicializando la interfase

Para cargar la interfase primero, se ubica el directorio del programa en el “current directory” de MATLAB; segundo, se escribe en la línea de comandos “erroranalysis” para inicializar la interfase. Aparecerá una ventana como en la figura 6.1.

En la pantalla principal se puede observar la ubicación de los archivos que se han cargado y el nombre bajo el cual se guardaron. Se puede ver los archivos de producción (*.prod), de pvt (*.pvt), de acuífero (*.aq), y de datos del yacimiento (*.res). También, cuando se guarda un caso, se muestra una previsualización del archivo debajo del cuadro de texto “Case name”. Los botones en la pantalla principal permiten acceder a ventanas relacionadas con: cargar los datos, calcular realizaciones, ó calcular diseños experimentales. Dependiendo de que datos se carguen, se podrá acceder a uno u otro. En la ventana de **Log Window** se muestra algunas de las operaciones realizadas, como por ejemplo mostrar que modelo de acuífero se va a utilizar. Por último la barra de herramientas muestra algunos de los botones convencionales como: nuevo caso, abrir caso, guardar caso, ayuda; y los restantes son los que aparecen en la pantalla principal.

6.2. Menú File

Este menú permite abrir, guardar ó cerrar un caso, y por supuesto, salir del programa. Los primeros cinco submenús están relacionados con abrir un caso nuevo (**New Case**), un caso existente (**Open Case**) o para guardar el caso actual (**Save Case** y **Save Case As...**), y cerrar el caso actual (**Close Case**). El submenú **Clear Historial** borra el contenido de la ventana **Log Window**. El último submenú, **Close**, cierra el programa; antes, mostrando una ventana preguntando si realmente desea salir del programa.

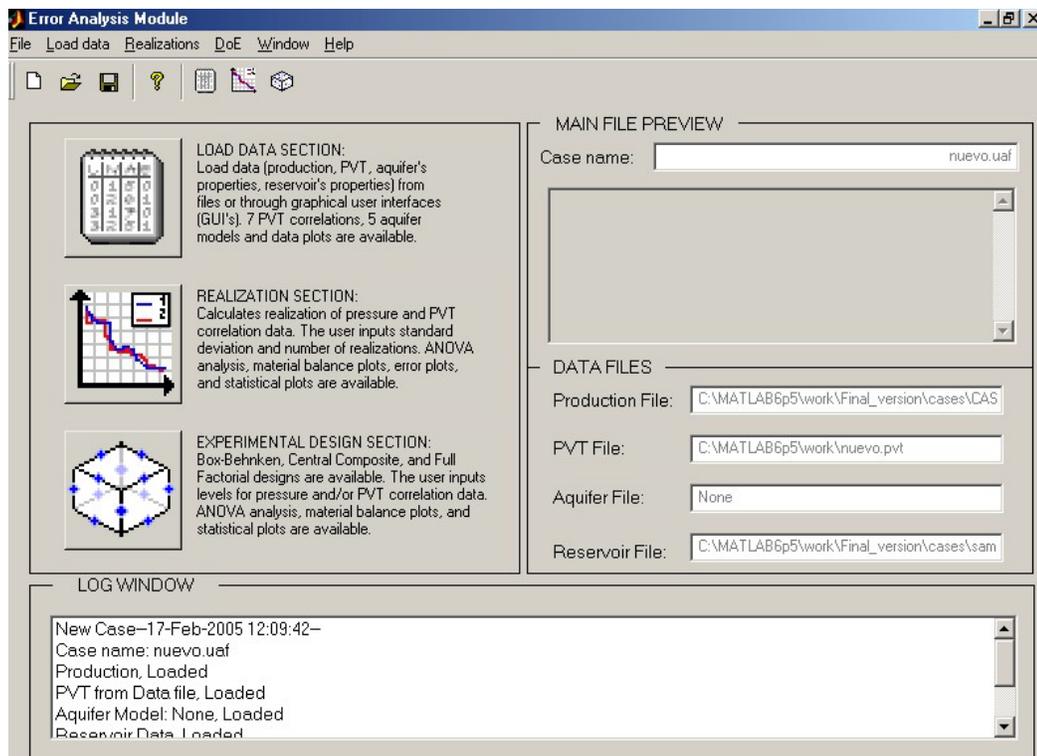


Figura 6.1: Pantalla principal de la interfase

6.3. Cargando y visualizando los datos

Los datos se pueden cargar a través del menú **Load data** o a través del botón **Load data-View data** (con forma de bloc de notas) en la pantalla principal. En la figura 6.2 se muestra la ventana que aparece al hacer clic en el botón **Load data-View data**.

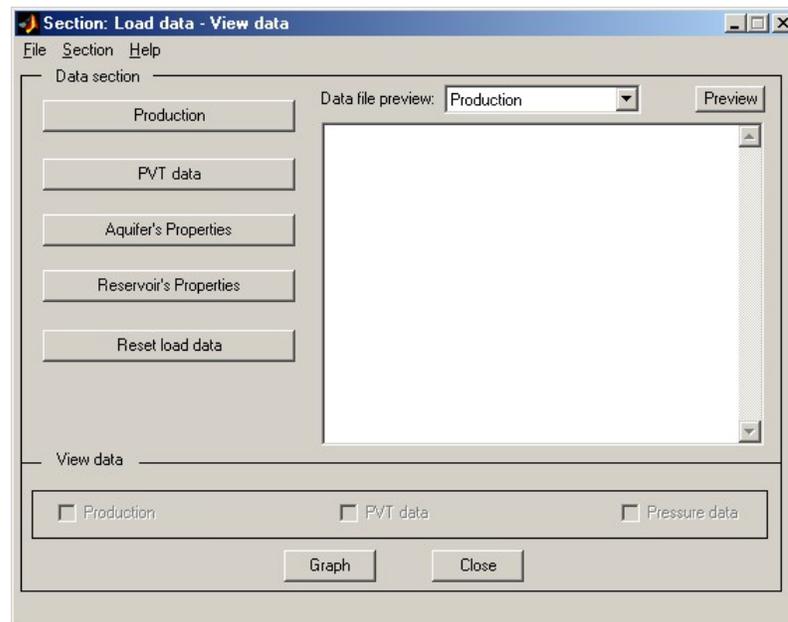


Figura 6.2: Ventana **Load data-View data**

Los botones en el lado izquierdo permiten cargar las opciones correspondientes. El último botón *Reset load data* permite volver a cargar algún dato en caso de haber cometido alguna equivocación o si se quiere simplemente cambiar los datos. Al hacer clic en este botón aparece la ventana de la figura 6.3.

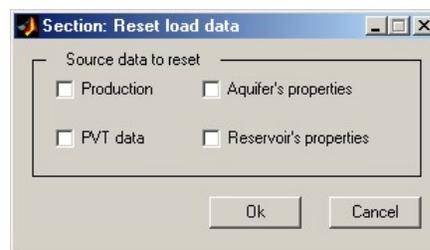


Figura 6.3: Ventana **Reset load data**

Después de que se cargan los datos exitosamente, en la parte inferior, en la sección de **View data**, se activan las opciones correspondientes, de acuerdo al tipo de datos cargados. Por ejemplo, después de cargar los datos de producción, se harán disponibles las opciones *Production* y *Pressure data*. También el botón *Graph* se hará disponible. Otras acciones que se pueden realizar en esta ventana es ver una previsualización del archivo que se cargó o generó. Esto se logra seleccionando la opción requerida en *Data file preview* y hacer clic en el botón *Preview*.

También desde el menú **Load data** se cargan los archivos necesarios para correr la interfase, estos consisten de archivos de producción (*.prod), archivos de pvt (*.pvt), archivos de acuíferos (*.aq), y archivos de las características del yacimiento (*.res); o se pueden introducir los datos a través de las cajas de dialogo correspondientes. Los submenús de que se disponen son **Production**, **PVT**, **Aquifer Model**, **Reservoir Data**, **Reset Load**, y **View data**.

6.3.1. Cargando los datos de producción

El submenú **Production** permite cargar los archivos de producción (*.prod) a través de la caja de dialogo abrir de Windows. En la figura 6.4 se puede observar el formato de un archivo PROD.

```
% PRODUCTION FILE
%
%Time[Year] Pressure[psia] Np [STB] Gp [SCF] Wp [STB]
0 4240.10 0.0000e+000 0.0000e+000 0.0000e+000
1 4116.90 9.8550e+005 1.8516e+009 1.2473e-001
2 4006.06 1.9710e+006 3.4905e+009 4.4411e-001
3 3896.18 2.9565e+006 5.1602e+009 9.9208e-001
4 3779.45 3.9420e+006 7.1044e+009 1.7948e+000
5 3650.89 4.9275e+006 9.5017e+009 2.9454e+000
6 3502.87 5.9130e+006 1.2451e+010 4.5818e+000
7 3332.98 6.8985e+006 1.6246e+010 6.9222e+000
8 3159.26 7.8840e+006 2.0817e+010 9.7513e+000
9 2924.77 8.8695e+006 2.7394e+010 1.4012e+001
10 2670.36 9.8550e+006 3.6834e+010 1.9791e+001
```

Figura 6.4: Archivo *.PROD

Los archivos PROD son del tipo delimitado por tabulaciones. La estructura de los archivos PROD es la descrita a continuación: la primera columna representa el tiempo y viene en años, la segunda columna es la presión y sus unidades son libras

por pulgada cuadrada (psia), la tercera columna es la producción acumulada de petróleo (N_p) y viene en barriles estándar (STB), la cuarta columna representa el gas acumulado producido (G_p) y sus unidades son pies cúbicos estándar (SCF), y la quinta y última columna es el agua acumulada producida (W_p) y sus unidades son millones de barriles estándar (STB). Este tipo de archivo se puede generar fácilmente con cualquier editor de texto, guardándolo bajo la extensión *.prod. La primera línea representa el encabezado del archivo. El signo de % indica que esa línea no será leída por el programa. Estos comentarios son igualmente válidos para los archivos de pvt, de acuíferos, y de datos del yacimiento.

6.3.2. Cargando los datos PVT

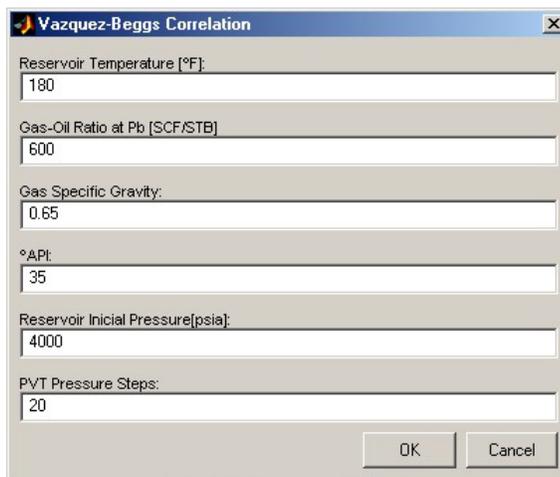
El submenú PVT se divide en **From File** y **Correlations**. **From File** permite cargar los datos PVT directamente desde un archivo (*.pvt) a través de la caja de dialogo abrir.

```
%PVT FILE
%Pressure [psia] Bo [bbl/STB] Bg [bbl/SCF] Rs [SCF/STB]
4500.00 1.76567 0.00074 1476.00000
4275.73 1.77423 0.00075 1476.00000
4240.08 1.77568 0.00075 1476.00000
3827.20 1.69525 0.00078 1306.99418
3602.94 1.65222 0.00080 1216.59195
3378.68 1.60970 0.00082 1127.23650
3154.41 1.56770 0.00085 1038.98404
2930.15 1.52625 0.00088 951.89801
. . . .
. . . .
. . . .
```

Figura 6.5: Archivo *.PVT

El archivo PVT es del tipo delimitado por tabulaciones. La estructura se describe a continuación: la primera columna es la presión en libras por pulgada cuadrada (psia), la segunda columna es el B_o y viene dado en barril de yacimiento por barril estándar (bbl/STB), la tercera columna es el B_g y sus unidades son barriles de yacimiento por pies cúbicos estándar (bbl/SCF), y la última columna es el R_s y viene en pies cúbicos estándar por barril estándar (SCF/STB). La figura 6.5 muestra el formato de un archivo PVT. **Correlations** permite elegir una correlación para generar el PVT. Las correlaciones PVT disponibles son Al-Marhoun, Dindoruk.-Christman, Glasø, Lasater, Petrosky-Farshad, Standing, y

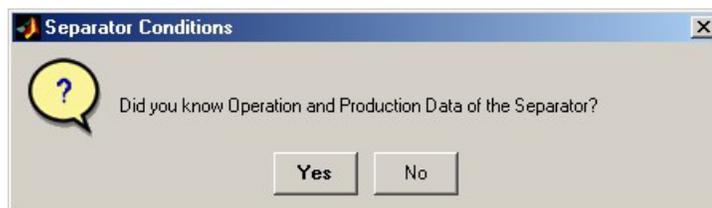
Vazquez-Beggs. Al seleccionar una correlación PVT aparece la caja de dialogo de la figura 6.6.



The image shows a software dialog box titled "Vazquez-Beggs Correlation". It has a standard Windows-style title bar with a close button. The dialog contains several input fields with the following values: "Reservoir Temperature [°F]" is 180, "Gas-Oil Ratio at Pb [SCF/STB]" is 600, "Gas Specific Gravity" is 0.65, "°API" is 35, "Reservoir Initial Pressure [psia]" is 4000, and "PVT Pressure Steps" is 20. At the bottom right, there are two buttons labeled "OK" and "Cancel".

Figura 6.6: Caja de dialogo para las correlaciones PVT

Esta caja de dialogo permite la entrada de los parámetros necesarios para generar el PVT para la correlación seleccionada. Los parámetros que deben introducirse son: temperatura del yacimiento en grados Fahrenheit ($^{\circ}$ F), la relación gas-petróleo en solución a la presión de burbujeo con unidades de pies cúbicos estándar de gas por barril estándar de petróleo (SCF/STB), la gravedad específica del gas, la gravedad API del petróleo, la presión inicial para el PVT a generar en libras por pulgada cuadrada (psia), y el paso de presión para generar el PVT, este consiste del número de valores de presión a los que se desea reportar el PVT. Para las correlaciones PVT de Al-Marhoun, Glasø, Lasater y Vazquez-Beggs, saldrá el mensaje de la figura 6.7, preguntando si las condiciones del separador son conocidas.



The image shows a software dialog box titled "Separator Conditions". It has a question mark icon in a speech bubble on the left. The text in the dialog asks "Did you know Operation and Production Data of the Separator?". At the bottom, there are two buttons labeled "Yes" and "No".

Figura 6.7: Mensaje emergente preguntando acerca de las condiciones de operación del separador

Si se selecciona **Yes**, aparece la caja de dialogo de la figura 6.8, en la cual se pueden introducir los valores de la temperatura del separador en grados Fahrenheit ($^{\circ}\text{F}$), y la presión del separador en libras por pulgada cuadrada (psia). Si en cambio, en el mensaje de las condiciones del separador se selecciona **No**, se asigna a la temperatura del separador (80°F) y a la presión del separador 60 (psia).



Figura 6.8: Caja de dialogo para las condiciones de operación del separador

6.3.3. Cargando el modelo de acuífero

El modelo de acuífero se puede cargar de un archivo existente (*.aq), o a través de las cajas de diálogo para cada uno de los modelos disponibles. También es posible seleccionar **None** y establecer que el yacimiento no tenga acuífero. Los modelos disponibles de acuíferos son: Pot Aquifer, Schiltuis, Hurst, Fetkovich, y Carter-Tracy. Es importante decir que cada opción genera un archivo de acuífero, por tanto existen cinco formatos diferentes del archivo AQ que dependerá del modelo elegido o del que se utilizará.

La caja de dialogo para el modelo de Pot Aquifer se presenta a continuación:

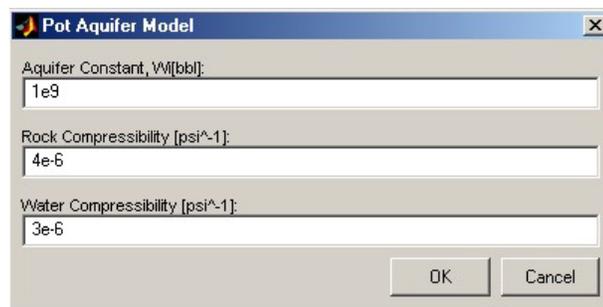


Figura 6.9: Caja de dialogo para el modelo Pot Aquifer

En ella se introduce los datos correspondientes a la constante del acuífero en barriles de yacimiento (bbf), la compresibilidad de la roca en (psi^{-1}), y la compresibilidad del agua en (psi^{-1}). El archivo que se genera tiene el formato mostrado en la figura 6.10.

```
% Aquifer Data File
% Aquifer Model
POT
% Aquifer Constant, Wf[bbf]
1e9
% Rock Compressibility [psi^-1]
4e-6
% Water Compressibility [psi^-1]
3e-6
```

Figura 6.10: Archivo AQ para el modelo Pot Aquifer

Para los modelos de Schiltuis y Hurst en la caja de dialogo se introduce, solamente, la constante del acuífero en unidades de barriles por día por libras por pulgada cuadrada (bbl/d/psi):



Figura 6.11: Caja de dialogo para los modelos de Hurst y Schiltuis

El archivo AQ que se genera para los modelos de Hurst y Schiltuis es mostrado en la figura 6.12.

```
% Aquifer Data File
% Aquifer Model
SCHILTUIS
% Aquifer Constant [bbl/d/psi]
200
```

Figura 6.12: Archivo AQ para los modelos de Schiltuis, y Hurst

La caja de dialogo del modelo de Fetkovich se muestra en la figura 6.13. En ella se introducen el siguiente conjunto de parámetros: compresibilidad de la roca en (psi^{-1}), compresibilidad del agua en (psi^{-1}), viscosidad del agua en centipoise (cp), radio externo del yacimiento en pies (ft), permeabilidad del acuífero en md, porosidad del acuífero en fracción, y espesor neto del acuífero en pies (ft). Dependiendo del tipo de flujo y la geometría del acuífero se deben introducir los datos correspondientes al ancho del acuífero en pies (ft), y largo del acuífero en pies (ft), para geometría lineal; y radio externo del acuífero en pies (ft), radio externo del yacimiento en pies (ft), y ángulo de invasión (encroachment angle, $^{\circ}$), para geometría radial.

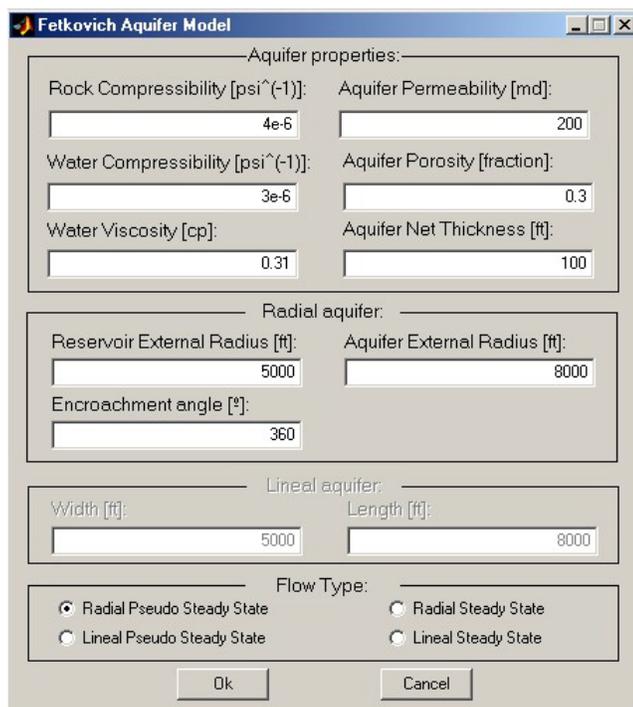


Figura 6.13: Caja de dialogo para el modelo de Fetkovich

El archivo AQ para el modelo de Fetkovich tiene el siguiente formato:

```
% Aquifer Data File
% Aquifer Model
FETKOVICH
% Rock Compressibility [psi^-1]
4e-006
% Water Compressibility [psi^-1]
3e-006
% Water Viscosity [cp]
0.31
% Aquifer Permeability [md]
200
% Aquifer Porosity [fraction]
0.3
% Aquifer Net Thickness [ft]
100
% Reservoir Flow Type
RPSS
% Aquifer External Radius [ft]
8000
% Reservoir External Radius [ft]
5000
% Encroachment Angle [degrees]
360
```

Figura 6.14: Archivo AQ para el modelo de Fetkovich

La caja de dialogo del modelo de Carter-Tracy se muestra en la figura 6.15. Esta permite introducir el siguiente conjunto de parámetros: compresibilidad de la roca en (psi^{-1}), compresibilidad del agua en (psi^{-1}), viscosidad del agua en centipoise (cp), radio externo del yacimiento en pies (ft), el ángulo de invasión en grados, permeabilidad del acuífero en md, porosidad del acuífero en fracción, espesor neto del acuífero en pies (ft), y la tasa entre el radio del acuífero y el radio del yacimiento.

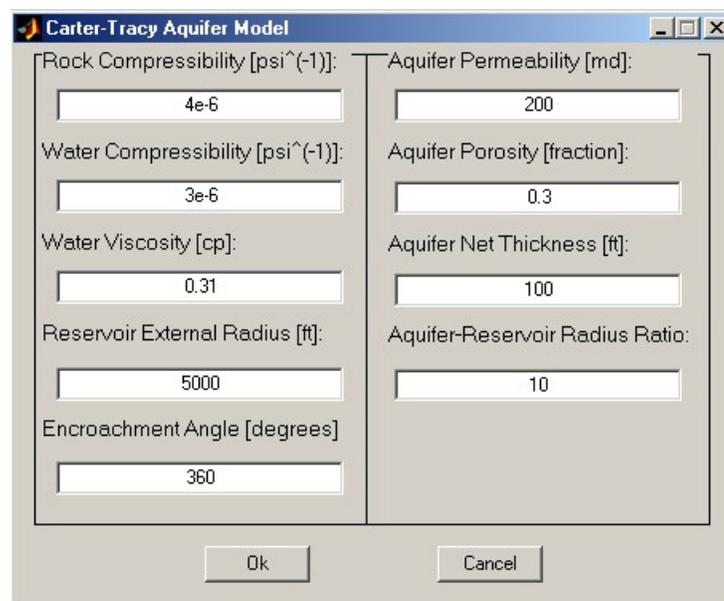


Figura 6.15: Caja de dialogo para el modelo de Carter-Tracy

El archivo AQ para este modelo posee el formato mostrado en la figura 6.16.

```
% Aquifer Data File
% Aquifer Model
CARTER-TRACY
% Rock Compressibility [psi^-1]
4e-6
% Water Compressibility [psi^-1]
3e-6
% Water Viscosity [cp]
0.31
% Aquifer Permeability [md]
200
% Aquifer Porosity [fraction]
0.3
% Aquifer Net Thickness [ft]
100
% Reservoir External Radius [ft]
5000
% Encroachment Angle [degrees] [0<F<1]
360
% Aquifer/Reservoir Radius Ratio
10
```

Figura 6.16: Archivo AQ para el modelo de Carter-Tracy

6.3.4. Cargando los datos del yacimiento

El siguiente submenú es el correspondiente a **Reservoir Data**, en él se cargan los datos correspondientes a la compresibilidad de la roca en (psi^{-1}), la compresibilidad del agua en (psi^{-1}), y la saturación del agua inicial en fracción, también permite seleccionar si el yacimiento posee capa de gas o no, y en caso positivo, permite especificar el tamaño de la capa de gas.

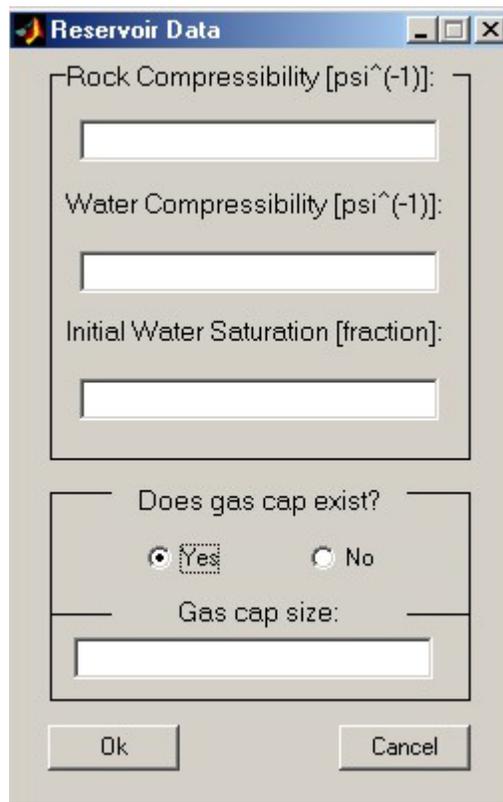


Figura 6.17: Caja de dialogo para los datos del yacimiento

El archivo RES posee el siguiente formato:

```
% Reservoir Data File
% Rock Compressibility [psi^-1]:
3e-6
% Water Compressibility [psi^-1]:
4e-6
% Initial Water Saturation [fraction]:
0.12
% Does gas cap exist?:
1
% Gas cap size:
0.25
```

Figura 6.18: Archivo *.RES

El siguiente submenú es **Reset Load** y se utiliza en casos donde se ha cargado erróneamente algún dato (por ejemplo un archivo PROD, PVT, AQ, y/o RES) o en la entrada mediante las cajas de diálogos. Específicamente lo que hace este submenú es que vuelve a poner a disposición el submenú elegido. Se puede seleccionar uno en específico o poner a disposición todos a la vez, seleccionando la opción **All parameters**.

6.3.5. Visualización de los datos cargados de producción y PVT

De esto se encarga el último submenú **View data**. Hay 3 submenús disponibles: **Production Data**, **PVT Data**, y **Reservoir Pressure Profile**. Cada submenú se explica a continuación:

Production Data: Este submenú se activa cuando se cargan los datos de producción. Permite visualizar la historia de producción por medio de dos gráficos. El primer gráfico es la producción acumulada. Permite observar la producción acumulada de gas, petróleo y agua en el mismo gráfico. El eje horizontal representa el tiempo en años, el eje de la izquierda es la producción líquida, y el de la derecha es la producción de gas. La producción líquida viene en unidades de millones de barriles estándar (MMSTB), y permite ubicar en el mismo eje la producción de petróleo y agua. Las unidades de la producción de gas son mil millones de pies cúbicos estándar (MMSCF). El segundo gráfico es la producción promedio por año calculada a partir de la historia de producción. Nuevamente, el eje horizontal es el tiempo en años, el eje de la izquierda es la producción líquida en barriles estándar por día (STB/day), y el eje de la derecha es la producción de gas en pies cúbicos estándar por día (SCF/day).

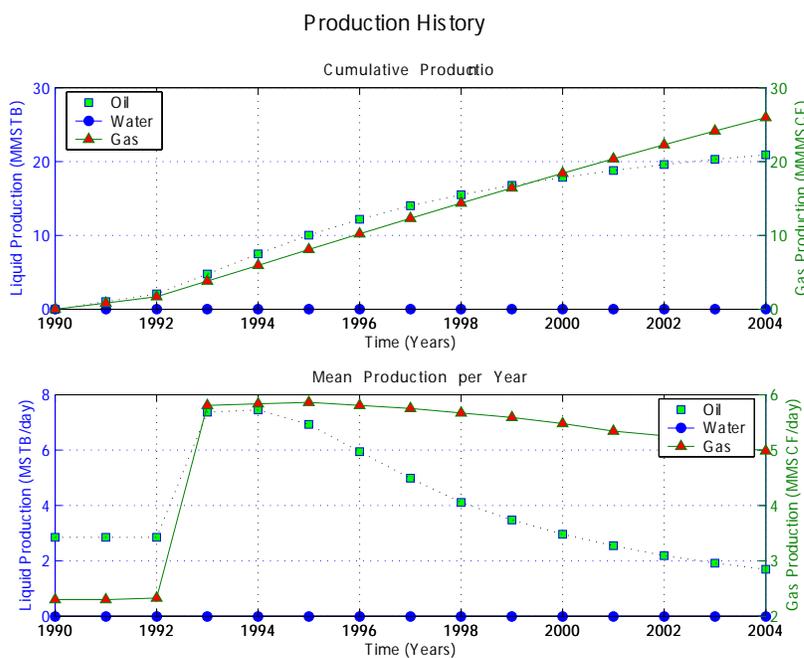


Figura 6.19: Visualizando los datos de producción

PVT Data: en este submenú se puede observar los datos PVT cargados o generados a partir de la correlación escogida. Particularmente, los gráficos que se muestran son: (de izquierda a derecha y de arriba a abajo) el primero es el correspondiente a *Oil FVF*, las unidades de B_o son bbl/STB y las correspondientes a la presión son psia, el segundo gráfico es *Gas FVF*, las unidades de B_g son bbl/SCF, y por último el de *Solution Gas-Oil Ratio*, con R_s en SCF/STB.

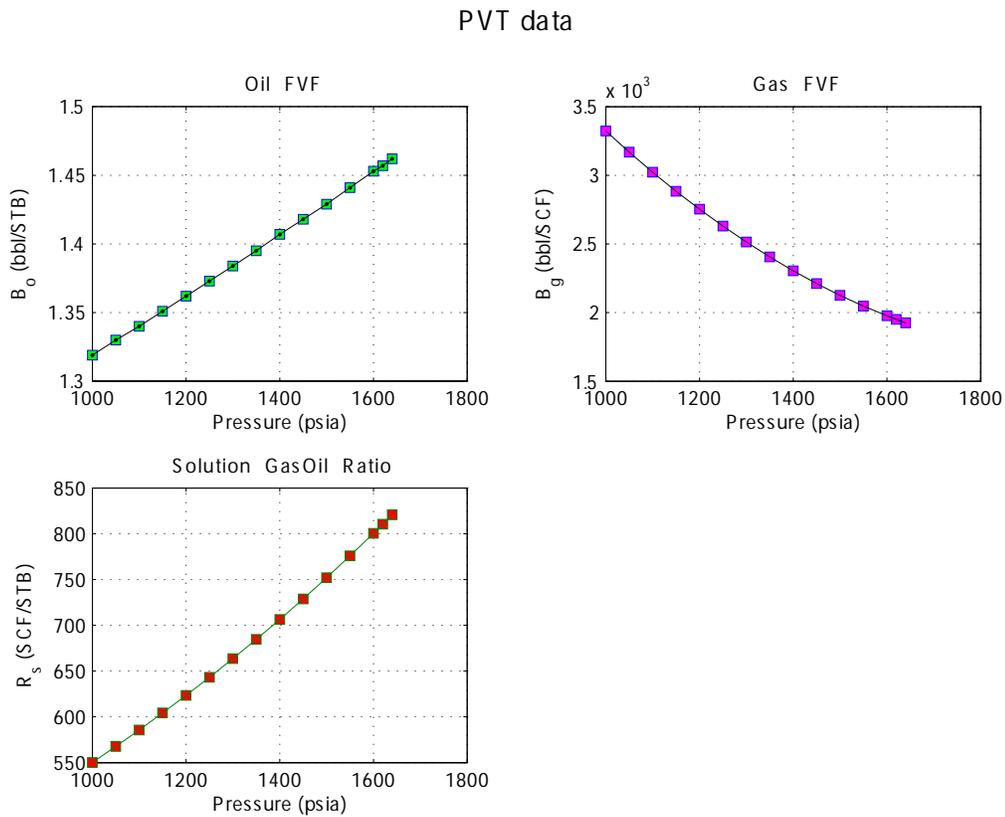


Figura 6.20: Visualizando los datos PVT

Reservoir Pressure Profile: en este submenú se puede observar el perfil de declinación de presión de la historia de producción. El eje vertical es la presión en psia, y el eje horizontal es el tiempo en años.

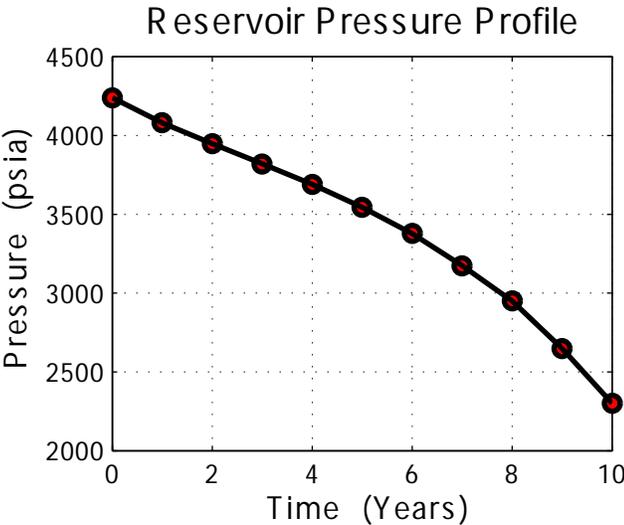


Figura 6.21: Visualizando el perfil de presiones del yacimiento

6.4. Realizaciones de presión y PVT

6.4.1. Introduciendo los parámetros para el cálculo de las realizaciones

Se puede acceder a esta sección mediante el segundo botón de la pantalla principal **Realizations** (con forma de una gráfica) o a través del menú **Realizations-Parameters**. En la figura 6.22 se muestra la ventana correspondiente a la sección **Realizations**.

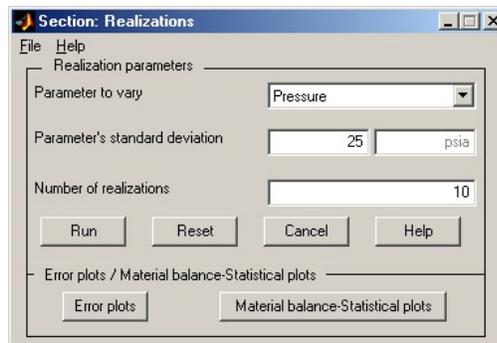


Figura 6.22: Ventana **Realizations**

En esta ventana se eligen, primero, el parámetro que se desea variar; luego, la desviación estándar y el número de realizaciones, y posteriormente se hacen los cálculos de realizaciones y balance de materiales. En la parte inferior de la ventana se pueden acceder a los gráficos resultantes de los cálculos realizados. Hay 2 tipos, los gráficos estadísticos y de balance de materiales, y los de error. El primer tipo de gráfico se obtiene con una simple corrida, mientras que para los gráficos de error se necesitan realizar varias corridas para obtener el gráfico requerido.

El menú **Realizations-Parameters** tiene los siguientes submenús: **Standard Deviation**, **Number of Realizations**, **Parameter to vary**, **Reset**, **Calculate Realizations**, **Results** y **Error Plots**.

Después de haber cargado los datos correspondiente a la producción y el PVT se puede pasar a este menú, aunque es recomendable terminar de cargar enteramente los datos del menú **Load data**. El submenú **Parameter to vary** permite seleccionar cuál es el parámetro que se desea variar, éste a la vez posee otros submenús, **Pressure** y **PVT**. Cuando se carga el PVT desde un archivo, es decir, se carga directamente los valores de las propiedades PVT, el submenú **PVT** de **Parameter to vary** se desactiva. La única opción disponible es **Pressure**. Sin embargo, cuando se genera el PVT a partir de alguna de la correlaciones, la opción

PVT se encuentra activada, permitiendo seleccionar entre las siguientes opciones: **Solution Gas Oil Ratio**, **Gas Specific Gravity**, **Reservoir Temperature**, y **Tank Oil API**.

Luego que se selecciona un parámetro, el submenú **Standard Deviation** se activa. Este submenú permite introducir la desviación estándar del parámetro que se quiere variar.



Figura 6.23: Caja de dialogo **Parameter standard deviation**

El siguiente submenú es **Number of Realizations**. Este submenú permite introducir el número de realizaciones que se harán con los parámetros seleccionados anteriormente.

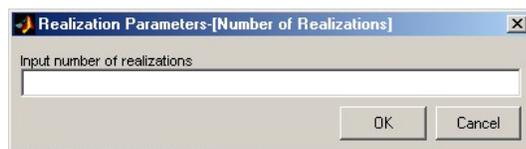


Figura 6.24: Caja de dialogo **Number of Realizations**

El siguiente submenú corresponde a la opción **Reset**, y permite activar algunos de los submenús desactivados o todos a la vez (mediante la opción **All Parameters**), con el fin de corregir algún error en la entrada de los datos.

El submenú **Calculate Realizations** realiza los cálculos de balance de materiales con el número de realizaciones indicadas, con los parámetros introducidos, y con los archivos cargados o generados. Luego que se ha seleccionado esta opción aparece un mensaje indicando el tiempo remanente de espera, y el porcentaje de cálculos realizados.

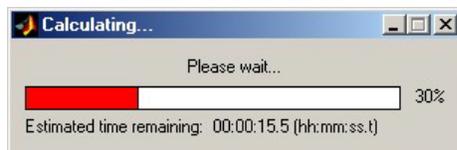


Figura 6.25: Ventana mostrando el tiempo remanente de espera y el porcentaje de cálculos realizados

6.4.2. Visualizando los resultados de las realizaciones

Los resultados se pueden observar mediante los botones *Error plots* y *Material balance-Statistical plots* de la ventana *Realizations* o a través del submenú **Results**. Los submenús de los que consta son: **One-Way ANOVA**, **Material Balance Plots**, **pdf's-CDF's**, y **Realizations**.

Botón *Material balance-Statistical plots*

La ventana para acceder a los gráficos de balance de materiales y los gráficos estadísticos se muestran en la figura 6.26. Cada botón permite ver el gráfico que éste indica. La función de cada botón se explica de acuerdo al submenú correspondiente.

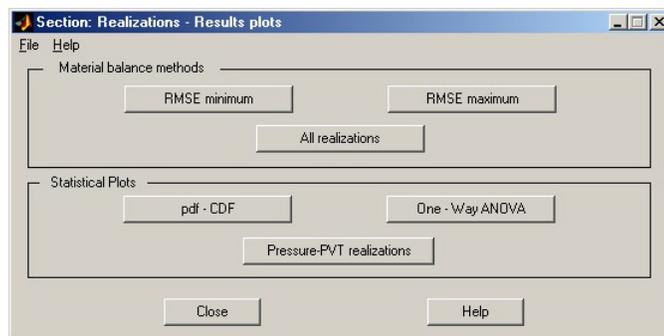


Figura 6.26: Ventana **Realizations - Results Plots**

Submenú Results

El submenú **One-Way ANOVA** realiza un análisis de varianza para los datos cuando se disponen de dos o los tres métodos para los cálculos de balance de materiales. Cuando se selecciona esta opción aparece una ventana como la figura 6.27.

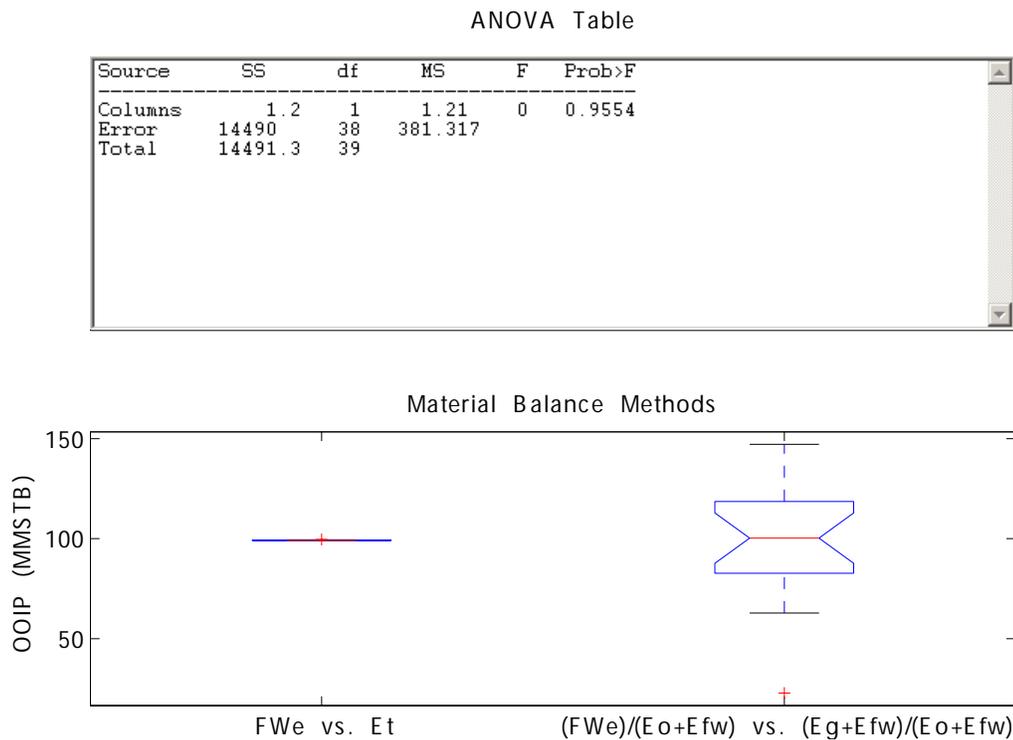


Figura 6.27: Tabla y diagrama de caja para One-Way ANOVA

La tabla muestra información acerca del análisis de varianza realizado en el conjunto de realizaciones de balance de materiales. La segunda columna indica la suma de cuadrados, la tercera columna indica los grados de libertad, la cuarta representa el error promedio (se obtiene dividiendo la segunda columna entre la tercera), la quinta columna es el valor F calculado, y la última columna representa el valor p de la prueba. El gráfico muestra un diagrama de caja para los métodos de balance de materiales a los que se le aplicó la prueba. En el eje vertical se encuentra el POES en MMSTB y en el eje horizontal se encuentran los distintos métodos utilizados.

El siguiente submenú es **Material Balance Plots**, el cual a su vez se divide en **All realizations**, **Minimum RMSE**, y **Maximum RMSE**. **All realizations** permite visualizar, para los métodos de balance de materiales escogidos, todas las realizaciones calculadas. La figura 6.28 aparece cuando se selecciona esta opción, permite seleccionar cual o cuáles de los tres métodos se va a visualizar.

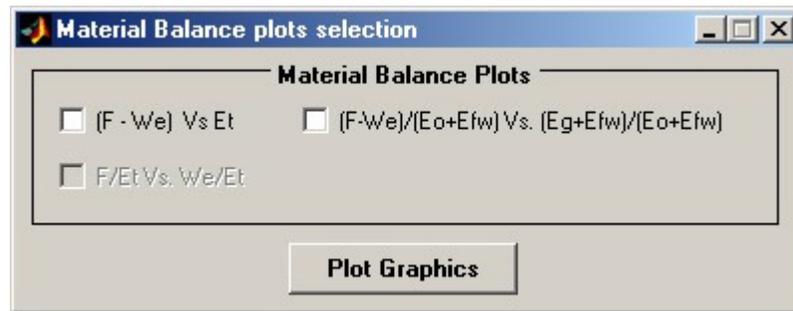


Figura 6.28: Ventana de selección para escoger los métodos que se van a graficar

Luego que se ha escogido los métodos de balance de materiales, se obtiene un gráfico como el mostrado en la figura 6.29.

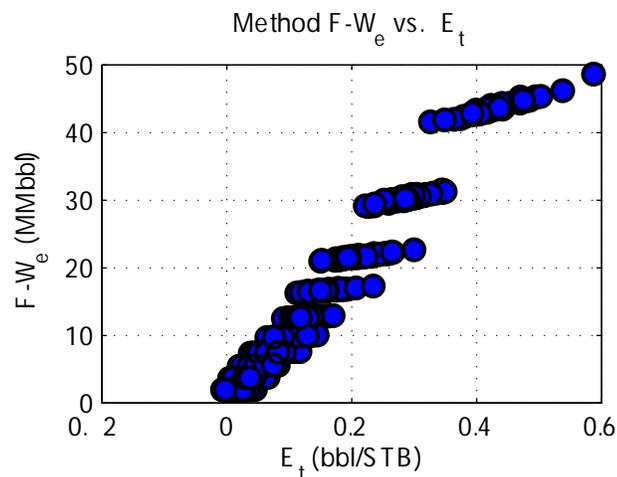


Figura 6.29: Métodos de balance de materiales mostrando todas las realizaciones calculadas

Para el submenú de **Minimum RMSE**, se selecciona de entre todas las realizaciones aquella que tuvo el menor valor para el RMSE (la raíz cuadrada del promedio del cuadrado de los errores). Antes de que aparezca la figura, primero se muestra la siguiente ventana de selección donde, además de seleccionar los métodos a graficar, permite seleccionar la opción **Energy Plot** para visualizar los gráficos correspondientes a los índices de producción.

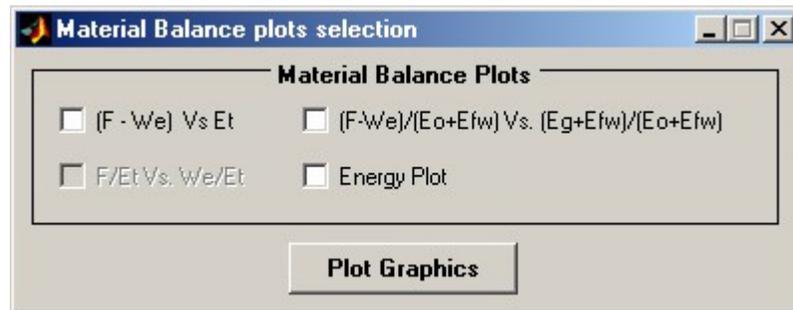


Figura 6.30: Ventana de selección con la opción Energy Plot, disponibles para los submenús Minimum y Maximum RMSE

Posteriormente se podrán ver los gráficos que aparecen en la figura 6.31. Como se puede observar, cada gráfico dispone de una caja flotante con información acerca de la regresión realizada para ese método de balance de materiales en específico. Particularmente, se observan el valor del POES (N), un intervalo de confianza para el POES (C), y el RMSE asociado a los cálculos. Cuando se selecciona **Maximum RMSE** el procedimiento es el mismo, sólo que se selecciona aquella realización que tiene el mayor valor de RMSE.

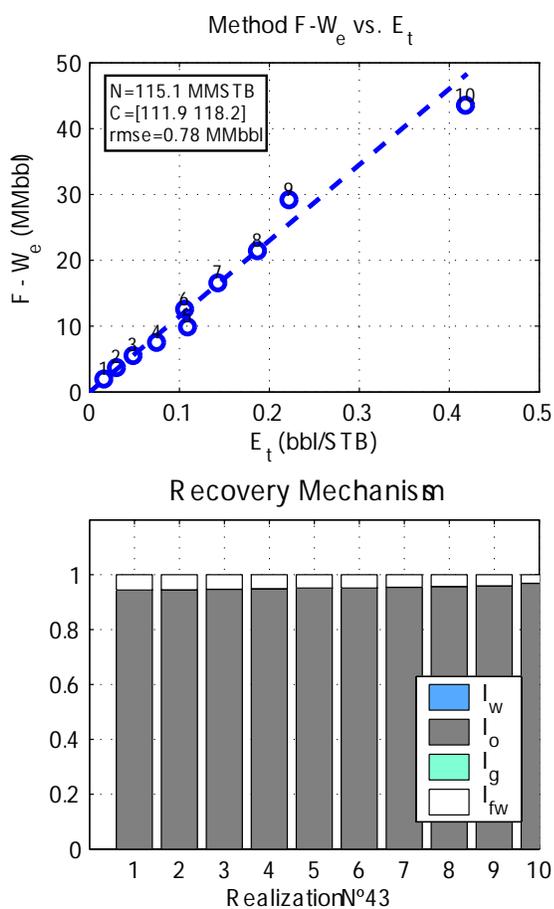


Figura 6.31: Gráfico de los métodos de balance de materiales junto con su correspondiente gráfico de los índices de producción

El siguiente submenú a explicar es **pdf's-CDF's**. Este submenú muestra un gráfico de la función de densidad de probabilidad o *probability density function* (pdf, en verde, y en el eje izquierdo), y de la función de densidad acumulada o *Cumulative Density Function* (CDF, la curva escalonada en azul con el eje a la derecha). Las curvas rojas representan la función de densidad de probabilidad y de densidad acumulada para una distribución normal calculada a partir de los datos, es decir, las curvas que se obtendrían para una muestra o población normal. El eje horizontal representa el POES en MMSTB. Además, este gráfico también posee una caja flotante con información acerca del promedio de los datos, la mediana, la desviación estándar de los datos, el mínimo y el máximo.

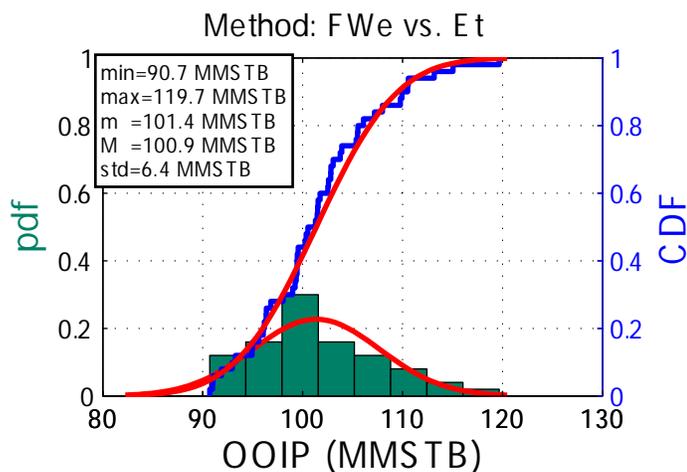


Figura 6.32: Ejemplo del gráfico pdf's-CDF

Por último se encuentra el submenú **Realizations** el cual permite visualizar las realizaciones. Si el parámetro escogido para variar fue la presión, entonces se mostrará un gráfico como el siguiente:

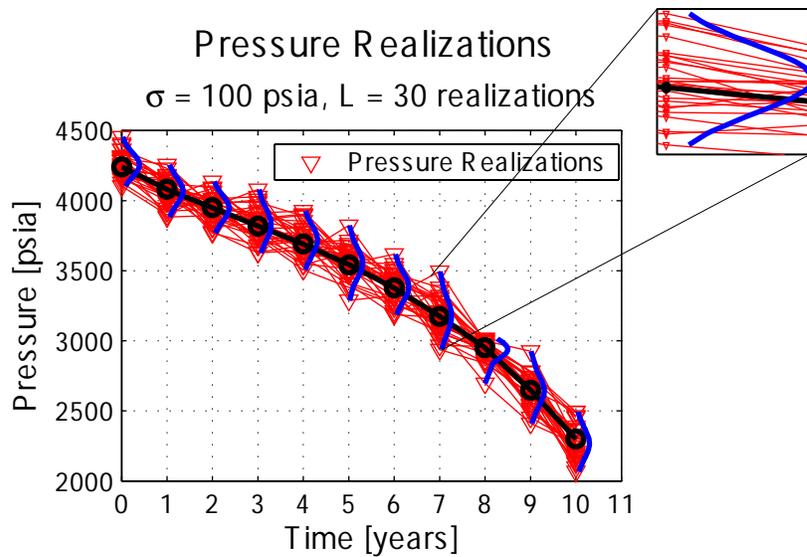


Figura 6.33: Ejemplo del gráfico de realizaciones de presión

El eje vertical es la presión en psia, y el eje horizontal es el tiempo en años. Las realizaciones aparecen en rojo, la línea negra representa la historia de presión sin error, y las líneas azules muestran la distribución de los valores de presión en cada punto de la historia, es decir, muestra la función de densidad de probabilidad para esa presión.

Cuando lo que se varía es algún parámetro del PVT, entonces los gráficos obtenidos son similares a la figura 6.34. Nuevamente las líneas en rojo representan las realizaciones y la línea negra representa la propiedad PVT sin error en los datos de entrada. En el título aparece que parámetro PVT se varió, cual fue su desviación estándar y el número de realizaciones calculadas.

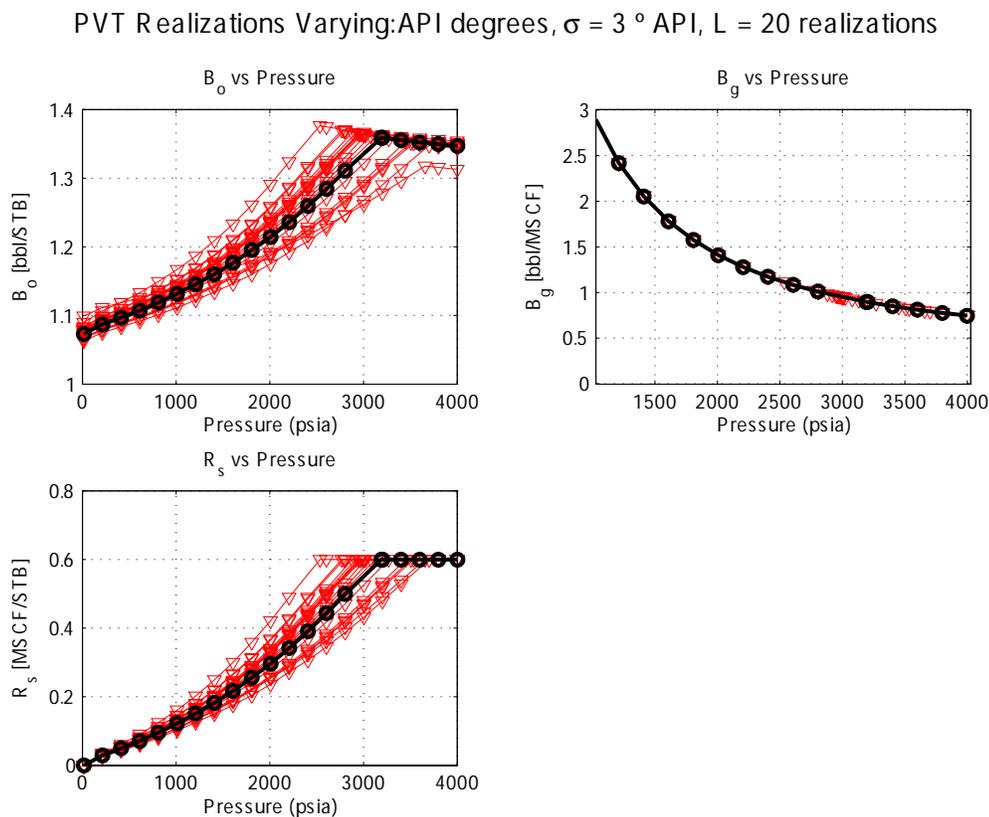


Figura 6.34: Ejemplo del gráfico de realizaciones de las propiedades PVT

Botón Error Plots

La ventana para acceder a los gráficos de error en las estimaciones de balance de materiales se muestra en la figura 6.35. Se selecciona el gráfico que se desea ver en la lista “Error plot”, y se hace clic en el botón *Graph*. Como se mencionó antes es necesario hacer varias corridas con diferentes valores de desviación estándar o número de realizaciones para obtener el gráfico requerido.

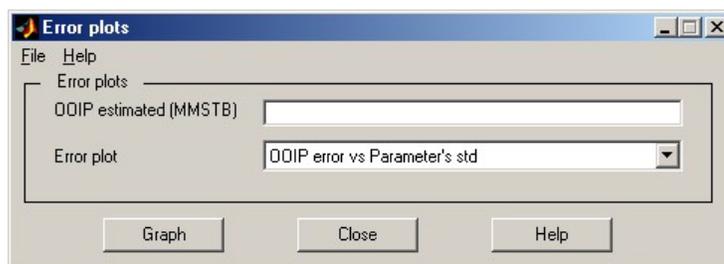


Figura 6.35: Ventana **Error plots**

Menú Error Plots

El menú posee una serie de submenús para observar el error estimado de las realizaciones, si se posee un valor estimado del POES. Por tanto el primer paso es seleccionar **OOIP estimated**, esto abrirá una caja de dialogo en la cual se puede introducir el valor del POES estimado en MMSTB.

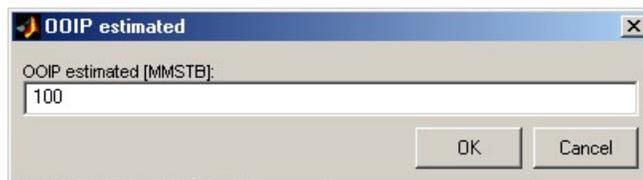


Figura 6.36: Caja de dialogo para el POES estimado

Después de esta operación los submenús de los gráficos se encontrarán disponibles para su selección, específicamente estos son: **Error vs Standard Deviation**, **Error vs N° Realizations**, y **Ternary Plot**. Para los submenús **Error vs Standard Deviation**, **Error vs N° Realizations** aparece una ventana de selección para elegir el gráfico del método de balance de materiales que se desea visualizar, idéntica a la figura 6.28 sólo que el título de la ventana cambia a *Error Plots*.

Para utilizar todos los submenús adecuadamente se deben seguir los pasos para cada submenú explicados a continuación.

Submenú Error vs Standard Deviation: Este submenú permite ver como ha variado el error, en los POES calculados, a medida que se va cambiando la desviación estándar del parámetro seleccionado. El programa realiza las siguientes operaciones: calcula un POES promedio por cada corrida realizada, con una desviación estándar especificada; luego se calcula el error relativo de este POES promedio con el POES estimado, introducido anteriormente en el submenú **OOIP estimated**, entonces se grafica el error obtenido vs. la desviación estándar, claro nada más se tendrá un punto en el gráfico. Si se realiza la misma operación para diferentes valores de desviación estándar se obtiene un gráfico parecido a la siguiente figura.

Error vs. Standard Deviation, Varying: Pressure

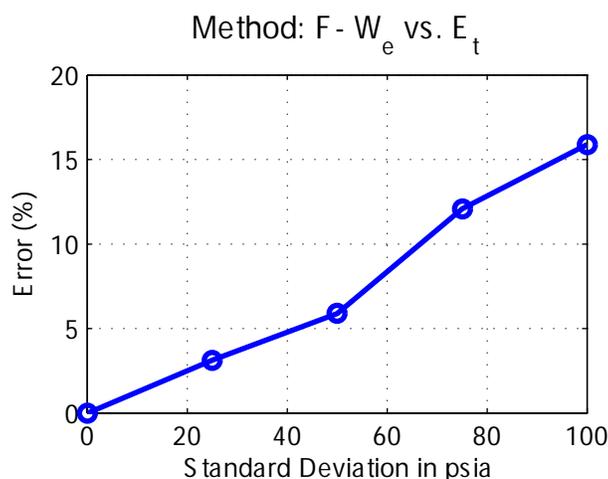


Figura 6.37: Gráfico del Error en el POES vs desviación estándar del parámetro escogido

El título del gráfico muestra el parámetro que se varió, el eje vertical de los gráficos es el error de los POES calculados en porcentaje, y el eje horizontal es la desviación estándar del parámetro seleccionado. En resumen, para construir estos gráficos se deben seguir los siguientes pasos:

- seleccionar cual parámetro es el que se desea variar, este parámetro no se deberá cambiar en las siguientes realizaciones.

- seleccionar el número de realizaciones que se desean calcular, igualmente este valor no se deberá cambiar en las siguientes realizaciones.
- introducir el valor de la desviación estándar para el parámetro seleccionado.
- correr el módulo.
- Cambiar la desviación estándar y correr de nuevo el módulo. Repetir tantas veces esta operación como puntos se desean en el gráfico de error vs desviación estándar.

Submenú Error vs N Realizations: La operación es análoga que para el submenú de **Error vs Standard Deviation**, a diferencia de que se mantiene la desviación estándar constante para todas las corridas y se varía el número de realizaciones a calcular por corrida. Por tanto, los pasos a seguir son los mismos, excepto que en vez de cambiar la desviación estándar, se cambia el número de realizaciones, y se deja la desviación estándar del parámetro escogido fija para todas las corridas. Es muy importante recordar que no se debe cambiar el parámetro a variar porque causará error en el programa. La gráfica obtenida para este submenú se presenta a continuación.

Error vs. N° Realizations, Varying: Pressure, $\sigma=100$ psia

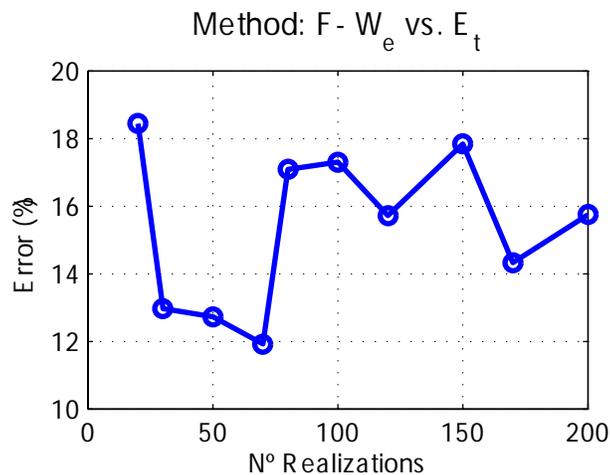


Figura 6.38: Gráfico del Error en el POES vs el número de realizaciones

Submenú Ternary Plot: El submenú **Ternary Plot** fue diseñado para evaluar como influye el mecanismo de empuje en el error de los POES calculados. Algunos comentarios con respecto a la construcción de este gráfico:

- Los valores de desviación estándar y el número de realizaciones debe permanecer constante para todas las corridas.
- Cada corrida representa un mecanismo de empuje diferente, por tanto, se tendrá que cargar los datos correspondientes de producción, acuífero, y de yacimiento.
- Se supone que los fluidos serán los mismos para todas las corridas por lo que el PVT debería permanecer constante.
- Como este gráfico representa una superficie, se necesitan por lo menos tres puntos para generar el gráfico, en otras palabras, se necesitan tres casos con mecanismos de empuje diferentes.

Después de seleccionar el submenú aparece la siguiente ventana de selección, para elegir qué método se va a visualizar, y si se graficará la superficie de error. Si no se selecciona la opción **Ternary Plot 3D**, únicamente se dibujará el diagrama de contornos para la superficie.

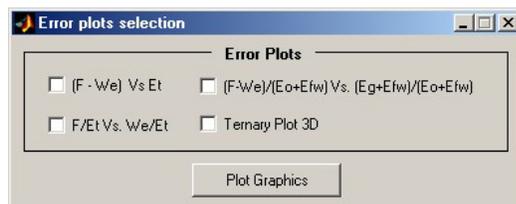


Figura 6.39: Ventana de selección para el gráfico ternario

Dependiendo de la ubicación del punto en el diagrama ternario, se tiene la contribución relativa de energía del yacimiento. En otras palabras, un punto en el diagrama ternario representa un contribución específica de mecanismos de empuje, los cuales caracterizan el yacimiento. Un ejemplo de diagrama ternario se muestra en la figura 6.40. Se debe tener cuidado al analizar este gráfico ya que utiliza el método de interpolación *Nearest Point*, es decir, no interpola o extrapola los puntos para la superficie restante, sino que utiliza el valor del punto para las áreas cercanas a ese punto para calcular la superficie. Por esta razón se recomienda utilizar este gráfico con la mayor cantidad de puntos posibles para obtener una buena interpretación y análisis. Para aquellos casos en que el yacimiento no posea capa de gas se debe establecer en la ventana de la figura 6.17 que existe capa

de gas, y se le asigna un valor de 0. Esto con el fin de que se pueda obtener una superficie de error para el método $(F - W_e)/(E_o + E_{fw})$ vs $(E_g + E_{fw})/(E_o + E_{fw})$. De la misma manera para aquellos casos en los que no exista acuífero se selecciona el modelo **Pot Acuífer**, por ejemplo, y se le asigna un W de 0 bbls, es decir, sin influjo de agua. También se realiza esto para obtener una superficie de error para el método F/E_t vs W_e/E_t . La barra de colores que aparece en el gráfico representa el valor del error en porcentaje.

Ternary Plot, Varying: Pressure, $\sigma = 10$ psia

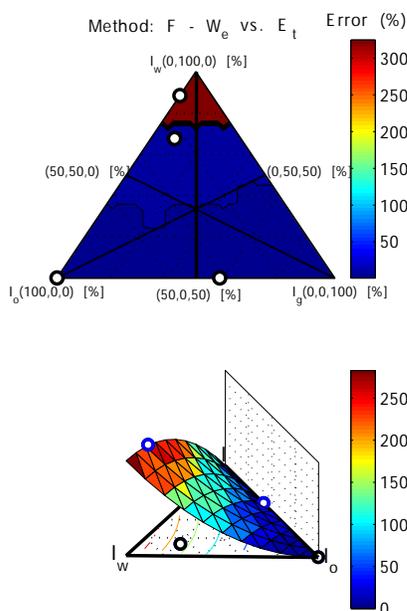


Figura 6.40: Ejemplo de gráfico ternario

El último submenú es el correspondiente a **Reset Error Plots**. Desactiva todos los submenús de los gráficos de error y ternario, y activa nuevamente el submenú de **OOIP estimated**. Esto para el caso de haber introducido un valor errado del POES, o cuando se desea hacer el análisis para otro valor del POES.

6.4.3. Archivo de salida para el cálculo de realizaciones

El programa genera un archivo de salida (*.mbc) con información de los cálculos realizados para las estimaciones de balance de materiales. Específicamente, proporciona los valores del vaciamiento, influjo de agua, términos de expansión de fluidos, índices de empuje; y los métodos utilizados junto con sus estimados del POES, intervalos de confianza, y el RMSE del ajuste para el POES estimado. El formato es como el mostrado a continuación:

```
%Material Balance Calculations, Run #: 1
%Material Balance Calculations Realization Set: 1
% Underground Withdrawals [MMbbl]
0.000000
1.989815
.
.
% Cumulative Aquifer Influx [STB]
0.000000
0.000000
.
.
% Expansion terms:
% E(:,1)=Eo , E(:,2)=Eg , E(:,3)=Efw , E(:,4)=Et , . . .
% E(:,6)=(F-We)/(Eo+Efw) , E(:,7)=We/Et , E(:,8)=F/Et . . .
0.00 0.00 0.00 0.00 0.00 . . .
0.02 0.03 0.00 0.02 1.44 . . .
. . . . .
. . . . .
% Drive Indices:
% I(:,1:4) for Methods: (F-We vs. Et)/(F/Et vs. We/Et);
% I(:,5:8) for Method: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw);
% I(:,1)=Eo I(:,2)=Eg I(:,3)=Efw I(:,4)=Et;
% I(:,5)=(Eg+Efw)/(Eo+Efw) I(:,6)=(F-We)/(Eo+Efw) . . .
0.00 0.94 0.00 0.06 0.00 . . .
0.00 0.95 0.00 0.05 0.00 . . .
. . . . .
. . . . .
```

Figura 6.41: Archivo *.MBC

```
% OOIPs estimated and its statistical results:
% Method (F-We vs. Et)
% OOIP [MMSTB]
    100.19
% Confidence interval for OOIP: (Inf limit, Sup limit)
(    99.92 ,    100.45 )
% RMSE for OOIP:
    0.07
% Methods ((F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)):
% OOIP [MMSTB]:
    107.00
% Confidence interval for OOIP: (Inf limit, Sup limit)
(    93.34 ,    120.67 )
% RMSE for OOIP:
    0.91
% Gas cap estimated:
    -0.04
```

Figura 6.42: Archivo *.MBC, continuación

6.5. Diseño experimental para los datos presión y PVT

6.5.1. Introduciendo los parámetros para el cálculo del diseño experimental

Para realizar diseños experimentales para los parámetros de PVT y presión, se debe generar el PVT mediante las correlaciones disponibles. Es de hacer notar que esta sección es completamente independiente de la correspondiente a las realizaciones. Se encuentran disponibles cinco tipos diferentes de diseños: Box-Behnken, Central Composite Circumscribed, Inscribed, Faced, y Full Factorial de tres niveles. Se puede acceder a esta sección mediante el último botón de la pantalla principal **DoE** (con forma de cubo) o a través del menú **DoE**. En la figura 6.43 se muestra la ventana correspondiente a la ventana **Experimental design (DoE)**. La diferencia entre acceder mediante el menú ó la ventana radica en que: en el menú **DoE** los diseños aparecen como submenús, mientras que en la ventana se seleccionan directamente.

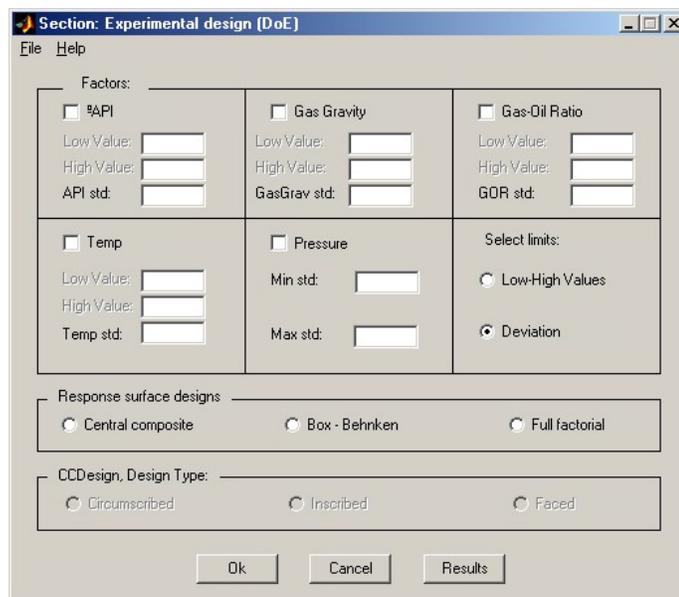


Figura 6.43: Ventana **DoE**

Esta ventana permite introducir los niveles para cada factor que se desea variar. Simplemente, se selecciona los parámetros que se desean variar y si se les asigna los niveles especificando, el valor bajo y alto, o se le asigna una desviación

para el parámetro (niveles = valor del parámetro +/- desviación). Es importante recordar que se calcularán tres niveles, excepto para los diseños *central composite circumscribed* y *inscribed*, en los que se calculan cinco niveles. El nivel medio viene dado por el promedio de los valores altos y bajos introducidos; o el valor del parámetro que se utilizó para generar el PVT, en el caso de asignar una desviación. Cuando se elige **Central Composite**, en el submenú o en la ventana de diseños, se activa la opción **CCDesign, Design Type** y se elige entre los diseños circunscrito (circumscribed), inscrito (inscribed), o de caras (faced). Hay que recordar que los diseños Box-Behnken fueron creados para diseños experimentales con más de tres factores, por tanto al intentar seleccionar el diseño Box-Behnken con dos factores aparece un mensaje de error, al igual que aparece cuando se intenta un diseño experimental de un factor para los otros diseños disponibles. Luego de seleccionar los parámetros a variar y seleccionar los límites a utilizar, aparece una ventana de progreso de los cálculos con el tiempo remanente estimado igual al de la figura 6.25.

6.5.2. Visualizando los resultados del diseño experimental

Los resultados de las corridas del diseño experimental se pueden ver o, haciendo clic en el botón *Results* de la ventana de selección del diseño, o a través del submenú **Plots** del menú **DoE**. Si se selecciona la primera opción aparece la siguiente ventana:

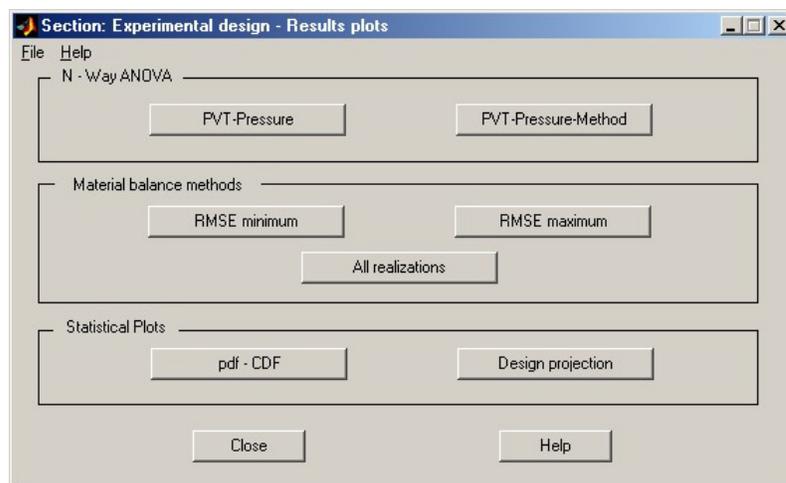


Figura 6.44: Ventana **Experimental design - Results plots**

El submenú **Plots** permite visualizar los resultados de los cálculos del diseño experimental y el diseño utilizado. Posee los submenús **All Realizations**, **Mi-**

nimum RMSE, Maximum RMSE, pdf's-CDF's y Design projection. Los primeros cuatro submenús funcionan idénticamente a aquellos del submenú **Result** de la sección de realizaciones.

El submenú **Design projection** ó el botón **Design projection** en la ventana **Experimental design - Results plots** abre una ventana donde se puede visualizar el diseño experimental escogido. Hay varias formas de representar el diseño experimental, estas son: diagrama 1-D, diagrama 2-D, diagrama 3-D, y diagrama 4-D. Estos diagramas se pueden seleccionar mediante la barra de herramientas que aparece en la parte superior de la ventana.

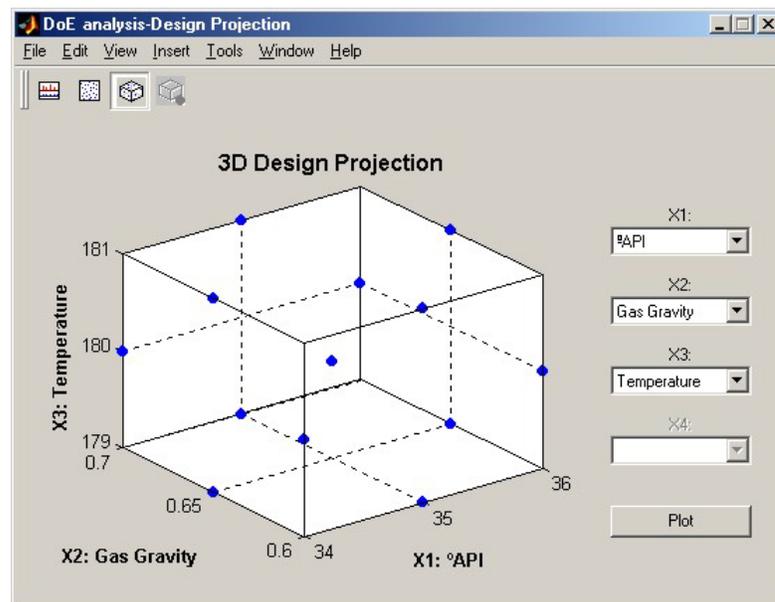


Figura 6.45: Ventana de visualización del diseño experimental

Diagrama 1-D: El primer botón de la barra de herramientas permite graficar un histograma del parámetro elegido de la lista X1.

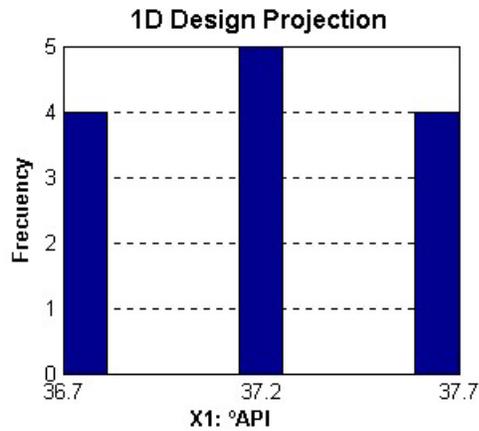


Figura 6.46: Ejemplo de diagrama 1-D

Diagrama 2-D: Corresponde al segundo botón de la barra de herramientas y grafica, en el plano, los parámetros seleccionados en las listas X1 y X2.

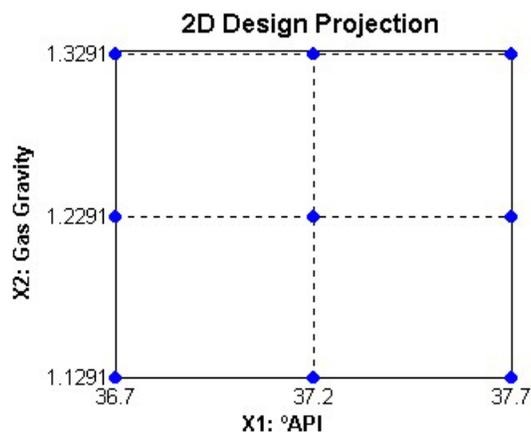


Figura 6.47: Ejemplo de diagrama 2-D

Diagrama 3-D: Permite visualizar en el espacio el diseño experimental elegido, los parámetros se escogen de las listas X1, X2, y X3.

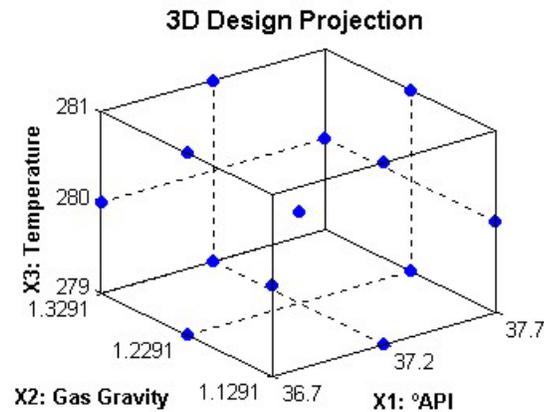


Figura 6.48: Ejemplo de diagrama 3-D

Diagrama 4-D: Se puede visualizar al seleccionar el cuarto botón de la barra de herramientas. Los parámetros se seleccionan de las listas X1, X2, X3, y X4. La barra de colores representa el valor que toma el parámetro X4 en los puntos graficados.

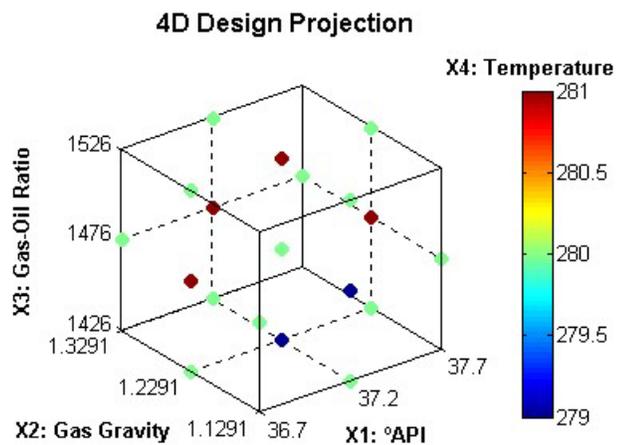
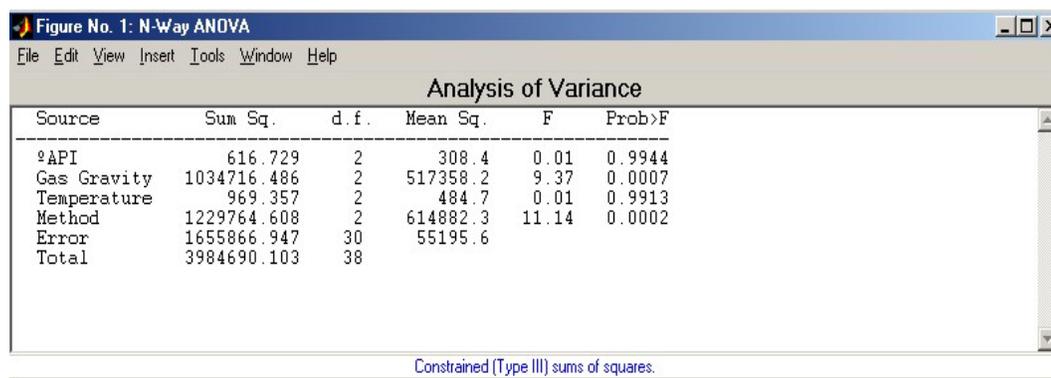


Figura 6.49: Ejemplo de diagrama 4-D

Si se elige el submenú **N-Way ANOVA** se realiza un análisis de varianza de múltiples factores para el diseño experimental calculado. Posee dos submenús **General** y **PVT-Pressure Data**. Estos submenús producen el mismo resultado que los botones correspondientes en la ventana **Experimental design - Results plots**.

Submenú **General**: Al seleccionar esta opción aparece la ventana de la figura 6.50. Ésta muestra una tabla, resultado del análisis de varianza aplicado a los datos para múltiples factores. Específicamente, el análisis que se realiza en esta prueba es evaluar de entre los parámetros que se variaron en el diseño experimental y el método de balance de materiales utilizado, cuál tiene más influencia en los resultados obtenidos. Este submenú sólo estará disponible si los tres métodos se encuentran disponibles para realizar los cálculos. También se aplican los comentarios hechos acerca de estos métodos en la sección correspondiente al diagrama ternario. Éstos son, seleccionar que existe capa de gas para utilizar el método $(F - W_e)/(E_o + E_{fw})$ vs $(E_g + E_{fw})/(E_o + E_{fw})$ y seleccionar algún modelo de acuífero con los valores de los parámetros apropiados para el método F/E_t vs W_e/E_t . Para la figura mostrada, se observa que los parámetros que no resultaron significativos fueron la gravedad API y la temperatura del yacimiento, dado que sus valores p son mayores a 0.05. Este submenú permite evaluar en forma global cuales son los factores que más influencia tienen sobre los resultados.

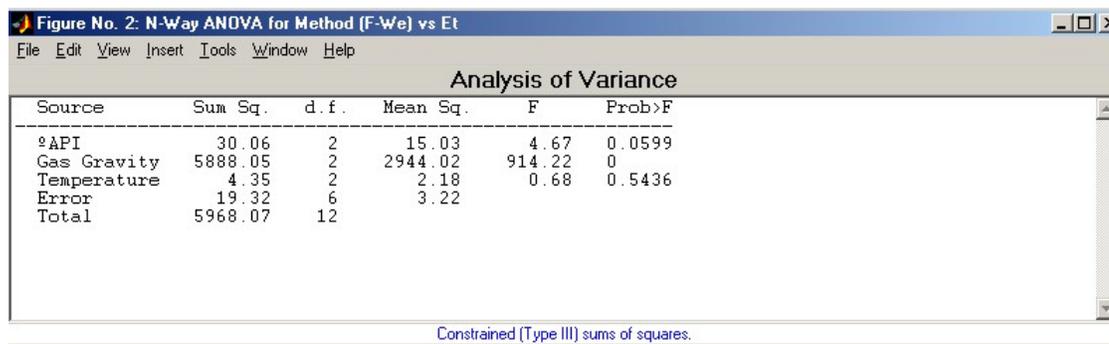


Source	Sum Sq.	d. f.	Mean Sq.	F	Prob>F
%API	616.729	2	308.4	0.01	0.9944
Gas Gravity	1034716.486	2	517358.2	9.37	0.0007
Temperature	969.357	2	484.7	0.01	0.9913
Method	1229764.608	2	614882.3	11.14	0.0002
Error	1655866.947	30	55195.6		
Total	3984690.103	38			

Constrained (Type III) sums of squares.

Figura 6.50: Ejemplo de una tabla N-Way ANOVA

Submenú **PVT-Pressure Data**: Posee los submenús correspondientes a cada uno de los métodos de balance de materiales, es decir $F - W_e$ vs E_t , $(F - W_e)/(E_o + E_{fw})$ vs $(E_g + E_{fw})/(E_o + E_{fw})$ y F/E_t vs W_e/E_t . Después de seleccionar cualquiera de los métodos, aparecerá la siguiente ventana.



Source	Sum Sq.	d. f.	Mean Sq.	F	Prob>F
%API	30.06	2	15.03	4.67	0.0599
Gas Gravity	5888.05	2	2944.02	914.22	0
Temperature	4.35	2	2.18	0.68	0.5436
Error	19.32	6	3.22		
Total	5968.07	12			

Figura 6.51: Ejemplo de una tabla N-Way ANOVA para el método $(F - W_e)$ vs E_t

Este conjunto de submenús permite realizar un análisis de varianza para múltiples factores en los resultados del diseño experimental. Es decir, permite concluir cuál o cuales de los parámetros que se variaron en el diseño experimental son más significativos. Para la figura mostrada el parámetro más significativo fue la gravedad específica del gas, ya que su valor p resultó ser cero. El método de balance de materiales utilizado se puede observar en el título de la figura.

6.5.3. Archivo de salida para los cálculos del diseño experimental

El programa genera un archivo de salida (*.mbc) con información de los cálculos realizados para las estimaciones de balance de materiales utilizando diseño experimental. Al igual que el archivo MBC para las realizaciones, proporciona los valores del vaciamiento, influjo de agua, términos de expansión de fluidos, índices de empuje; y los métodos utilizados junto con sus estimados del POES, intervalos de confianza, y el RMSE del ajuste para el POES estimado. Además proporciona la matriz de diseño (donde se especifican los niveles) utilizada para los cálculos. En la siguiente figura se muestra la primera parte del archivo, el resto es igual al mostrado en la sección 6.4.3.

```
%Material Balance Calculations, Run #: 1
% Experimental Design Calculations
% Design: Box-Behnken
%
% Design matrix:
%(API) (GasGrav) (Rs)      (Tr)      (dp)
36.70    1.33  1476.00  280.00    0.00
37.70    1.13  1476.00  280.00    0.00
37.70    1.33  1476.00  280.00    0.00
36.70    1.23  1476.00  279.00    0.00
36.70    1.23  1476.00  281.00    0.00
37.70    1.23  1476.00  279.00    0.00
37.70    1.23  1476.00  281.00    0.00
37.20    1.13  1476.00  279.00    0.00
37.20    1.13  1476.00  281.00    0.00
37.20    1.33  1476.00  279.00    0.00
37.20    1.33  1476.00  281.00    0.00
37.20    1.23  1476.00  280.00    0.00
```

Figura 6.52: Archivo *.MBC para diseño experimental

Capítulo 7

Metodología

La metodología de trabajo sigue un conjunto de pasos para el desarrollo de un programa que permite evaluar distintos métodos de balance de materiales bajo la acción de la incertidumbre en la presión y los datos de entrada a las correlaciones PVT (es decir, relación gas petróleo en solución, gravedad API, gravedad específica del gas, y temperatura del yacimiento). La serie de pasos se detallan a continuación:

7.1. Revisión bibliográfica

Esta revisión abarcó varios campos. Primero se indagó acerca de investigaciones precedentes referente al problema (en otras palabras, investigaciones sobre la influencia de errores en los cálculos de balance de materiales). Los resúmenes de las principales investigaciones se pueden encontrar en el capítulo 4, denominado *antecedentes del problema*. Este fue el punto de partida para dirigir la investigación ya que permitió discernir las áreas en las que se necesitaba ahondar en conocimientos. Segundo, se realizó la revisión de libros de textos y reportes acerca de los métodos de balance de materiales a utilizar. Esta revisión comprende el capítulo 1, *la ecuación de balance de materiales*. Tercero, debido a que el análisis a realizar necesita tener una base estadística se hizo una compilación de los conceptos necesarios para su desarrollo y entendimiento. Los capítulos correspondientes son el capítulo 2, *fundamentos de estadística*, y el capítulo 3, *herramientas estadísticas a utilizar*. Por último, la investigación se dirige a delimitar los márgenes de error en la medición de los parámetros a analizar. Esta información se encuentra en el capítulo 5, *error en la medición de la presión y los parámetros PVT*.

7.2. Determinación del análisis a utilizar

De las investigaciones previas se estableció que se implementará un análisis estadístico para evaluar los rangos de incertidumbre en los cálculos de balance de materiales. El análisis estadístico involucra herramientas como análisis de varianza (ANOVA), y diseño experimental, por mencionar los más importantes (véase capítulo 3). Los rangos de incertidumbre a evaluar se delimitarán en base al capítulo 5 referente a *error en la medición de la presión y los parámetros PVT*. Los métodos de balance de materiales a analizar se designaron en base a las investigaciones previas.

7.3. Modelos de yacimientos

Con el fin de estudiar como influye la incertidumbre de los datos de presión y PVT en los cálculos de balance de materiales, se decidió extender el análisis a un amplio rango de mecanismos de empuje. En total se presentan 15 casos de yacimientos con diferentes proporciones de los mecanismos de empuje: expansión del petróleo más gas disuelto, expansión del gas libre, e influjo de agua. Todos los yacimientos poseen las mismas propiedades de la roca y los fluidos (es decir, el mismo PVT, compresibilidad de la roca, permeabilidad, porosidad, etc). También es importante hacer notar que todos los yacimientos poseen el mismo POES (las dimensiones de la capa de petróleo permanecen inalterables para todos los casos). Las características que cambian de un yacimiento a otro corresponde al tamaño de la capa de gas (en caso de que exista) y la fuerza del acuífero (igualmente, en caso de que exista). Las propiedades de la roca y los fluidos se presentan en el apéndice A. Las propiedades PVT simulan aquellas del campo Santa Rosa [36], éstas se pueden generar a partir de la correlación de Vazquez Beggs mediante los siguientes parámetros de entrada:

Gravedad API 37.2

Gravedad específica del gas @ Pb 1.2291

Relación gas-petróleo en solución @ Pb 1476 SCF/STB

Temperatura del yacimiento 280°F

Presión del separador 114.7 psia

Temperatura del separador 60°F

En la figura 8.1 se puede observar la configuración del tamaño de la capa de gas y el tamaño relativo del acuífero para cada uno de los casos evaluados.

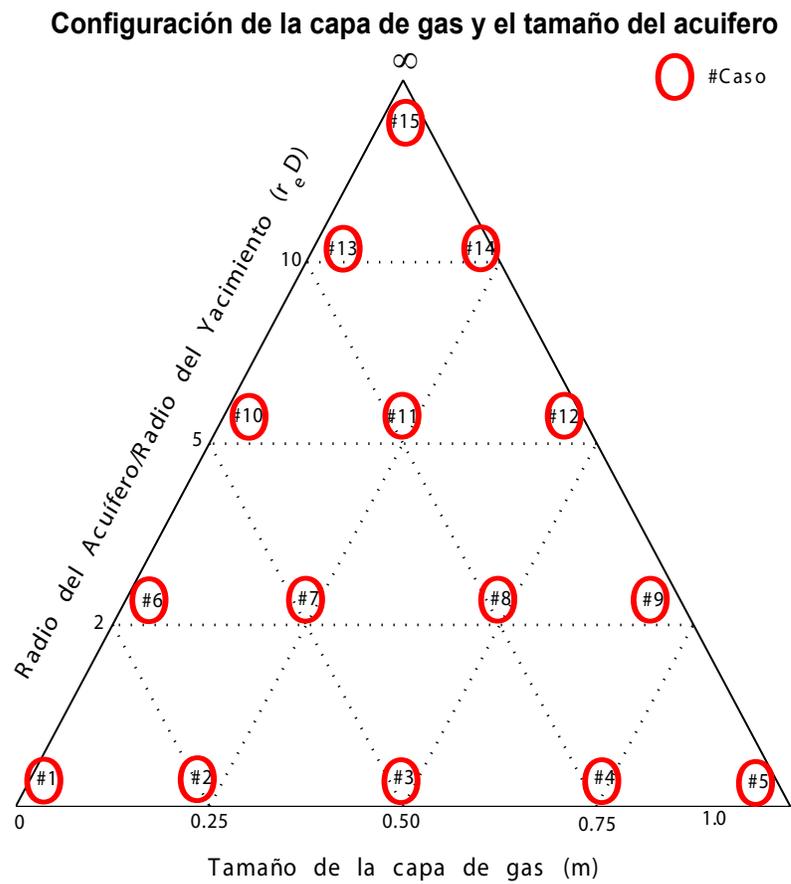


Figura 7.1: Diagrama mostrando la configuración del tamaño de la capa de gas y el tamaño relativo del acuífero

7.4. Rangos de incertidumbre a evaluar

Los rangos de incertidumbre a evaluar de acuerdo a la revisión bibliográfica del capítulo 5, *error en la medición de la presión y los parámetros PVT* son:

Presión , [2-200 psi], ya sea para el cálculo de las realizaciones o el diseño experimental.

Gravedad API , [0.1-0.5 grados API].

Gravedad específica del gas , el rango a utilizar es [0.01 % - 1 %] de la lectura realizada. Esto corresponde a [0.0001-0.0123 unidades de gravedad específica].

Relación gas-petróleo en solución , [0.1 %, - 2 %] de la lectura realizada. Lo cual es equivalente a [1.476-29.52 SCF/STB].

Temperatura del yacimiento , [0.04-0.9°F]. Esto es equivalente en °C aproximadamente a [0.02-0.5°C].

Como ya se mencionó estos rangos se utilizarán para el cálculo de las realizaciones o en la aplicación del diseño experimental.

Ahora bien, la distribución probabilística que se utilizará para calcular las realizaciones es una distribución normal. Se utilizará esta distribución debido a que la mayoría de las variables observadas (mediciones) están normalmente distribuidas (es la razón por la que la distribución normal representa una característica general de la realidad empírica) [37]. En una distribución normal, el 95 % de las observaciones se encuentran a $\pm 2\sigma$ del promedio [38]. Por tanto, las desviaciones reportadas anteriormente para la presión y los parámetros PVT equivaldrán a 2σ . Falta definir el tamaño de la muestra, es decir, el número de valores aleatorios a tomar (en el rango [valor real de la medición - 2σ , valor real de la medición + 2σ]) para el cálculo de la distribución. De la experiencia se dice que con un tamaño de muestra de 50 o más, los sesgos son poco probables, y con un tamaño de muestra de 100 no debería existir preocupación acerca de las suposiciones de normalidad [39]. Entonces, de acuerdo a esto se tomarán [50, 100, 150, 200, y 250] valores aleatorios en el rango establecido para el cálculo de las realizaciones. Es de hacer notar que el número de realizaciones será siempre igual al número de valores aleatorios para el cálculo de las distribuciones.

7.5. Proceso de cálculo

Se seleccionaron dos enfoques para el análisis. El primero es probabilístico y se basa en que los errores de los parámetros poseen una distribución normal con una desviación estándar conocida. En base a esta distribución y al parámetro seleccionado se hacen los cálculos de realizaciones. Si el parámetro es la presión, se modifica la historia de presión vs. tiempo reportada. Para los parámetros PVT se toma un valor aleatorio dentro de la distribución normal especificada y se introducen en los cálculos del PVT mediante las correlaciones físicas de fluidos. Posteriormente se toman éstos resultados y se realizan cálculos de balance de materiales. El segundo enfoque es determinístico, es decir, en vez de tomar números aleatorios dentro de una distribución, se utiliza diseño experimental para definir un conjunto de posibles valores para los parámetros (niveles), que generalmente varía entre 2 y 3 valores diferente por parámetro. Del mismo modo que en el enfoque probabilístico, se toman éstos resultados y se realizan cálculos de balance de materiales. La ventaja del diseño experimental sobre el enfoque probabilístico es que maximiza la cantidad de información con un menor número de experimentos realizados como se mencionó en la sección 3.3. Además es ideal para cuando existen una gran cantidad de parámetros con incertidumbre.

Sin importar que enfoque se haya elegido ahora se tendrán un conjunto de estimaciones de balance de materiales. El siguiente paso es medir la tendencia central y la dispersión los datos, además de generar algunos gráficos estadísticos que se mencionarán más adelante. Por último se aplicará la técnica ANOVA para establecer cuál de los parámetros y/o métodos de balance de materiales tuvo más influencia en los resultados obtenidos. Por otra parte, si el enfoque utilizado es el probabilístico, se pueden generar gráficos de la incertidumbre del parámetro vs. el error en los POES calculados. El esquema completo del proceso se muestra de forma esquemática en la figura 7.2.

Esquema del Proceso

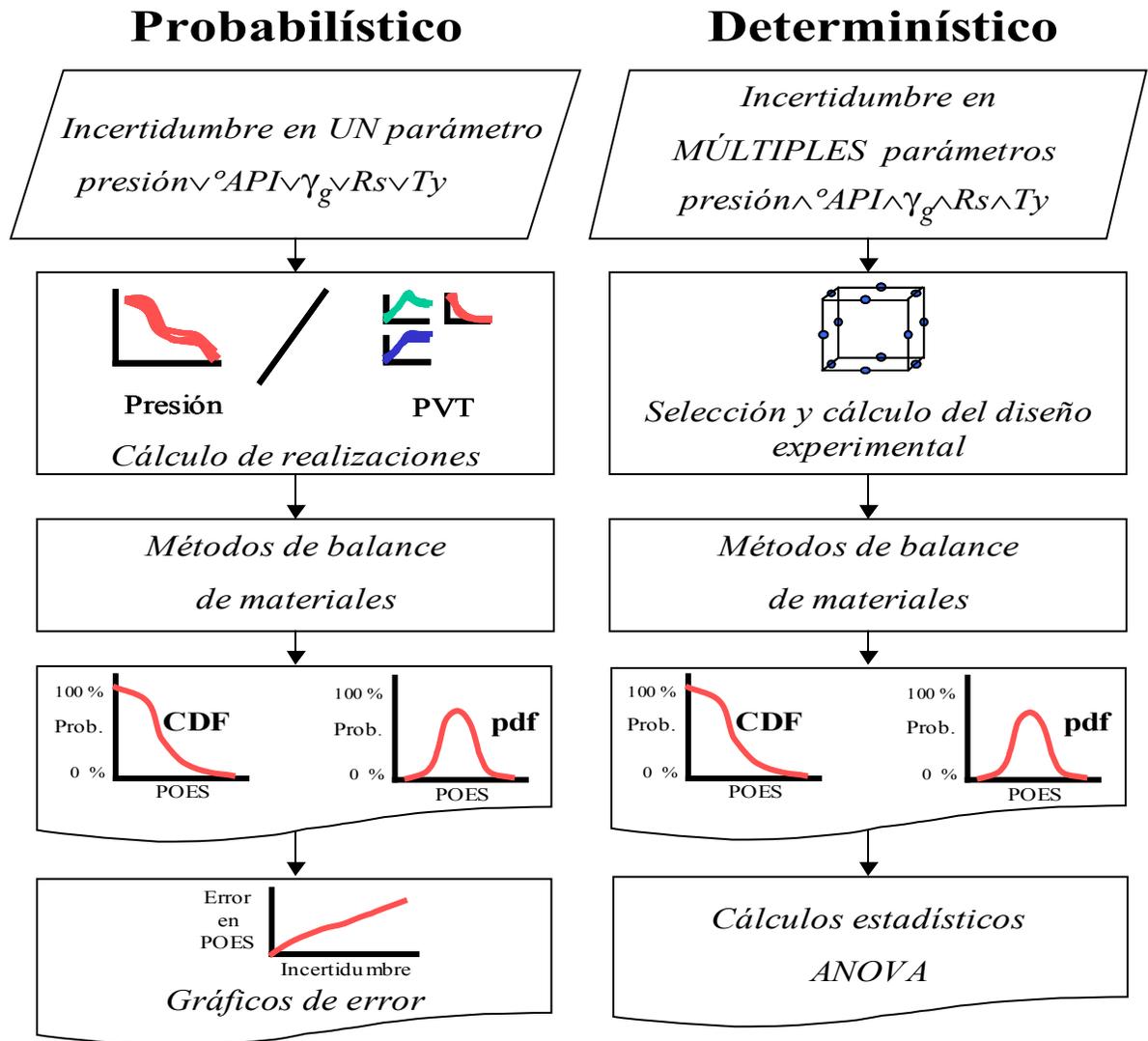


Figura 7.2: Diagrama mostrando el esquema del proceso de cálculo

7.5.1. Realizaciones

La incertidumbre de los datos de presión y PVT en los cálculos de balance de materiales se evaluarán mediante dos métodos. El primero corresponde a calcular realizaciones para la historia de declinación de presión ó para los parámetros PVT a introducir en las correlaciones de las propiedades físicas de fluidos.

Realizaciones de presión

Las realizaciones se calcularán de la siguiente manera: para cada punto de la historia de declinación de presión original se generará una distribución normal (la razón se explicó en la sección anterior) con la desviación estándar correspondiente al caso a evaluar. De esta manera, la historia de declinación de presión resultante se obtiene uniendo valores aleatorios de cada una de las distribuciones normales calculadas. La figura 8.2 ilustra el método a seguir.

Realizaciones PVT

La manera como se calculan las realizaciones para los parámetros PVT se explica a continuación: se escoge entre la gravedad API, la gravedad específica del gas, la relación gas-petróleo en solución, ó la temperatura del yacimiento, el parámetro que se desea variar. Ahora se genera una distribución normal para el parámetro seleccionado (especificando valores para la desviación estándar y el promedio, que corresponde al caso base sin error). Luego se seleccionan puntos aleatorios de esta distribución y junto con el valor del caso base para cada uno de los parámetros PVT restantes se calculan las propiedades PVT mediante las correlaciones de propiedades físicas de fluidos disponibles. El proceso completo se ilustra en la figura 8.3.

Gráficos y herramientas estadísticas para el análisis de las realizaciones

Para analizar los resultados obtenidos se utilizarán varios gráficos y herramientas estadísticas:

- Se genera una tabla ANOVA indicando si algún método arroja resultados significativamente diferente de los demás. Le acompaña un diagrama de caja que permite la comparación gráfica de los métodos de balance de materiales.
- Se visualizan todas las realizaciones en el gráfico de balance de materiales correspondiente así como aquella realización con el mayor o menor RMSE (la raíz del error cuadrado promedio).
- Se genera un gráfico con la función de distribución de probabilidad (pdf) y la función de distribución de probabilidad acumulativa (CDF) de los POES

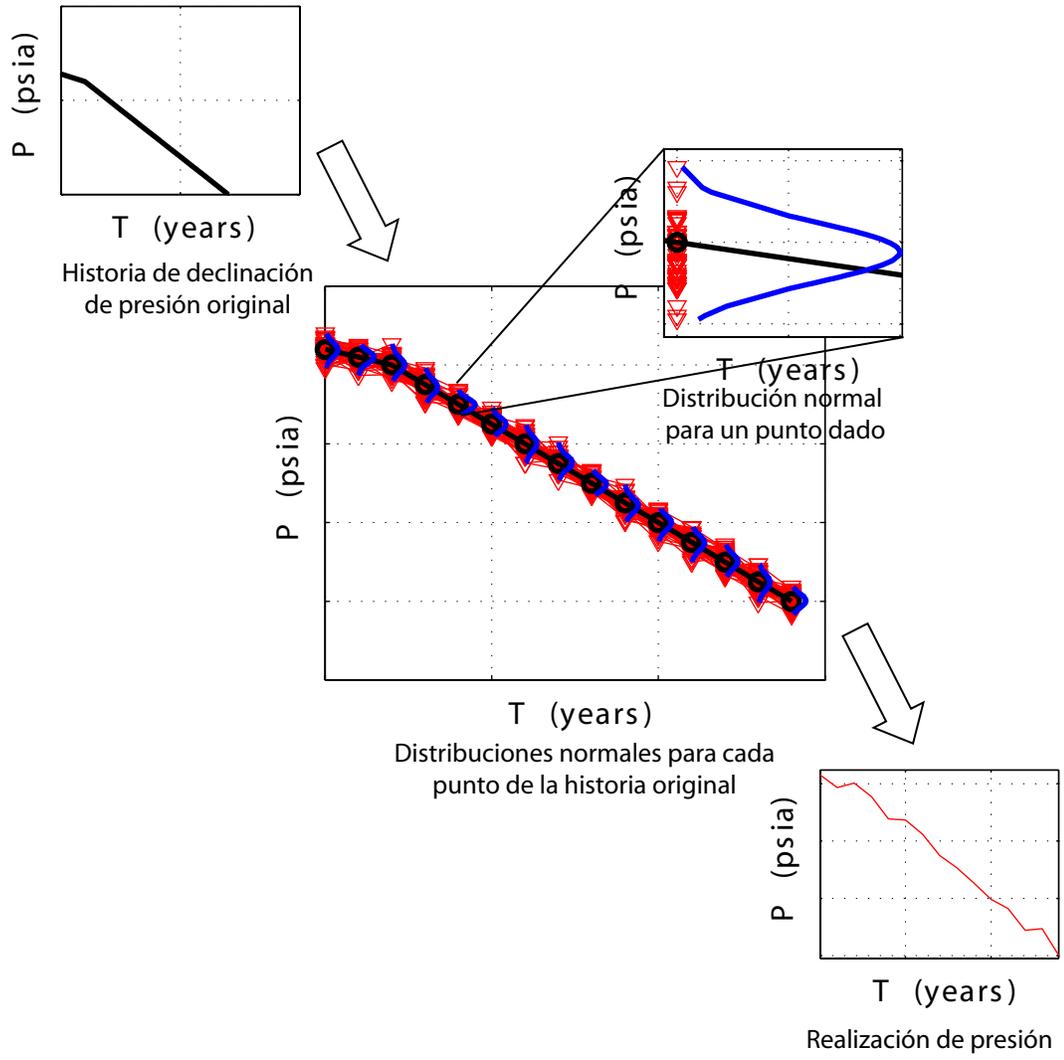
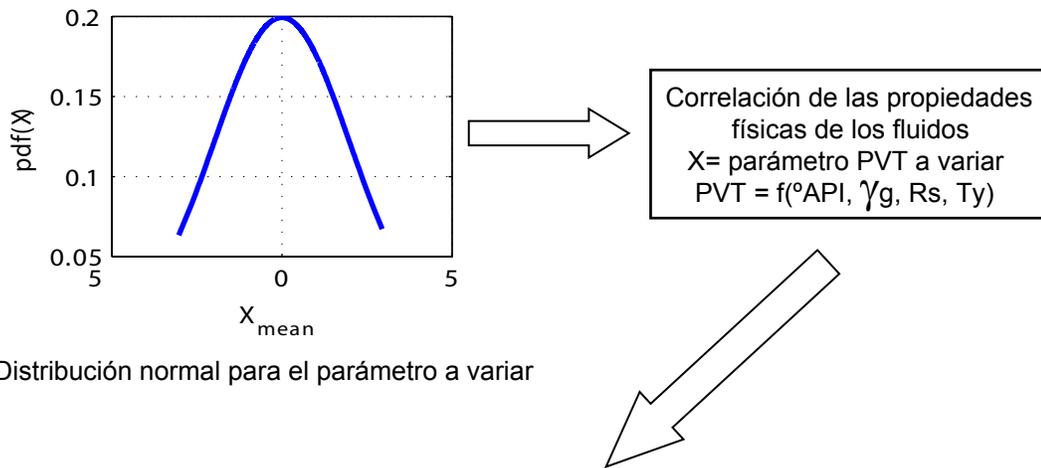


Figura 7.3: Cálculo de las realizaciones de presión



Realizaciones de las propiedades PVT

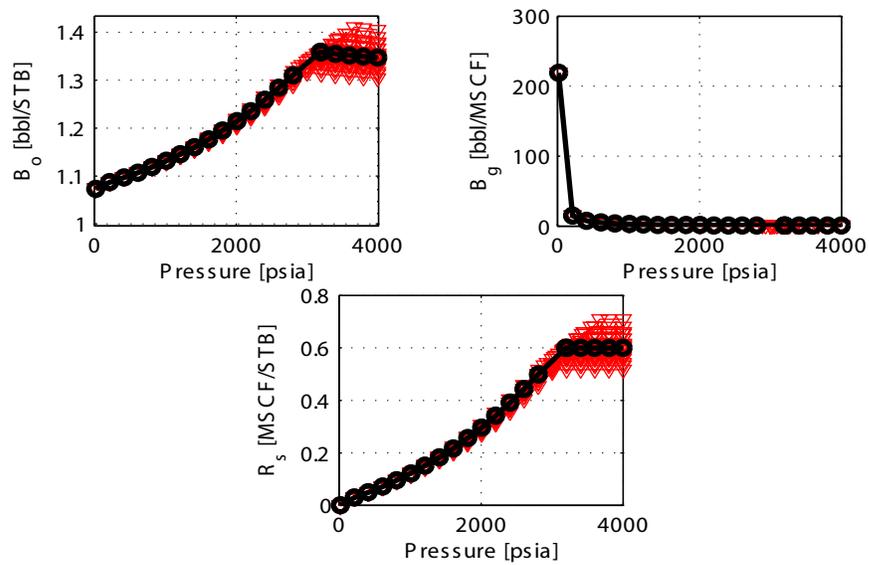


Figura 7.4: Cálculo de las realizaciones PVT

calculados a partir de las realizaciones, además de dar el valor mínimo calculado, el valor máximo, el promedio, la mediana, y la desviación estándar del POES.

- Y por último, se visualizan las realizaciones calculadas, ya sean para la historia de declinación de presión ó de las propiedades PVT.

A continuación se presentan otros métodos de análisis para las realizaciones basado en el error relativo de los POES calculados. Para esto se necesita conocer el valor del POES real, de esta manera se calcula el error relativo a partir de las realizaciones. Este método genera tres tipos de gráficos.

El primer gráfico que se genera es el error relativo vs. la desviación estándar del parámetro variado con el fin de analizar como influye la incertidumbre del parámetro sobre los POES calculados. Para realizar esto se calcula una corrida (es decir, se calculan el número de realizaciones especificadas) con una desviación estándar dada, y se calcula un RMSE de los POES calculados con respecto al POES real. Ahora se calcula otra corrida cambiando el valor de la desviación estándar a uno mayor y se continúa así hasta calcular todas las corridas deseadas. De esta manera se puede generar un gráfico del error relativo vs. la desviación estándar.

El segundo gráfico es el correspondiente al error relativo vs. el número de corridas y su fin es determinar cuantas corridas se necesitan para obtener unos resultados adecuados. El procedimiento es similar al anterior, sólo que en vez de cambiar el valor de desviación estándar para cada corrida se varía el número de corridas (mientras la desviación estándar permanece constante para todas las corridas).

El último gráfico es para determinar como influye la misma cantidad de incertidumbre de un parámetro bajo diferentes mecanismos de empuje. El gráfico se genera de la siguiente manera: Se realiza un diagrama ternario en base a los índices de empuje por agua [I_w], gas libre [I_g] y expansión del petróleo más su gas en solución [I_o] (el índice correspondiente a la compresibilidad de la roca y el agua connata se considera despreciable). Lo que se desea obtener es una superficie que representa el error relativo de los POES calculados a partir de las realizaciones. El diagrama ternario en el plano representa las contribuciones relativas de energía del yacimiento mientras el eje vertical representa el error relativo de los POES calculados. Con este gráfico se quiere identificar que mecanismos de empuje son menos afectados por la incertidumbre (aquellas zonas del gráfico donde la superficie de error alcance un valor mínimo). Ya que se trata de una superficie, se necesitan al menos tres casos con mecanismos de empuje diferentes para generar este gráfico. La figura 8.4 muestra un esquema del gráfico ternario.

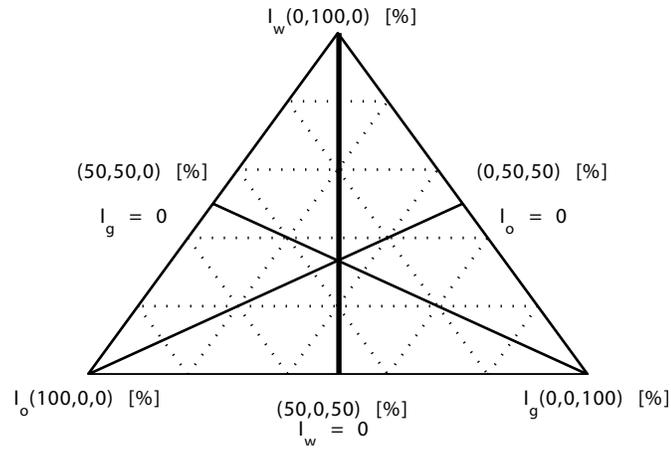


Figura 7.5: Gráfico ternario para la superficie de error

7.5.2. Diseño experimental

El segundo método para evaluar los rangos de incertidumbre es mediante el diseño experimental. Ya que el diseño de experimentos es una metodología para analizar como influye varios parámetros sobre una respuesta, al menos se pueden evaluar dos parámetros (para el diseño de Box Behnken se necesitan al menos tres parámetros). El proceso consiste de:

1. seleccionar los parámetros a evaluar
2. establecer los niveles para cada parámetro
3. seleccionar el tipo de diseño experimental a utilizar
4. generar la matriz de diseño basado en el tipo de diseño seleccionado
5. realizar las corridas de acuerdo a la matriz de diseño obtenida

Hay que mencionar que para variar la historia de presión se implementará el siguiente método. La desviación establecida en la matriz de diseño se utilizará como la desviación estándar para generar una realización de la historia de presión. Entonces se utiliza esta realización para los cálculos de balance de materiales. Por tanto la desviación máxima que habrá en la historia de presión será 2 veces la desviación establecida en la matriz de diseño.

Gráficos y herramientas estadísticas para el análisis del diseño experimental

Para analizar los resultados del diseño experimental se dispone de las siguientes herramientas:

- Se genera una tabla N-ANOVA indicando cual ó cuales de los parámetros variados influyó más significativamente sobre los resultados.
- Se visualizan todas las realizaciones así como aquella realización con el mayor o menor RMSE (la raíz del error cuadrado promedio).
- Se genera un gráfico con la función de distribución de probabilidad (pdf) y la función de distribución de probabilidad acumulativa (CDF) de los POES calculados a partir de las realizaciones, además de dar el valor mínimo calculado, el valor máximo, el promedio, la mediana, y la desviación estándar del POES.
- Y por último, se visualiza la matriz de diseño de una forma gráfica, dependiendo del número de parámetros variados.

7.6. Procedimiento para el análisis

Realizaciones

Para los casos 1, 5, 11, y 15 (ver apéndice A) se calcularán 100 realizaciones, con las desviaciones estándar indicadas a continuación:

Presión , las desviaciones estándar son [1 y 100 psi]

Gravedad API , las desviaciones estándar son [0.05 y 0.25 grados API].

Gravedad específica del gas , las desviaciones estándar a utilizar son [0.00005 y 0.00615 unidades de gravedad específica].

Relación gas - petróleo en solución , las desviaciones a aplicar son [0.738 y 14.76 SCF/STB].

Temperatura del yacimiento , se evaluarán las siguientes desviaciones estándar en la temperatura [0.02 y 0.45° F].

Para cada resultado se presentará una tabla resumen, y en aquellos casos posibles, el análisis ANOVA (para determinar si algún método de balance de materiales influye en los resultados).

Gráficos de error vs. desviación estándar y error vs. número de corridas

Para el gráfico de error vs. desviación estándar se calcularán 100 realizaciones por parámetro, y se utilizarán las siguientes desviaciones estándar:

Presión , las desviaciones estándar son [1, 25, 50, 75, y 100 psi]

Gravedad API , las desviaciones estándar son [0.05, 0.0625, 0.125, 0.1875, y 0.25 grados API].

Gravedad específica del gas , las desviaciones estándar a utilizar son [0.00005, 0.0015, 0.0030, 0.0045, y 0.00615 unidades de gravedad específica].

Relación gas - petróleo en solución , las desviaciones a aplicar son [0.738, 3.69, 7.38, 11.07, y 14.76 SCF/STB].

Temperatura del yacimiento , se evaluarán las siguientes desviaciones estándar en la temperatura [0.02, 0.1125, 0.2250, 0.3375, y 0.45° F].

Para el gráfico de error vs. el número de corridas se calcularán [50, 100, 150, y 200] realizaciones utilizando los siguientes valores de desviación estándar:

Presión [50 psi]

Gravedad API [0.25 grados API].

Gravedad específica del gas [0.0062 unidades de gravedad específica].

Relación gas - petróleo en solución [14.76 SCF/STB].

Temperatura del yacimiento [0.45° F].

Gráfico ternario de error

Para el gráfico ternario del error se utilizarán los 15 casos de yacimientos. Hay que recordar que para este gráfico el número de realizaciones y la desviación estándar del parámetro a variar deben ser iguales para todos los casos a evaluar. El número de realizaciones será de 100. Se evaluarán dos desviaciones estándar por parámetro para este gráfico, las cuales se muestran a continuación:

Presión [1 y 100 psi]

Gravedad API [0.05 y 0.25 grados API].

Gravedad específica del gas [0.00006 y 0.0062 unidades de gravedad específica].

Relación gas - petróleo en solución [0.738 y 14.76 SCF/STB].

Temperatura del yacimiento [0.02 y 0.45°F].

Diseño experimental

Para el análisis con diseño experimental se utilizarán los diseños full factorial y Box-Behnken. Para cada uno de estos tipos de diseños se van a variar todos los parámetros juntos. Se utilizarán dos diseños, uno con poca desviación con respecto al caso base (diseño N°1), y otro con desviaciones más grandes (diseño N°2).

El diseño N°1 posee las siguientes desviaciones:

Presión [10 psi]

Gravedad API [0.2 grados API].

Gravedad específica del gas [0.00123 unidades de gravedad específica].

Relación gas - petróleo en solución [7.38 SCF/STB].

Temperatura del yacimiento [0.09 °F].

El diseño N°2 con las desviaciones más altas se presenta a continuación:

Presión [100 psi]

Gravedad API [0.5 grados API].

Gravedad específica del gas [0.0123 unidades de gravedad específica].

Relación gas - petróleo en solución [29.52 SCF/STB].

Temperatura del yacimiento [0.9°F].

También se utilizarán estos valores para generar diseños donde sólo se varíen los parámetros PVT. Los diseños se generan de la siguiente manera: donde aparezca un 1 ó -1, se sustituye el valor de las desviaciones dadas por el nivel especificado (1 ó -1), y donde aparezca 0, se sustituye el valor de la propiedad sin error. En el siguiente capítulo se presentan los resultados de cada uno de los diseños en forma de una tabla resumen, junto con información del análisis ANOVA realizado a los resultados del diseño experimental.

Capítulo 8

Análisis de resultados

Este capítulo está comprendido en dos secciones, primero se analizan los resultados de las realizaciones de la historia de presión y de los parámetros PVT de los casos evaluados; y la segunda sección presenta los resultados de los diseños experimentales aplicados. Los resultados se presentan en forma de cuadros resúmenes con información acerca del promedio, mediana, desviación estándar, y otras estadísticas importantes. En el apéndice B y apéndice C se muestran casos ejemplos para las realizaciones y el diseño experimental para el caso 5, respectivamente. Al final de cada sección se presenta un resumen de los aspectos más importantes de los resultados obtenidos en esa sección.

8.1. Realizaciones

A continuación se presentan los resultados de las realizaciones para cada uno de los parámetros PVT y la historia de presión. Para cada uno de los modelos de yacimiento se presenta una tabla resumen de los resultados, junto con una gráfica del error en el POES vs. la desviación estándar del parámetro evaluado, y una gráfica del error vs. el número de realizaciones calculadas. Luego de esta sección se muestra los gráficos ternarios del error para cada uno de los parámetros PVT y la historia de presión como se establece en la metodología.

8.1.1. Historia de presión

Caso 1

Método: (F-We vs. Et)	$\sigma = 1$ psi	$\sigma = 100$ psi
POES (promedio) [MMSTB]	100.03	100.68
POES (mediana) [MMSTB]	100.02	99.52
POES (máximo) [MMSTB]	100.19	120.37
POES (mínimo) [MMSTB]	99.88	83.43
POES (desviación estándar) [MMSTB]	0.07	7.34
RMSE (mínimo) [MMSTB]	0.00	0.31
RMSE (máximo) [MMSTB]	0.01	1.04

Cuadro 8.1: Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero

El caso 1 representa un yacimiento sin capa de gas y sin acuífero. Por tanto, el único método utilizado fue (F-We vs. Et). Los cálculos de balance de materiales muestran poca dispersión para $\sigma = 1$ psi, alcanzando un valor de desviación estándar de 0.07 MMSTB. Para un $\sigma = 100$ psi se observa que la desviación estándar aumentó hasta 7.34 MMSTB para la desviación estándar. La calidad del ajuste en los datos viene dado por el valor del RMSE, mientras este valor sea pequeño o cerca de cero se tiene un buen ajuste, a diferencia de cuando el valor del RMSE es grande en cuyo caso el ajuste a los datos se puede calificar como pobre. Para este caso, con un $\sigma = 1$ psi, los valores de RMSE obtenidos muestran un muy buen ajuste a los datos, mientras que para $\sigma = 100$ psi, disminuye la calidad del ajuste, aunque todavía sigue habiendo un muy buen ajuste a los datos. El gráfico de error en el POES versus la desviación estándar (gráfico 8.1), indica que el error esperado para el POES con una desviación estándar de 100 psi en la historia de presión es alrededor de 7%. El valor del error se muestra estable para todos los valores de realizaciones utilizados, alcanzando aproximadamente 7%, para un $\sigma = 100$ psi (gráfico 8.2).

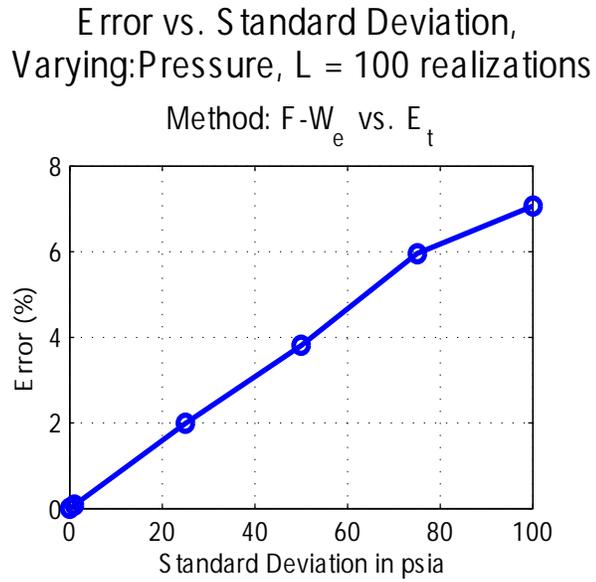


Figura 8.1: Gráfico del error en el POES vs. la desviación estándar de la historia de presión para el caso 1

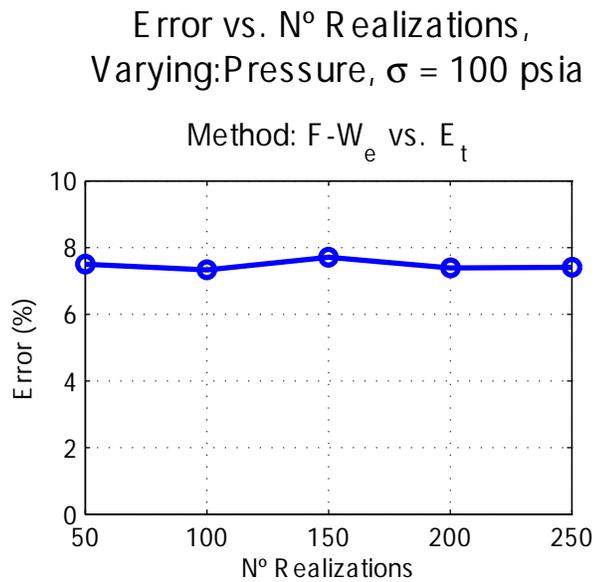


Figura 8.2: Gráfico del error en el POES vs. el número de realizaciones para el caso 1

Caso 5

Método: (F-We vs. Et)	$\sigma = 1$ psi	$\sigma = 100$ psi
POES (promedio) [MMSTB]	100.03	100.35
POES (mediana) [MMSTB]	99.99	96.99
POES (máximo) [MMSTB]	100.50	157.72
POES (mínimo) [MMSTB]	99.81	70.74
POES (desviación estándar) [MMSTB]	0.13	16.60
RMSE (mínimo) [MMSTB]	0.01	0.45
RMSE (máximo) [MMSTB]	0.02	1.94
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	$\sigma = 1$ psi	$\sigma = 100$ psi
POES (promedio) [MMSTB]	120.90	178.95
POES (mediana) [MMSTB]	110.40	789.44
POES (máximo) [MMSTB]	313.65	10566.30
POES (mínimo) [MMSTB]	25.27	-63652.52
POES (desviación estándar) [MMSTB]	55.01	7388.01
RMSE mínimo [MMSTB]	0.36	34.92
RMSE máximo [MMSTB]	4.24	5938.50

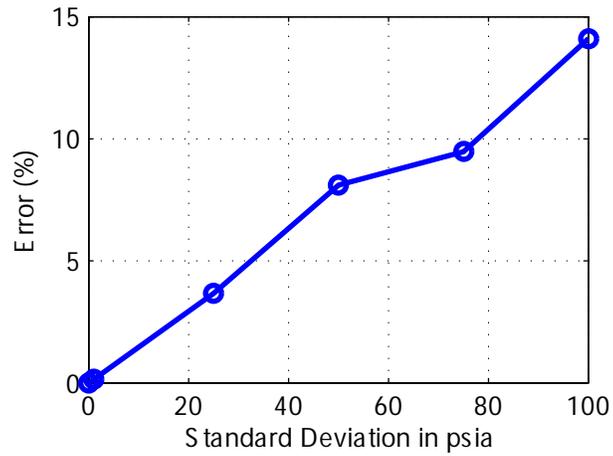
Cuadro 8.2: Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero

El caso 5 simula un yacimiento con una gran capa de gas ($m = 1$), y sin acuífero. Los métodos utilizados para el balance de materiales fueron: (F-We vs. Et) y (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) (método de la capa de gas). Para un $\sigma = 1$ psi se obtuvo una desviación estándar de 0.13 MMSTB para el método de (F-We vs. Et) y una desviación estándar de 55.01 MMSTB para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw). Para un $\sigma = 100$ psi la desviación estándar fue 16.60 MMSTB para el método (F-We vs. Et), y de 7388.01 MMSTB para el método de la capa de gas. Es de hacer notar el valor obtenido como el POES mínimo calculado para el método de la capa de gas con un $\sigma = 100$ psi, claramente refleja un resultado no físico (el valor es negativo), además de la gran diferencia en magnitud con respecto al POES real de 100 MMSTB. En cuanto a los valores del RMSE obtenidos se tiene que el método (F-We vs. Et) muestra un muy buen ajuste a los datos para los dos valores de desviación evaluados. Para el método de la capa de gas, con $\sigma = 1$ psi se consiguen buenos ajustes a los datos; sin embargo, para un $\sigma = 100$ psi, los valores de RMSE son altos (y aún más alto el valor obtenido para el máximo RMSE) indicando que resultaron ajustes muy pobres a los datos. El gráfico 8.3 señala que para el método (F-We vs. Et) con un $\sigma = 100$ psi se obtendrá un error promedio en el POES de 14 %, mientras que para el otro método se tendría un error de 4000 % con la misma desviación en la presión ($\sigma = 100$ psi). El gráfico 8.4 indica que para el primer método (F-We

vs. Et) el error permanece estable para todos los valores de realizaciones tomados (alrededor de 14-15%). El método de la capa de gas muestra poca estabilidad en el error para los valores de realizaciones tomados, el promedio de los errores obtenidos está alrededor de 6000%.

Error vs. Standard Deviation,
Varying: Pressure, L = 100 realizations

Method: $F - W_e$ vs. E_t



Method: $(F - W_e)/(E_o + E_{fw})$ vs. $(E_g + E_{fw})/(E_o + E_{fw})$

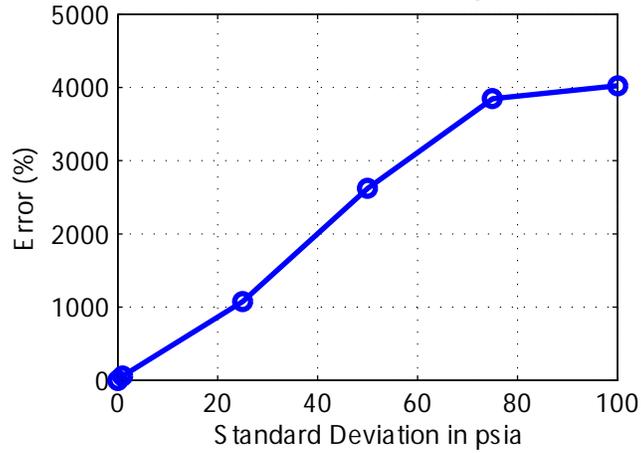
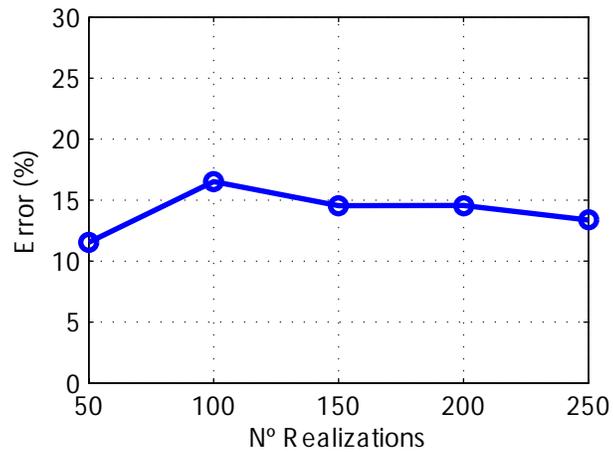


Figura 8.3: Gráfico del error en el POES vs. la desviación estándar de la historia de presión para el caso 5

Error vs. N° Realizations,
 Varying: Pressure, $\sigma = 100$ psia

Method: $F - W_e$ vs. E_t



Method: $(F - W_e)/(E_o + E_{fw})$ vs. $(E_g + E_{fw})/(E_o + E_{fw})$

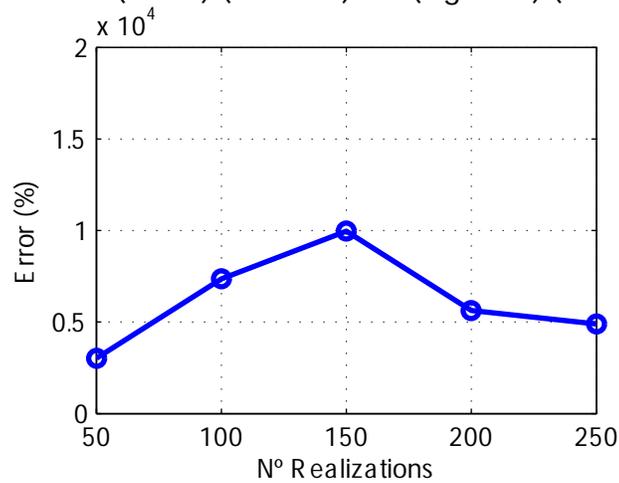


Figura 8.4: Gráfico del error en el POES vs. el número de realizaciones para el caso 5

Caso 11

Método: (F-We vs. Et)	$\sigma = 1$ psi	$\sigma = 100$ psi
POES (promedio) [MMSTB]	99.75	72.63
POES (mediana) [MMSTB]	99.75	65.01
POES (máximo) [MMSTB]	102.98	264.47
POES (mínimo) [MMSTB]	96.24	-89.08
POES (desviación estándar) [MMSTB]	1.30	73.29
RMSE (mínimo) [MMSTB]	0.01	0.50
RMSE (máximo) [MMSTB]	0.03	2.85
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	$\sigma = 1$ psi	$\sigma = 100$ psi
POES (promedio) [MMSTB]	82.58	142.44
POES (mediana) [MMSTB]	85.11	533.92
POES (máximo) [MMSTB]	355.54	24290.99
POES (mínimo) [MMSTB]	-248.03	-31863.80
POES (desviación estándar) [MMSTB]	95.49	6806.53
RMSE mínimo [MMSTB]	0.75	60.13
RMSE máximo [MMSTB]	8.57	12639.43
Método: (F/Et vs. We/Et)	$\sigma = 1$ psi	$\sigma = 100$ psi
POES (promedio) [MMSTB]	99.18	119.80
POES (mediana) [MMSTB]	99.14	82.15
POES (máximo) [MMSTB]	106.31	9753.23
POES (mínimo) [MMSTB]	93.27	-3815.39
POES (desviación estándar) [MMSTB]	2.65	1227.72
RMSE mínimo [MMSTB]	0.78	55.46
RMSE máximo [MMSTB]	6.93	28449.76

Cuadro 8.3: Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$

El caso 11 representa un yacimiento con una capa de gas de 0.25, y un acuífero con un $r_e D = 5$. Por tanto, los métodos de balance de materiales utilizados fueron: (F-We vs. Et), $(F-We)/(E_o+E_{fw})$ vs. $(E_g+E_{fw})/(E_o+E_{fw})$, y (F/Et vs. We/Et). Con un $\sigma = 1$ psi se obtuvo una valor de desviación estándar para el método (F-We vs. Et) de 1.30 MMSTB, 95.49 MMSTB para el método de la capa de gas, y de 2.65 MMSTB para el método (F/Et vs. We/Et). Para un $\sigma = 100$ psi se tiene que los valores de desviaciones estándar obtenidos fueron: 73.29 MMSTB para el método (F-We vs. Et), 6806.53 MMSTB para el método de la capa de gas, y 1227.72 MMSTB para el método (F/Et vs. We/Et). Se hace resaltar que para una desviación de $\sigma = 100$ psi, los valores del POES mínimo calculado son valores negativos para todos los métodos, específicamente, -89.08 MMSTB para el método (F-We vs. Et), -31863.80 MMSTB para el método de la capa de gas, y -3815.39 MMSTB para el método (F/Et vs. We/Et). Aún mas, el POES mínimo calculado para el método de la capa de gas, con un $\sigma = 1$ psi fue de -248.03 MMSTB. Para los valores del RMSE reportados, el método (F-We vs. Et) muestra valores pequeños indicando muy buenos ajustes a los datos. El método de capa de gas, con $\sigma = 1$ psi, indica ajustes aceptables a los datos (se hace notar que la calidad del ajuste para este método resulta inferior si se compara con el método (F-We vs. Et), debido a las diferencias en los valores del RMSE); a diferencia de cuando $\sigma = 100$ psi, en donde se tienen ajustes muy pobres a los datos. Para el método (F/Et vs. We/Et), para un $\sigma = 1$ psi, se tienen ajustes aceptables a los datos; entretanto para un $\sigma = 100$ psi, los ajustes resultan incluso de calidad inferior al método de capa de gas (pudiéndose observar en los valores máximos del RMSE). El gráfico 8.5 muestra que el error promedio para el método (F-We vs. Et) con un $\sigma = 100$ psi es de aproximadamente 70 %. Para el método $(F-We)/(E_o+E_{fw})$ vs. $(E_g+E_{fw})/(E_o+E_{fw})$ el error está alrededor de 6000-6500 %, mientras que para el último método el error alcanzó un valor de 1000 % (para un $\sigma = 100$ psi). El gráfico 8.6 muestra el error vs. el número de realizaciones para este caso. Se puede observar que todos los casos tuvieron errores bastante estables. El método (F-We vs. Et) obtuvo un error promedio de 80 %, $(F-We)/(E_o+E_{fw})$ vs. $(E_g+E_{fw})/(E_o+E_{fw})$ con un error promedio de 6000 %, y (F/Et vs. We/Et) con un error promedio alrededor de 1200-1300 %.

Error vs. Standard Deviation, Varying: Pressure, L = 100 realizations

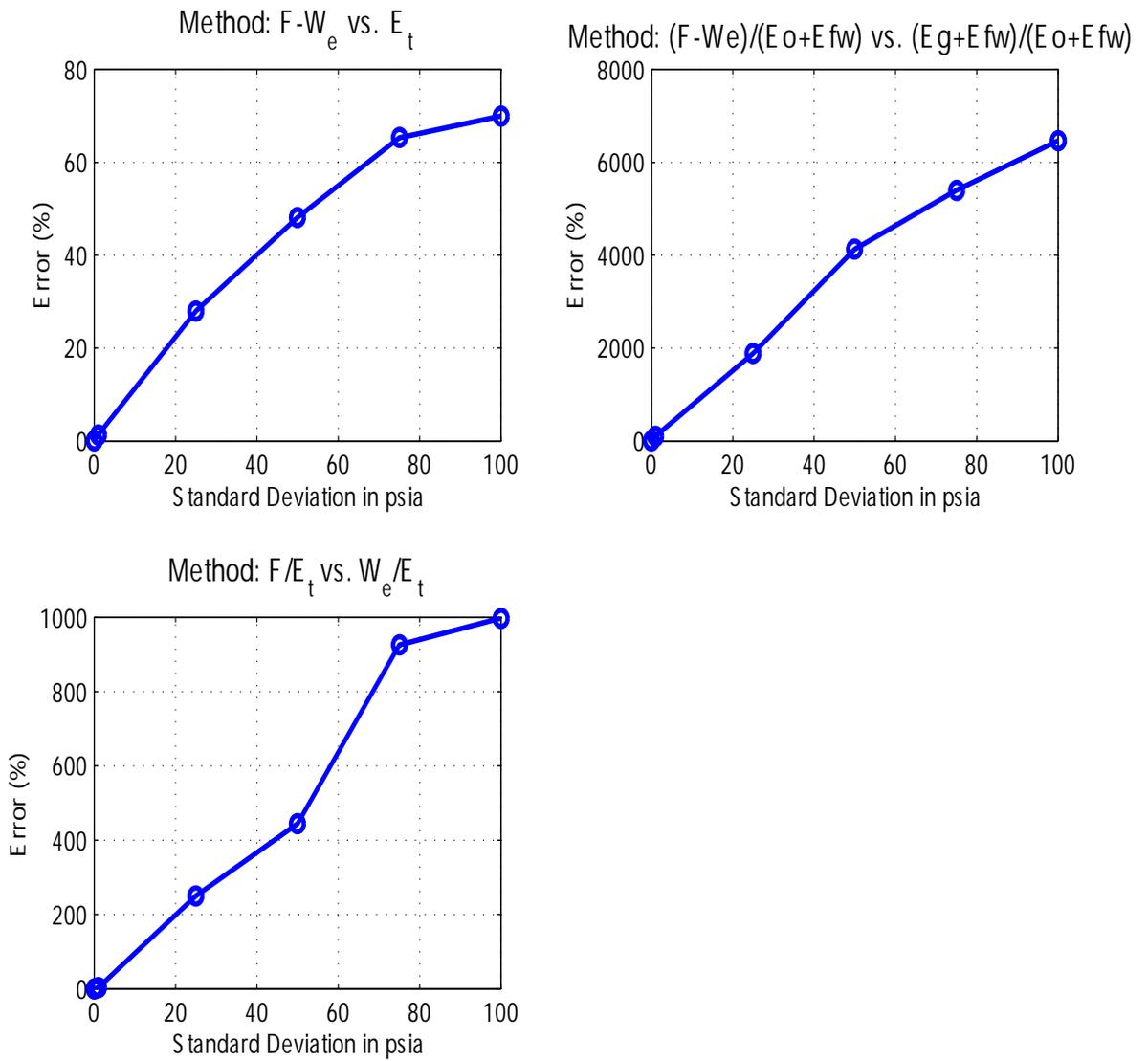


Figura 8.5: Gráfico del error en el POES vs. la desviación estándar de la historia de presión para el caso 11

Error vs. N° Realizations, Varying: Pressure, $\sigma = 100$ psia

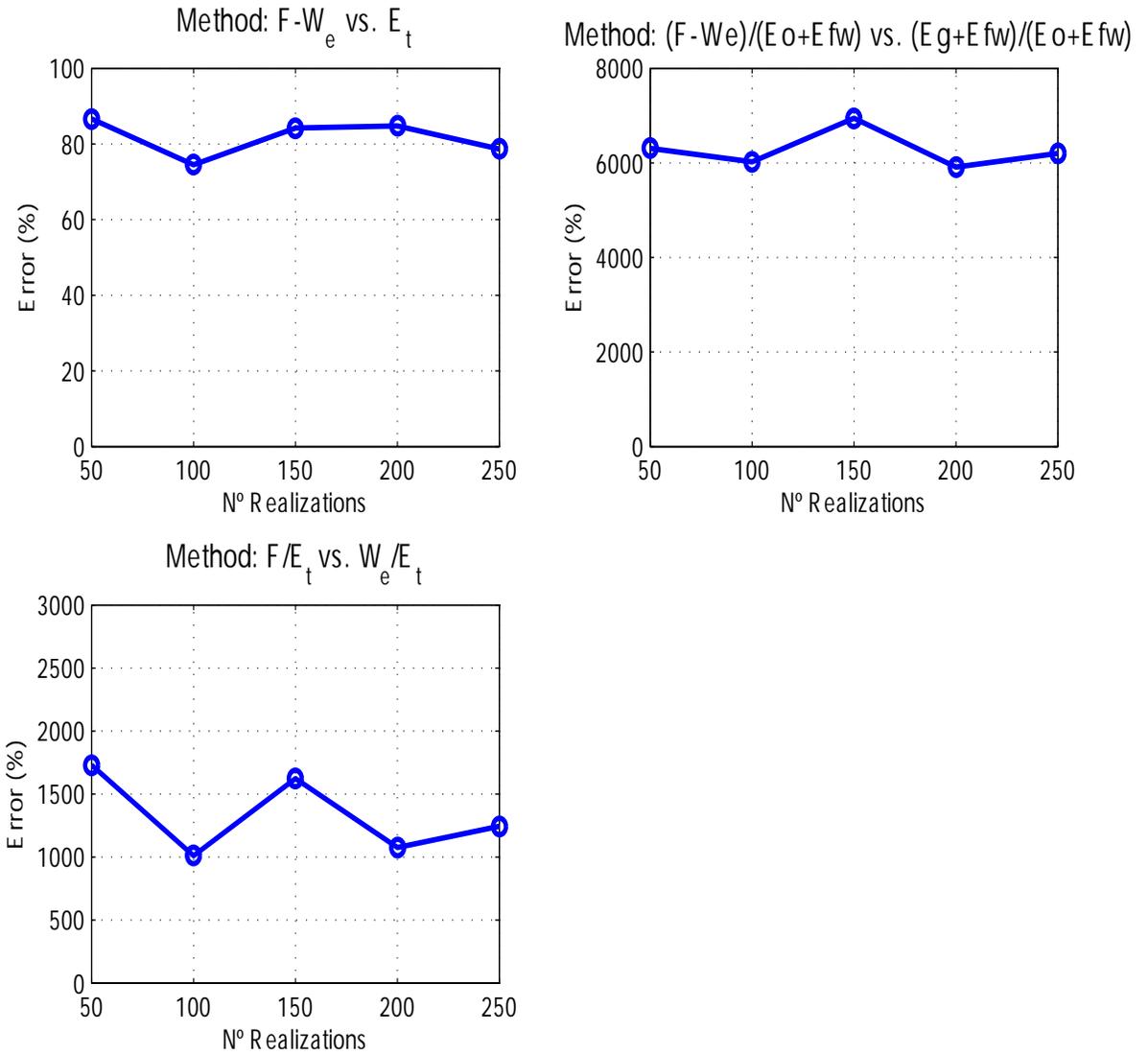


Figura 8.6: Gráfico del error en el POES vs. el número de realizaciones para el caso 11

Caso 15

Método: (F-We vs. Et)	$\sigma = 1$ psi	$\sigma = 100$ psi
POES (promedio) [MMSTB]	96.51	-756.54
POES (mediana) [MMSTB]	98.12	-512.35
POES (máximo) [MMSTB]	167.72	1590.05
POES (mínimo) [MMSTB]	19.05	-3604.78
POES (desviación estándar) [MMSTB]	26.05	1216.53
RMSE (mínimo) [MMSTB]	0.01	0.46
RMSE (máximo) [MMSTB]	0.06	11.44
Método: (F/Et vs. We/Et)	$\sigma = 1$ psi	$\sigma = 100$ psi
POES (promedio) [MMSTB]	96.58	-59.12
POES (mediana) [MMSTB]	98.42	-765.34
POES (máximo) [MMSTB]	161.42	146655.27
POES (mínimo) [MMSTB]	26.04	-43541.51
POES (desviación estándar) [MMSTB]	24.06	16407.87
RMSE mínimo [MMSTB]	2.22	232.69
RMSE máximo [MMSTB]	23.14	461176.99

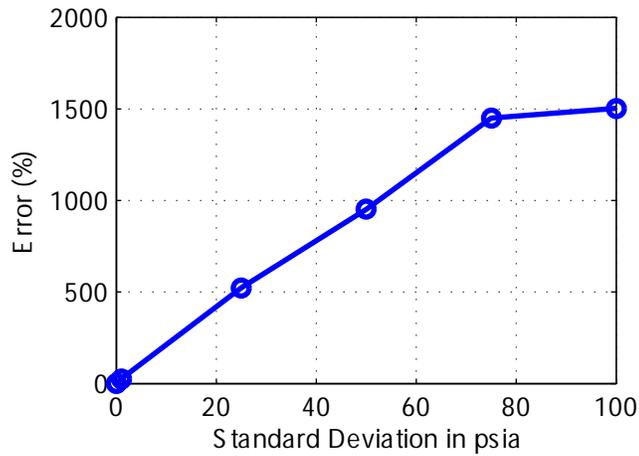
Cuadro 8.4: Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento, $r_e D = \infty$

El modelo de yacimiento del caso 15 no posee capa de gas ($m = 0$), pero sí un acuífero con un $r_e D = \infty$. Es por ello que se utilizaron los métodos (F-We vs. Et) y (F/Et vs. We/Et). Con un $\sigma = 1$ psi en la historia de presión se obtuvo los siguientes valores de desviación estándar para los POES: 26.05 MMSTB para el método (F-We vs. Et), y 24.06 MMSTB para el método (F/Et vs. We/Et). Para un $\sigma = 100$ psi los valores de desviación estándar fueron: 1216.53 MMSTB para el método (F-We vs. Et) y de 16407.87 para el método (F/Et vs. We/Et). Nuevamente, para este caso se obtuvieron valores negativos para los POES mínimos calculados, -3604.78 MMSTB para el método (F-We vs. Et), y -43541.51 MMSTB para el método (F/Et vs. We/Et). Es importante señalar que la tendencia central para los dos métodos, con un $\sigma = 100$ psi, es hacia valores negativos reflejándose en los valores de los promedios de los POES; -56.54 MMSTB para el método (F-We vs. Et), y -59.12 MMSTB para el método (F/Et vs. We/Et). Entre los dos métodos el más robusto fue (F-We vs. Et), mientras que el otro método, para un $\sigma = 100$ psi, superó en más de un orden de magnitud la desviación estándar del método (F-We vs. Et). Los valores del RMSE, en general, muestran para ambos métodos un aumento en comparación a los otros casos indicando ajustes más pobres a los datos, y se acentúa más para el método (F/Et vs. We/Et) con $\sigma = 100$

psi. El método (F-We vs. Et) también muestra una disminución en la calidad del ajuste si se compara a los casos anteriores con $\sigma = 100$ psi. En el gráfico 8.7 se muestra el error promedio versus la desviación estándar en la historia de presión. El método (F-We vs. Et) obtuvo un error promedio de 1500 %, mientras que el método (F/Et vs. We/Et) obtuvo un error de 10000 %. El gráfico 8.8 muestra el error promedio en los POES versus el número de realizaciones. El método (F-We vs. Et) obtuvo un error promedio de 1500 % para los valores de realizaciones tomados. Para el método (F/Et vs. We/Et) el error promedio varió entre 10000 y 11000 %.

Error vs. Standard Deviation,
Varying: Pressure, L = 100 realizations

Method: $F - W_e$ vs. E_t



Method: F/E_t vs. W_e/E_t

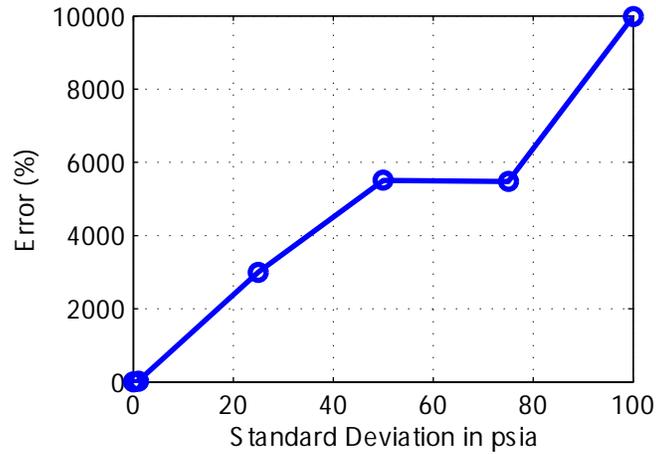


Figura 8.7: Gráfico del error en el POES vs. la desviación estándar de la historia de presión para el caso 15

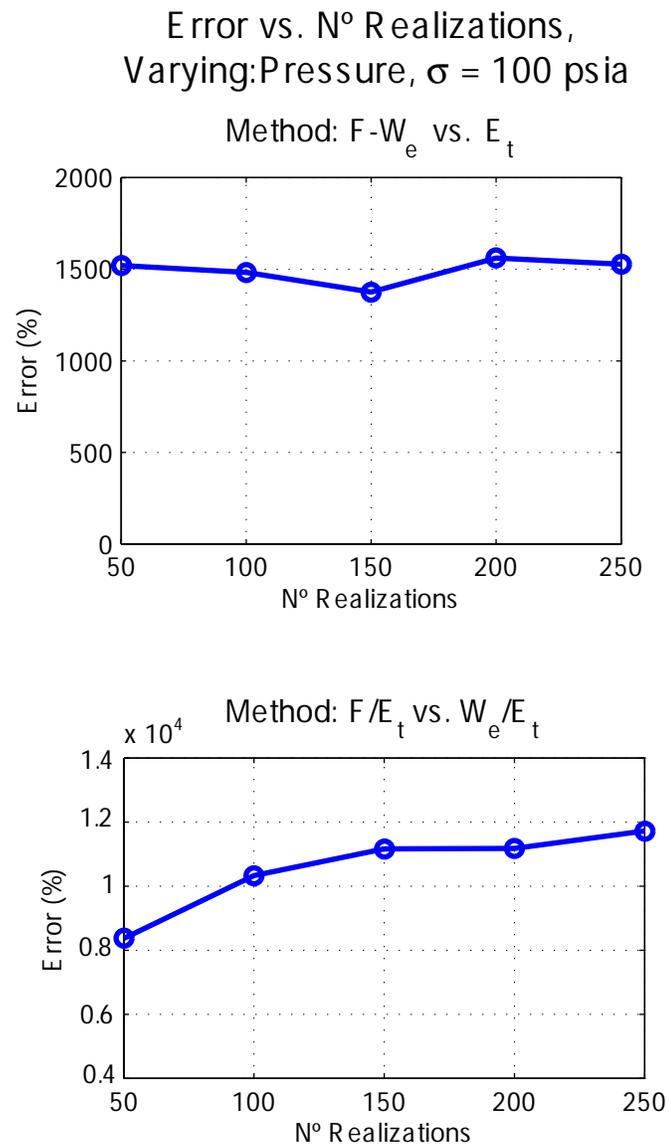


Figura 8.8: Gráfico del error en el POES vs. el número de realizaciones para el caso 15

Análisis final para la historia de presión

En el siguiente cuadro se presenta el promedio de las desviaciones estándar obtenidas para cada método de acuerdo a la desviación utilizada en las realizaciones de la historia de presión. En base a este resultado se clasifican los métodos de balance de materiales de acuerdo a su sensibilidad a la desviación en la historia de presión.

Método de balance de materiales	Método 1	Método 2	Método 3
Promedio de la desviación estándar de los POES para un $\sigma = 1$ psi	6.89	75.25	13.355
Promedio de la desviación estándar de los POES para un $\sigma = 100$ psi	328.44	7097.27	8817.80
Método 1: F-We vs. Et			
Método 2: $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$			
Método 3: F/Et vs. We/Et			

Cuadro 8.5: Desviación estándar de los POES para la realizaciones de la historia de presión

De acuerdo a estos resultados el método más robusto fue (F-We vs. Et), para todos los casos. El método (F/Et vs. We/Et) también muestra poca dispersión en los POES para desviaciones bajas. El método $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$ muestra gran dispersión en los POES con apenas $\sigma = 1$ psi en la historia de presión. Para desviaciones más grandes éstos dos métodos de balance de materiales se muestran muy sensibles a las desviaciones en la historia de presión. La tendencia en cuanto a los valores del RMSE es que a medida que aumenta el tamaño de la capa de gas o el influjo de agua aumenta el valor del RMSE del ajuste, indicando que se obtendrá un ajuste má pobre en los datos. También se observa que aún para grandes desviaciones en la historia de presión se pueden obtener valores del RMSE suficientemente bajos como para tener un buen ajuste a los datos no implicando que el valor calculado del POES sea el correcto. EL mejor ejemplo de ello se puede observar en el caso 15 para el método (F-We vs. Et) con un $\sigma = 100$ psi, donde el promedio del POES calculado fue negativo aún cuando se obtuvieron valores del RMSE tan bajos como 0.46 MMSTB.

8.1.2. Gravedad API

Caso 1

Método: (F-We vs. Et)	$\sigma = 0,05 \text{ } ^\circ\text{API}$	$\sigma = 0,25 \text{ } ^\circ\text{API}$
POES (promedio) [MMSTB]	100.11	100.60
POES (mediana) [MMSTB]	100.11	100.52
POES (máximo) [MMSTB]	100.32	102.08
POES (mínimo) [MMSTB]	99.99	100.02
POES (desviación estándar) [MMSTB]	0.08	0.49
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.11

Cuadro 8.6: Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero

De la misma manera como para la historia de presión, el único método utilizado fue (F-We vs. Et). Los cálculos de balance de materiales muestran un valor de la desviación estándar en los POES calculados de 0.08 MMSTB, con $\sigma = 0,05 \text{ } ^\circ\text{API}$. Para un $\sigma = 0,25 \text{ } ^\circ\text{API}$ se observa que la desviación estándar aumentó a 0.49 MMSTB. Los valores del RMSE obtenidos para ambos valores de desviación indican que los ajustes realizados fueron muy buenos. El gráfico de error en el POES versus la desviación estándar (gráfico 8.9), indica que el error esperado para el POES con una desviación estándar de $0,05 \text{ } ^\circ\text{API}$ es alrededor de 0.8%. Para un $\sigma = 0,25 \text{ } ^\circ\text{API}$ el valor del error se muestra estable para todos los valores de realizaciones utilizados alcanzando aproximadamente 0.8% (gráfico 8.10).

Error vs. Standard Deviation,
Varying: API degrees, L = 100 realizations

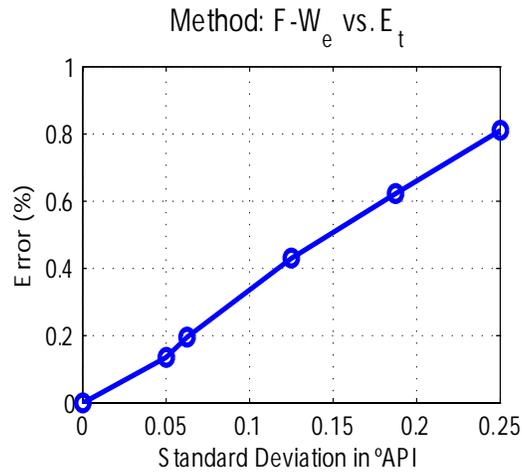


Figura 8.9: Gráfico del error en el POES vs. la desviación estándar de la gravedad API para el caso 1

Error vs. Nº Realizations,
Varying: API degrees, $\sigma = 0.25$ °API

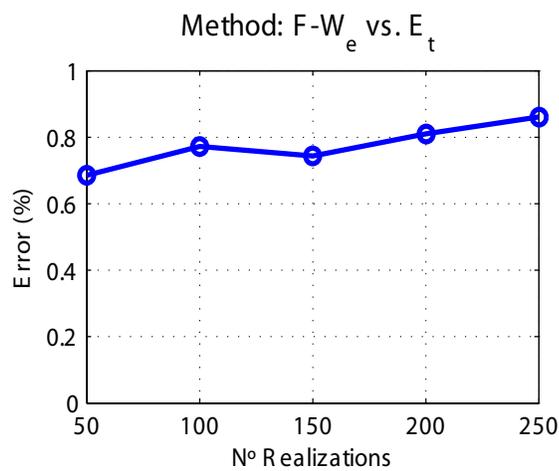


Figura 8.10: Gráfico del error en el POES vs. el número de realizaciones para el caso 1

Caso 5

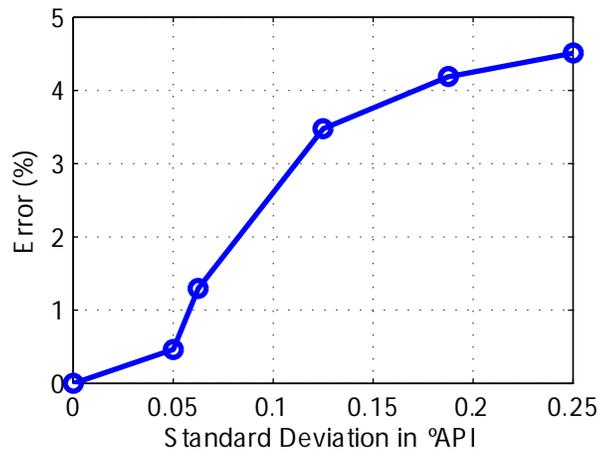
Método: (F-We vs. Et)	$\sigma = 0,05$ °API	$\sigma = 0,25$ °API
POES (promedio) [MMSTB]	100.35	103.92
POES (mediana) [MMSTB]	100.29	103.11
POES (máximo) [MMSTB]	101.44	109.09
POES (mínimo) [MMSTB]	99.95	100.03
POES (desviación estándar) [MMSTB]	0.30	3.11
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.04	0.14
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	$\sigma = 0,05$ [°API]	$\sigma = 0,25$ [°API]
POES (promedio) [MMSTB]	76.07	118.52
POES (mediana) [MMSTB]	93.11	119.92
POES (máximo) [MMSTB]	128.89	234.95
POES (mínimo) [MMSTB]	-134.66	-138.68
POES (desviación estándar) [MMSTB]	51.46	92.32
RMSE mínimo [MMSTB]	0.00	0.03
RMSE máximo [MMSTB]	5.73	6.02

Cuadro 8.7: Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero

Para un $\sigma = 0,05$ °API se obtuvo una desviación estándar de 0.30 MMSTB para el método de (F-We vs. Et); y una desviación estándar de 51.46 MMSTB para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw). Para un $\sigma = 0,25$ °API la desviación estándar fue 3.11 MMSTB para el método (F-We vs. Et), y de 92.32 MMSTB para el método de la capa de gas. Los valores del RMSE para ambos métodos con los dos valores de desviación arrojan que, en general, se obtuvieron ajustes bastante aceptables en los datos. El gráfico 8.11 indica que para el método (F-We vs. Et), con un $\sigma = 0,25$ °API, se obtendrá un error promedio de 4.5 %, mientras que para el otro método se tendría un error de, aproximadamente, 90 % con la misma desviación ($\sigma = 0,25$ °API). El gráfico 8.12 indica que para el primer método (F-We vs. Et) el error permanece entre 4.5 y 5 % para todos los valores de realizaciones tomados. El método de la capa de gas varía entre 80-90 % del error para los valores de realizaciones tomados.

Error vs. Standard Deviation,
Varying: API degrees, L = 100 realizations

Method: $F - W_e$ vs. E_t



Method: $(F - W_e)/(E_o + E_{fw})$ vs. $(E_g + E_{fw})/(E_o + E_{fw})$

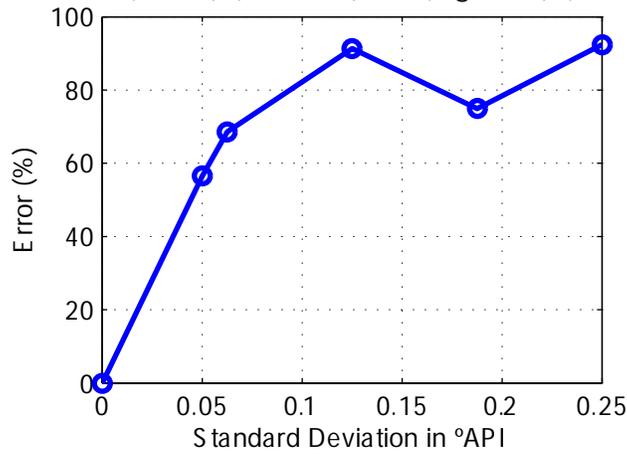
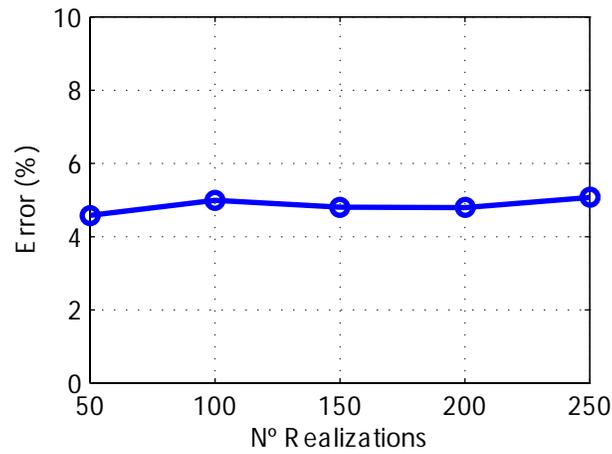


Figura 8.11: Gráfico del error en el POES vs. la desviación estándar de la gravedad API para el caso 5

Error vs. N° Realizations,
 Varying: API degrees, $\sigma = 0.25^\circ \text{API}$
 Method: $F - W_e$ vs. E_t



Method: $(F - W_e) / (E_o + E_{fw})$ vs. $(E_g + E_{fw}) / (E_o + E_{fw})$

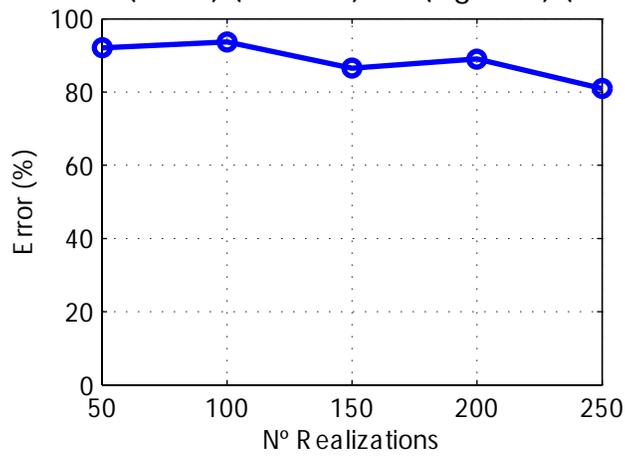


Figura 8.12: Gráfico del error en el POES vs. el número de realizaciones para el caso 5

Caso 11

Método: (F-We vs. Et)	$\sigma = 0,05$ °API	$\sigma = 0,25$ °API
POES (promedio) [MMSTB]	100.30	103.02
POES (mediana) [MMSTB]	100.21	101.80
POES (máximo) [MMSTB]	101.71	109.42
POES (mínimo) [MMSTB]	99.63	99.67
POES (desviación estándar) [MMSTB]	0.39	2.81
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.04
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	$\sigma = 0,05$ [°API]	$\sigma = 0,25$ [°API]
POES (promedio) [MMSTB]	99.62	113.69
POES (mediana) [MMSTB]	100.83	107.34
POES (máximo) [MMSTB]	186.91	187.30
POES (mínimo) [MMSTB]	38.76	41.14
POES (desviación estándar) [MMSTB]	31.05	24.74
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	2.50	2.63
Método: (F/Et vs. We/Et)	$\sigma = 0,05$ °API	$\sigma = 0,25$ °API
POES (promedio) [MMSTB]	100.11	102.60
POES (mediana) [MMSTB]	100.12	101.80
POES (máximo) [MMSTB]	103.48	113.95
POES (mínimo) [MMSTB]	98.45	97.86
POES (desviación estándar) [MMSTB]	0.82	3.22
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	2.92	9.14

Cuadro 8.8: Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$

Para un $\sigma = 0,05$ °API se obtuvo una valor de desviación estándar para el método (F-We vs. Et) de 0.39 MMSTB, 31.05 MMSTB para el método de la capa de gas, y de 0.82 MMSTB para el método (F/Et vs. We/Et). Con un $\sigma = 0,25$ °API, los valores obtenidos de desviaciones estándar fueron: 2.81 MMSTB para el método (F-We vs. Et), 24.74 MMSTB para el método de la capa de gas, y 3.22 MMSTB para el método (F/Et vs. We/Et). Para el método (F-We vs. Et) los valores del RMSE indican que los ajustes realizados fueron muy buenos. De igual forma, para el método de la capa de gas y el método (F/Et vs. We/Et) los ajustes realizados fueron buenos. El ajuste más pobre obtenido fue para el método (F/Et vs. We/Et) con una desviación de $\sigma = 0,25$ °API y un RMSE de 9.14 MMSTB. El gráfico 8.13 muestra que el error promedio para el método (F-We vs. Et) es menos que 4.5 %, para un $\sigma = 0,25$ °API. Para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) el error está alrededor de 30 %, mientras que para el último método el error alcanzó un valor de 5 % (para un $\sigma = 0,25$ °API). El gráfico 8.14 muestra que todos los casos tuvieron errores bastante estables. El método (F-We vs. Et) obtuvo un error promedio de 4.5 %, el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) con un error promedio de 30 %, y (F/Et vs. We/Et) con un error promedio alrededor de 4-5 %.

Error vs. Standard Deviation, Varying: API degrees, L = 100 realizations

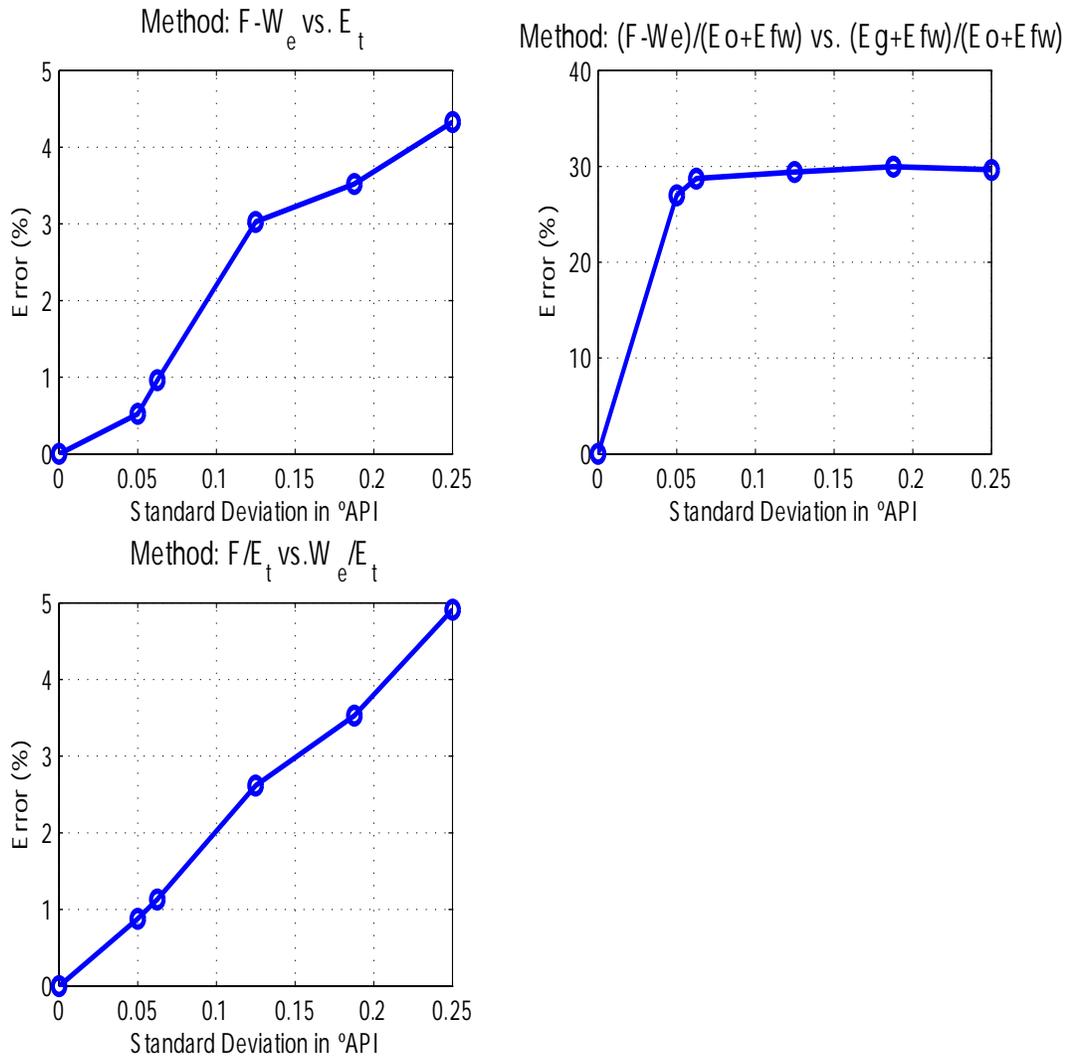


Figura 8.13: Gráfico del error en el POES vs. la desviación estándar de la gravedad API para el caso 11

Error vs. N° Realizations, Varying: API degrees, $\sigma = 0.25^\circ \text{API}$

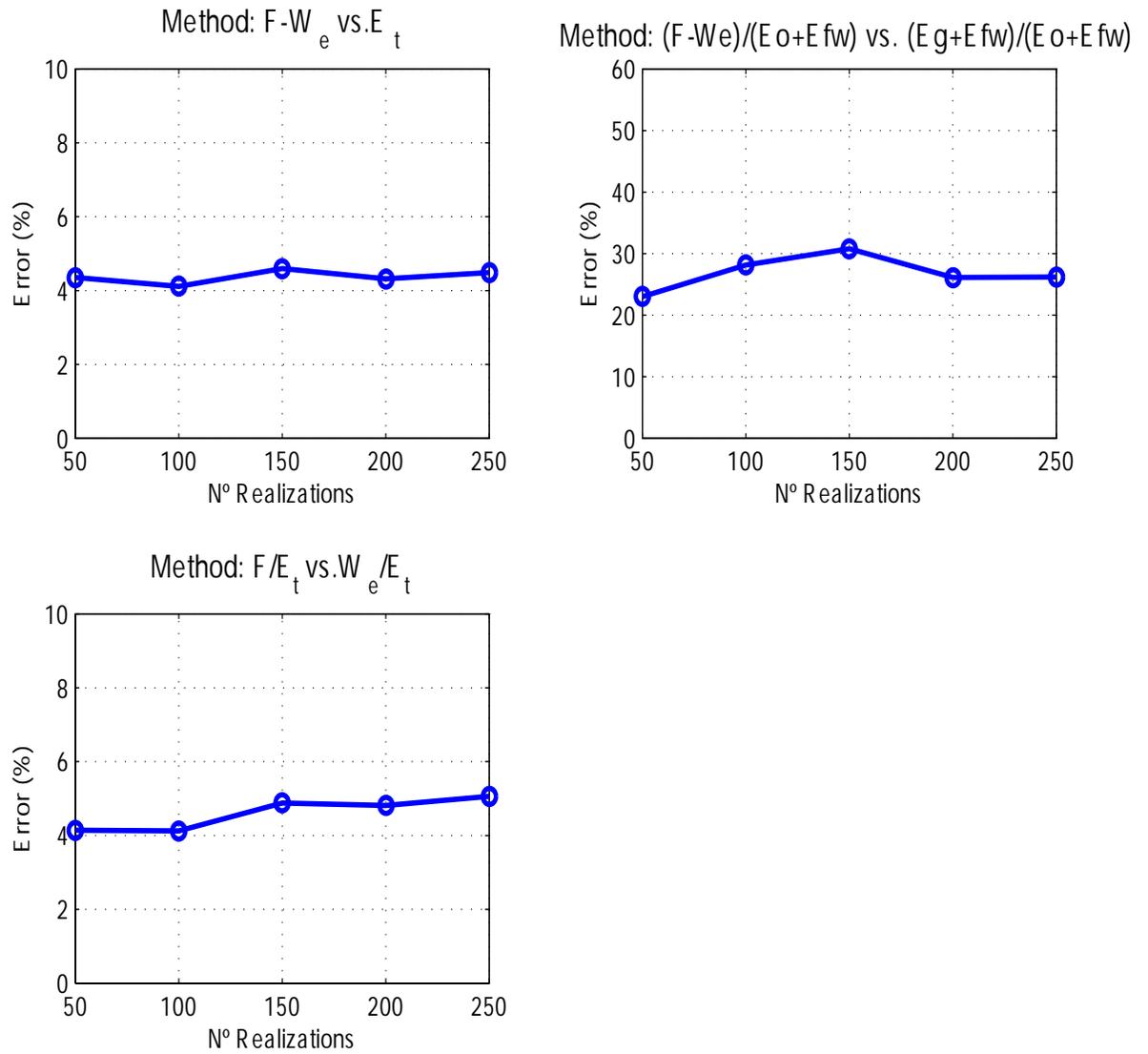


Figura 8.14: Gráfico del error en el POES vs. el número de realizaciones para el caso 11

Caso 15

Método: (F-We vs. Et)	$\sigma = 0,05$ °API	$\sigma = 0,25$ °API
POES (promedio) [MMSTB]	99.41	109.88
POES (mediana) [MMSTB]	100.14	102.39
POES (máximo) [MMSTB]	104.68	242.58
POES (mínimo) [MMSTB]	93.79	94.52
POES (desviación estándar) [MMSTB]	2.28	21.88
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.04
Método: (F/Et vs. We/Et)	$\sigma = 0,05$ °API	$\sigma = 0,25$ °API
POES (promedio) [MMSTB]	99.29	111.41
POES (mediana) [MMSTB]	100.13	102.05
POES (máximo) [MMSTB]	105.28	243.98
POES (mínimo) [MMSTB]	93.17	93.95
POES (desviación estándar) [MMSTB]	2.50	25.59
RMSE mínimo [MMSTB]	0.00	0.01
RMSE máximo [MMSTB]	1.83	39.49

Cuadro 8.9: Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento,

$$r_e D = \infty$$

Con un $\sigma = 0,05$ °API, se obtuvo los siguientes valores de desviación estándar para los POES: 2.28 MMSTB para el método (F-We vs. Et), y 2.50 MMSTB para el método (F/Et vs. We/Et). Para un $\sigma = 0,25$ °API los valores de desviación estándar fueron: 21.88 MMSTB para el método (F-We vs. Et) y de 25.59 para el método (F/Et vs. We/Et). Los valores del RMSE para el método (F-We vs. Et) indican que todos los ajustes resultaron bastante buenos, mientras que el peor ajuste obtenido fue para el método (F/Et vs. We/Et) con una desviación de $\sigma = 0,25$ °API con un RMSE de 39.49 MMSTB, un valor mucho mayor en comparación con todos los valores obtenidos en los casos anteriores. En el gráfico 8.15 se muestra el error promedio versus la desviación estándar para la gravedad API. El método (F-We vs. Et) obtuvo un error promedio de 18 %, mientras que el método (F/Et vs. We/Et) obtuvo un error de 20 %. El gráfico 8.16 muestra que el método (F-We vs. Et) obtuvo un error promedio de 23 % para los valores de realizaciones tomados. Para el método (F/Et vs. We/Et) el error promedio varió entre 25 y 30 %. Los gráficos de error vs. número de realizaciones se obtuvieron en diferentes realizaciones que el gráfico 8.15, por tal razón la diferencia en los valores de error reportados.

Error vs. Standard Deviation,
Varying: API degrees, L = 100 realizations

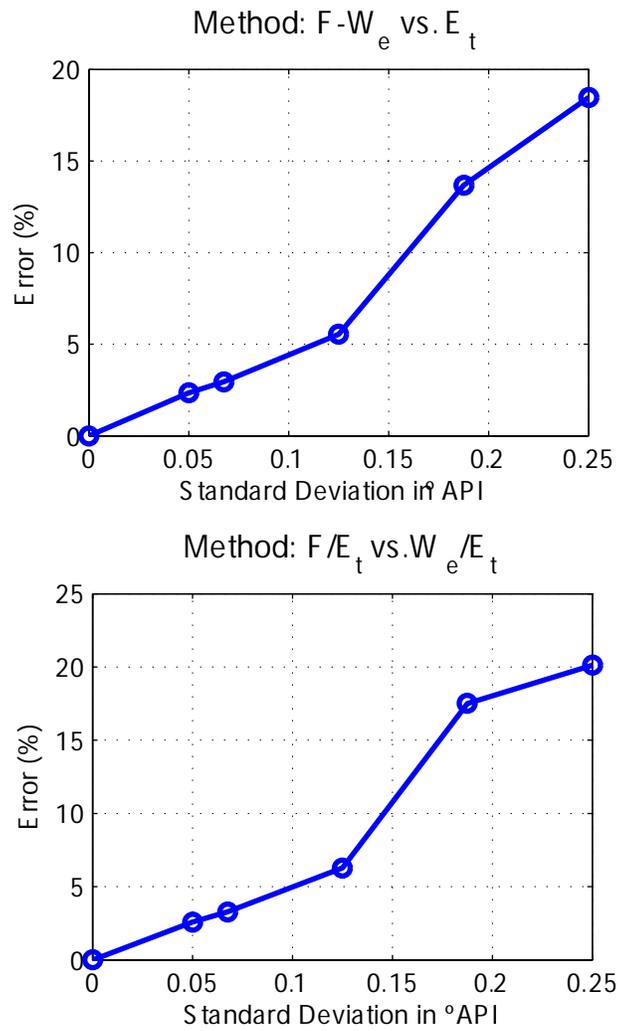
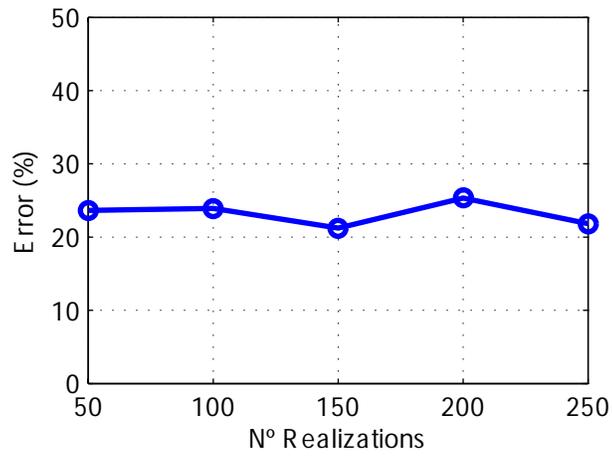


Figura 8.15: Gráfico del error en el POES vs. la desviación estándar de la gravedad API para el caso 15

Error vs. N° Realizations,
Varying: API degrees, $\sigma = 0.25 \text{ } ^\circ\text{API}$

Method: $F - W_e$ vs. E_t



Method: F/E_t vs. W_e/E_t

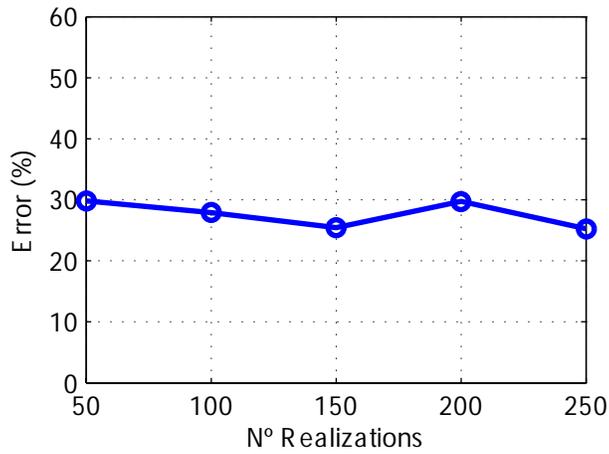


Figura 8.16: Gráfico del error en el POES vs. el número de realizaciones para el caso 15

Análisis final para la gravedad API

Como en el caso anterior de la historia de presión, en el siguiente cuadro se presenta el promedio de las desviaciones estándar obtenidas para cada método de acuerdo a la desviación utilizada en las realizaciones. En base a este resultado se clasifican los métodos de balance de materiales de acuerdo a su sensibilidad.

Método de balance de materiales	Método 1	Método 2	Método 3
Promedio de la desviación estándar de los POES para un $\sigma = 0,05$ °API	0.763	41.26	1.66
Promedio de la desviación estándar de los POES para un $\sigma = 0,25$ °API	7.07	58.53	14.41
Método 1: F-We vs. Et			
Método 2: $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$			
Método 3: F/Et vs. We/Et			

Cuadro 8.10: Desviación estándar de los POES para la realizaciones de la gravedad API

El método más robusto fue (F-We vs. Et), para todos los casos. El método (F/Et vs. We/Et) también muestra poca dispersión para todos los casos; sin embargo, la dispersión es 2 veces mayor que el método anterior. El método $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$ muestra bastante dispersión en los POES para todos los casos evaluados. En cuanto a los valores del RMSE se tiene que el método (F-We vs. Et) siempre obtuvo los valores más bajos de este parámetro mientras que los otros dos métodos obtuvieron resultados aceptables. El peor ajuste se obtuvo para el método (F/Et vs. We/Et) con una desviación de $\sigma = 0,25$ °API para el caso 15.

8.1.3. Gravedad específica del gas

Caso 1

Método: (F-We vs. Et)	$\sigma = 5e - 05$	$\sigma = 0,00615$
POES (promedio) [MMSTB]	100.00	100.26
POES (mediana) [MMSTB]	100.00	100.14
POES (máximo) [MMSTB]	100.01	101.80
POES (mínimo) [MMSTB]	99.99	99.51
POES (desviación estándar) [MMSTB]	0.01	0.51
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.05

Cuadro 8.11: Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero

Los cálculos de balance de materiales muestran un valor de la desviación estándar de 0.01 MMSTB, para un $\sigma = 5e - 05$. Para un $\sigma = 0,00615$ se observa que la desviación estándar aumenta hasta 0.51 MMSTB. Los valores del RMSE muestran que los ajustes realizados para ambos valores de desviación fueron excelentes. El gráfico de error en el POES versus la desviación estándar 8.17, indica que el error esperado para el POES con una desviación estándar de 0.00615 es de 0.5 %. El gráfico 8.18 muestra que el valor del error es estable para todos los valores de realizaciones utilizados, 0.6 % (para un $\sigma = 0,00615$).

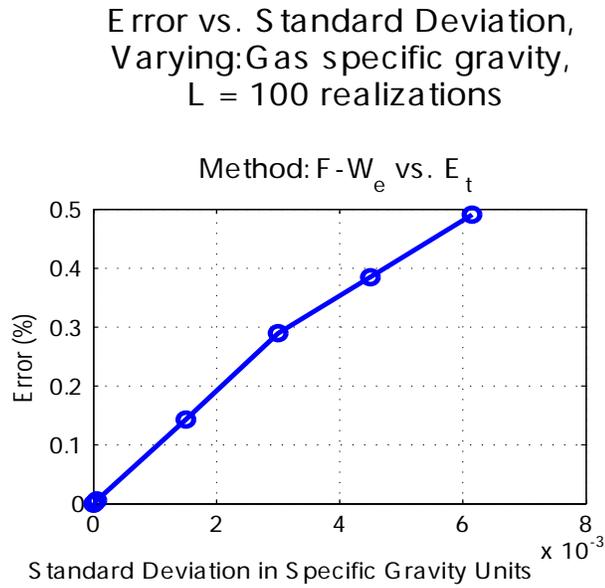


Figura 8.17: Gráfico del error en el POES vs. la desviación estándar de la gravedad específica del gas para el caso 1

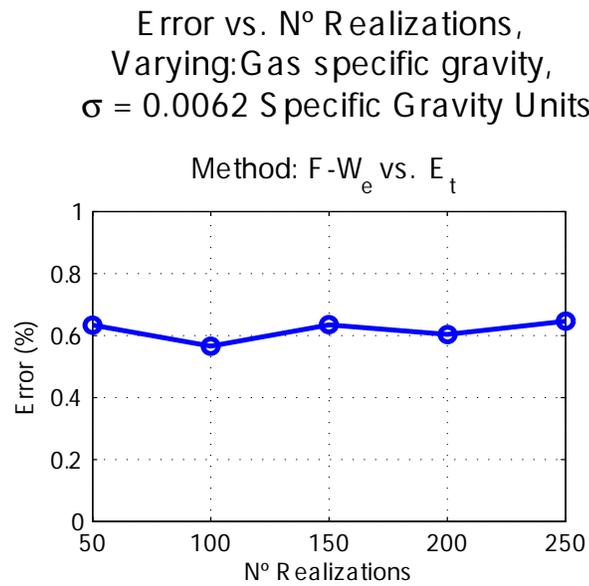


Figura 8.18: Gráfico del error en el POES vs. el número de realizaciones para el caso 1

Caso 5

Método: (F-We vs. Et)	$\sigma = 5e - 05$	$\sigma = 0,00615$
POES (promedio) [MMSTB]	100.00	102.36
POES (mediana) [MMSTB]	100.00	100.74
POES (máximo) [MMSTB]	100.04	108.16
POES (mínimo) [MMSTB]	99.95	99.98
POES (desviación estándar) [MMSTB]	0.02	2.77
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.24
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	$\sigma = 5e - 05$	$\sigma = 0,00615$
POES (promedio) [MMSTB]	99.93	153.29
POES (mediana) [MMSTB]	99.94	167.53
POES (máximo) [MMSTB]	100.54	501.27
POES (mínimo) [MMSTB]	99.36	-147.87
POES (desviación estándar) [MMSTB]	0.23	135.00
RMSE mínimo [MMSTB]	0.00	0.04
RMSE máximo [MMSTB]	0.03	7.06

Cuadro 8.12: Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero

Se obtuvo una desviación estándar de 0.02 MMSTB para el método de (F-We vs. Et) y una desviación estándar de 0.23 MMSTB para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw), con un $\sigma = 5e - 05$. Para un $\sigma = 0,00615$ la desviación estándar fue 2.77 MMSTB para el método (F-We vs. Et), y de 135.00 MMSTB para el método de la capa de gas. Los valores del RMSE para ambos métodos muestran ajustes bastante buenos, siendo el mayor valor del RMSE de 7.06 MMSTB para el método de la capa de gas (para una desviación de $\sigma = 0,00615$). El gráfico 8.19 señala que para el método (F-We vs. Et), con $\sigma = 0,00615$, se obtendrá un error promedio en el POES de 4.8%. Mientras que para el otro método se tendría un error de 160% con la misma desviación en la gravedad específica del gas ($\sigma = 0,00615$). El gráfico 8.20 indica que para el método (F-We vs. Et) el error permanece mas o menos estable para los valores de realizaciones tomados (alrededor de 2.5-3.5%). El método de la capa de gas muestra valores de error alrededor de 150-160%.

Error vs. Standard Deviation,
 Varying: Gas specific gravity, L = 100 realizations

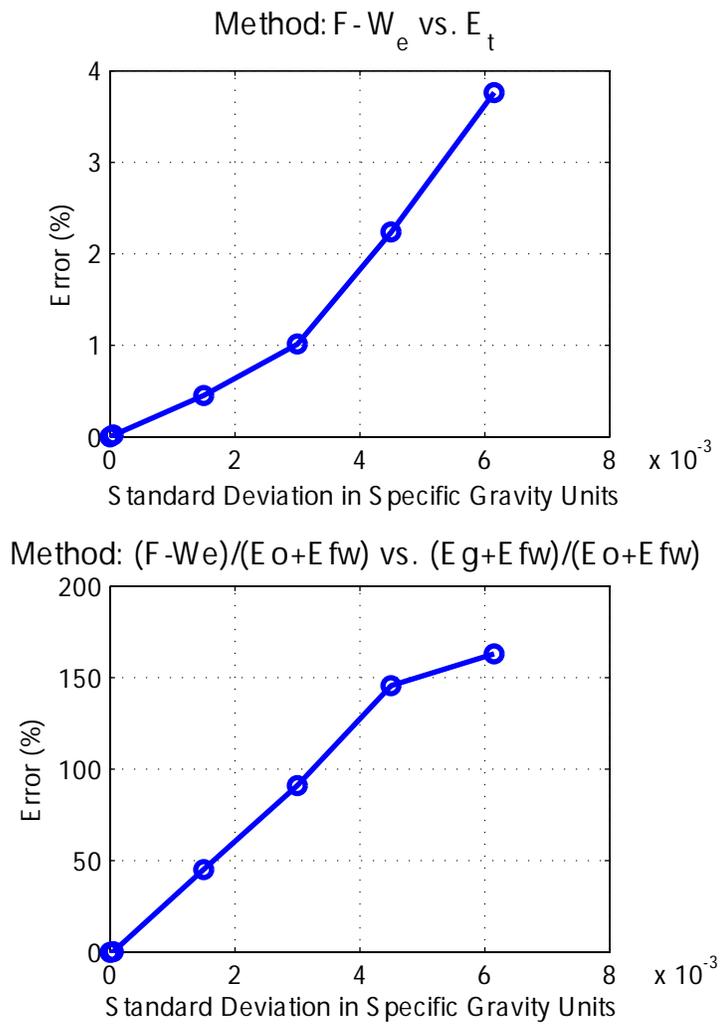
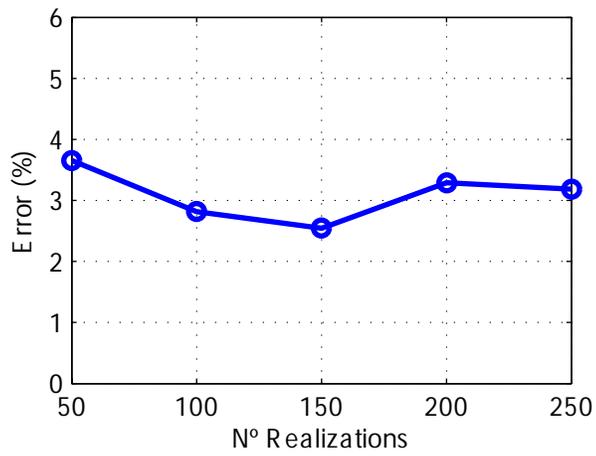


Figura 8.19: Gráfico del error en el POES vs. la desviación estándar de la gravedad específica del gas para el caso 5

Error vs. N° Realizations,
 Varying: Gas specific gravity,
 $\sigma = 0.0062$ Specific Gravity Units

Method: $F - W_e$ vs. E_t



Method: $(F - W_e)/(E_o + E_{fw})$ vs. $(E_g + E_{fw})/(E_o + E_{fw})$

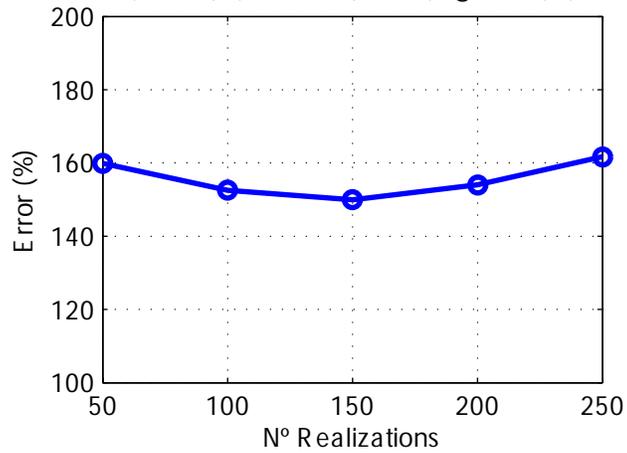


Figura 8.20: Gráfico del error en el POES vs. el número de realizaciones para el caso 5

Caso 11

Método: (F-We vs. Et)	$\sigma = 5e - 05$	$\sigma = 0,00615$
POES (promedio) [MMSTB]	100.00	101.82
POES (mediana) [MMSTB]	100.00	101.21
POES (máximo) [MMSTB]	100.03	107.62
POES (mínimo) [MMSTB]	99.97	99.63
POES (desviación estándar) [MMSTB]	0.02	1.99
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.04
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	$\sigma = 5e - 05$	$\sigma = 0,00615$
POES (promedio) [MMSTB]	100.00	118.80
POES (mediana) [MMSTB]	99.99	105.18
POES (máximo) [MMSTB]	100.09	229.95
POES (mínimo) [MMSTB]	99.91	41.25
POES (desviación estándar) [MMSTB]	0.03	36.67
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	0.01	2.50
Método: (F/Et vs. We/Et)	$\sigma = 5e - 05$	$\sigma = 0,00615$
POES (promedio) [MMSTB]	100.00	101.95
POES (mediana) [MMSTB]	100.00	101.60
POES (máximo) [MMSTB]	100.03	108.08
POES (mínimo) [MMSTB]	99.97	98.20
POES (desviación estándar) [MMSTB]	0.01	2.19
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	0.01	7.00

Cuadro 8.13: Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$

Para un $\sigma = 5e-05$ se obtuvo una valor de desviación estándar para el método (F-We vs. Et) de 0.02 MMSTB, 0.03 MMSTB para el método de la capa de gas, y de 0.01 MMSTB para el método (F/Et vs. We/Et). Con un $\sigma = 0,00615$ los valores obtenidos de desviaciones estándar fueron: 1.99 MMSTB para el método (F-We vs. Et), 36.67 MMSTB para el método de la capa de gas, y 2.19 MMSTB para el método (F/Et vs. We/Et). En cuanto a los valores del RMSE para todos los métodos se obtuvieron resultados bastante aceptables. El peor ajuste se obtuvo para el método (F/Et vs. We/Et) con $\sigma = 0,00615$. El gráfico 8.21 muestra que el error promedio para el método (F-We vs. Et) se encuentra alrededor de 3%, para un $\sigma = 0,00615$. Para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) el error se encuentra alrededor de 50%, mientras que para el último método el error alcanzó un valor de 3.5%. El gráfico 8.22 muestra que el error versus el número de realizaciones para todos los casos estuvieron bastante estables. El método (F-We vs. Et) obtuvo un error promedio de 2.5%, (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) con un error promedio de 45%, y el método (F/Et vs. We/Et) con un error promedio alrededor de 3-3.5%.

Error vs. Standard Deviation, Varying: Gas specific gravity, L = 100 realizations

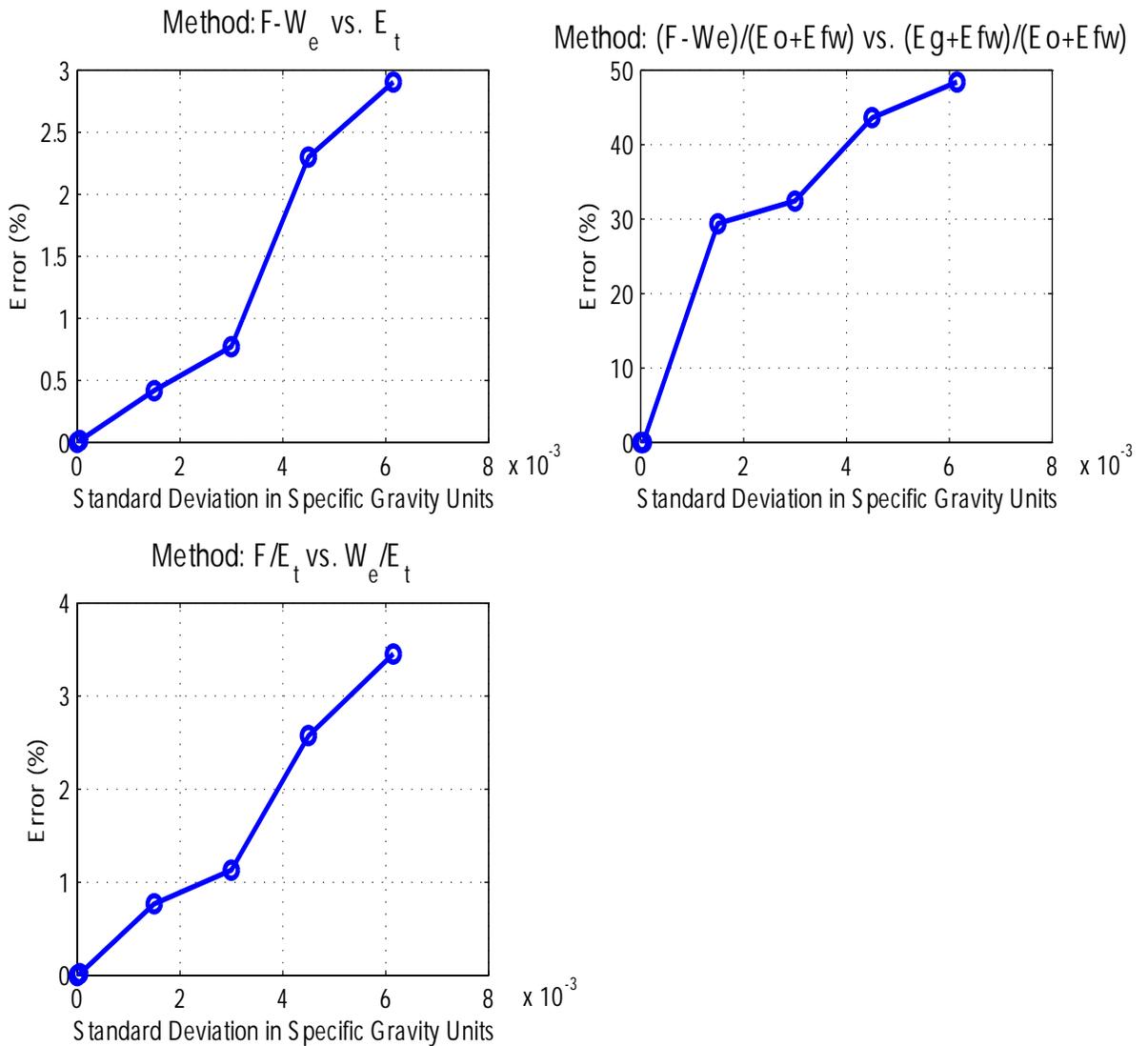


Figura 8.21: Gráfico del error en el POES vs. la desviación estándar de la gravedad específica del gas para el caso 11

Error vs. N° Realizations, Varying: Gas specific gravity, $\sigma = 0.0062$ Specific Gravity Units

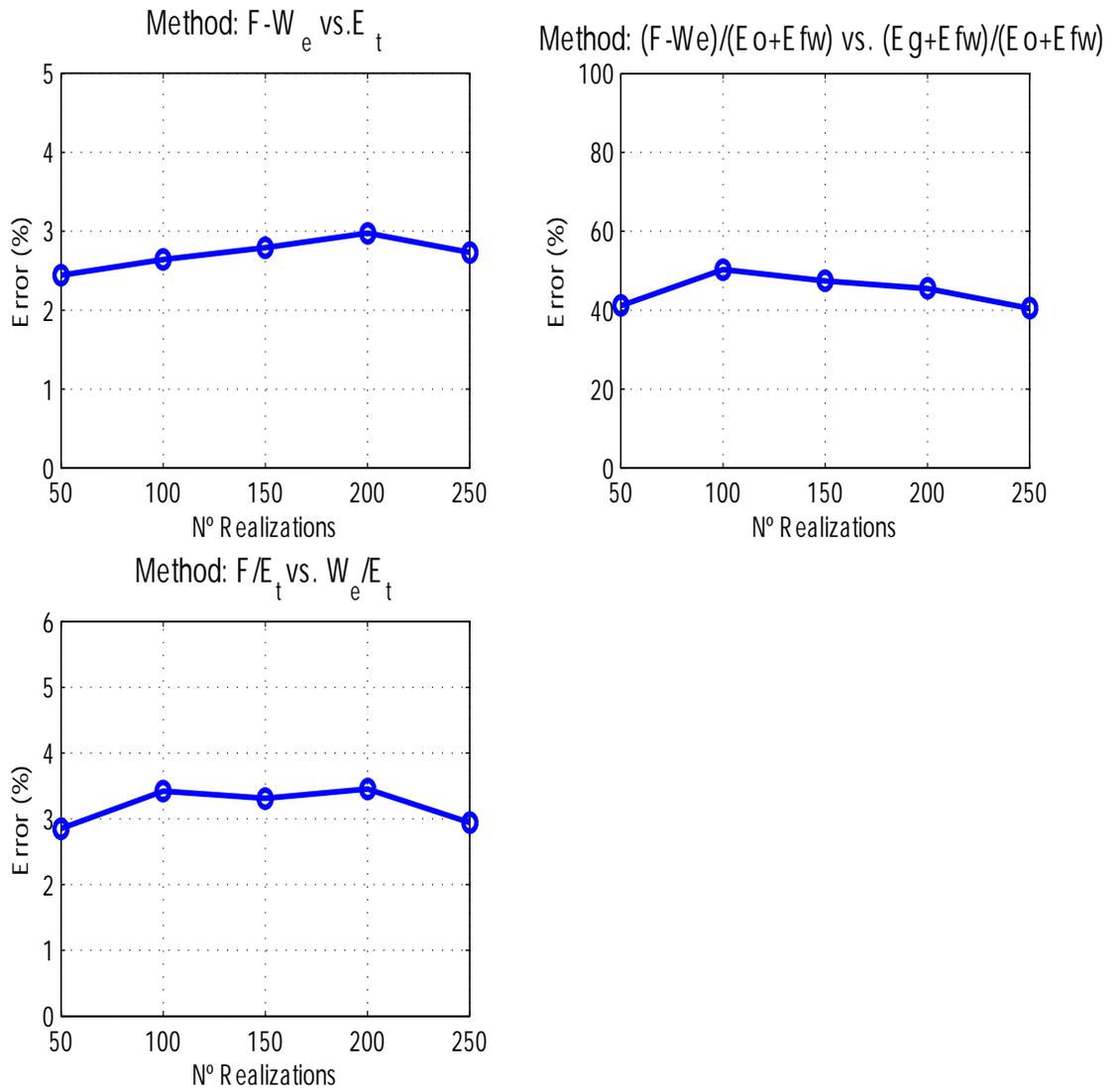


Figura 8.22: Gráfico del error en el POES vs. el número de realizaciones para el caso 11

Caso 15

Método: (F-We vs. Et)	$\sigma = 0,00005$	$\sigma = 0,00615$
POES (promedio) [MMSTB]	100.00	1252.27
POES (mediana) [MMSTB]	100.00	1182.34
POES (máximo) [MMSTB]	100.03	2024.80
POES (mínimo) [MMSTB]	99.96	1097.21
POES (desviación estándar) [MMSTB]	0.02	159.05
RMSE (mínimo) [MMSTB]	0.00	0.79
RMSE (máximo) [MMSTB]	0.01	1.17
Método: (F/Et vs. We/Et)	$\sigma = 0,00005$	$\sigma = 0,00615$
POES (promedio) [MMSTB]	100.00	1073.59
POES (mediana) [MMSTB]	100.00	1018.42
POES (máximo) [MMSTB]	100.02	1695.03
POES (mínimo) [MMSTB]	99.96	952.75
POES (desviación estándar) [MMSTB]	0.02	125.83
RMSE mínimo [MMSTB]	0.00	475.96
RMSE máximo [MMSTB]	0.01	653.38

Cuadro 8.14: Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento,

$$r_e D = \infty$$

Con un $\sigma = 5e - 05$ para la gravedad específica del gas se obtuvo los siguientes valores de desviación estándar para los POES: 0.02 MMSTB para el método (F-We vs. Et), y 0.02 MMSTB para el método (F/Et vs. We/Et). Para un $\sigma = 0,00615$ los valores de desviación estándar fueron: 159.05 MMSTB para el método (F-We vs. Et) y de 125.83 para el método (F/Et vs. We/Et). Para el método (F-We vs. Et) se obtuvieron ajustes excelentes para ambas desviaciones. De la misma forma los valores del RMSE reportados para el método (F/Et vs. We/Et) con una desviación de $\sigma = 5e - 05$ indican que los ajustes fueron excelentes; más no fue así para una desviación de $\sigma = 0,00615$ donde claramente se observa que los ajustes realizados fueron bastante pobres. En el gráfico 8.23 se muestra el error promedio versus la desviación estándar para la gravedad específica del gas. El método (F-We vs. Et) obtuvo un error promedio de 8%, mientras que el método (F/Et vs. We/Et) obtuvo un error de 9.5-10%. El gráfico 8.24 muestra el error promedio en los POES versus el número de realizaciones. El método (F-We vs. Et) obtuvo un error promedio de 9% para los valores de realizaciones tomados. Para el método (F/Et vs. We/Et) el error promedio fue de 10%.

Error vs. Standard Deviation,
 Varying: Gas specific gravity, L = 100 realizations

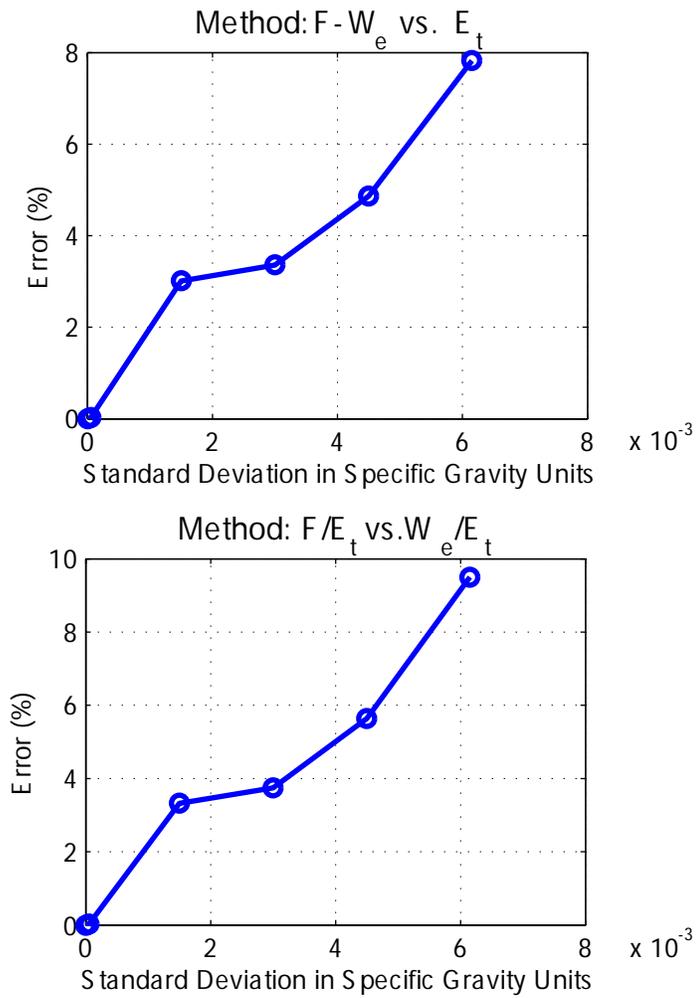


Figura 8.23: Gráfico del error en el POES vs. la desviación estándar de la gravedad específica del gas para el caso 15

Error vs. N° Realizations,
Varying: Gas specific gravity,
 $\sigma = 0.0062$ Specific Gravity Units

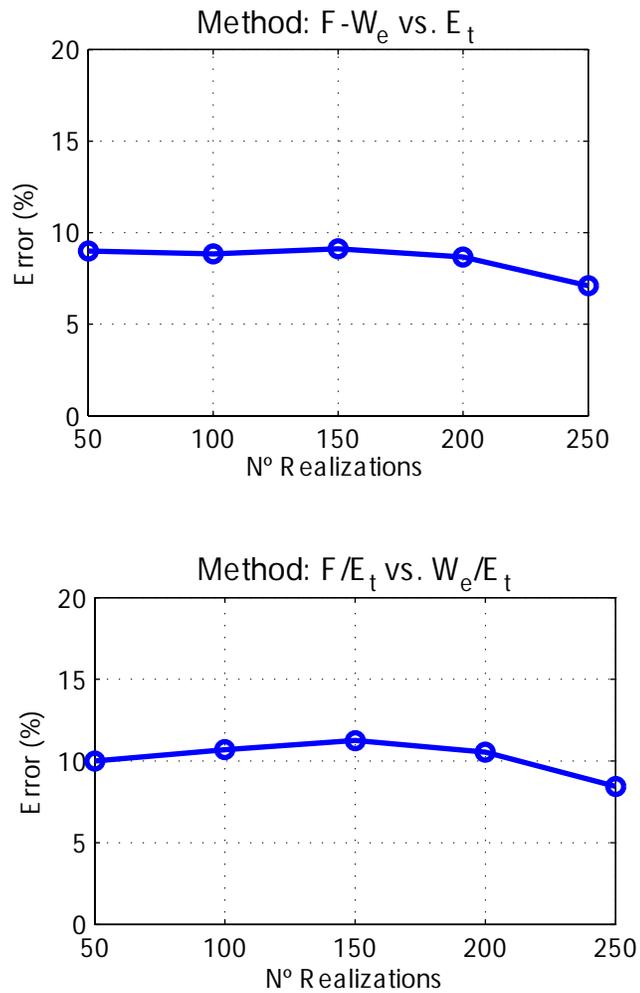


Figura 8.24: Gráfico del error en el POES vs. el número de realizaciones para el caso 15

Análisis final para la gravedad específica del gas

Método de balance de materiales	Método 1	Método 2	Método 3
Promedio de la desviación estándar de los POES para un $\sigma = 5e - 5$	0.02	0.13	0.02
Promedio de la desviación estándar de los POES para un $\sigma = 0,0062$	54.77	85.84	64.01
Método 1: F-We vs. Et			
Método 2: $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$			
Método 3: F/Et vs. We/Et			

Cuadro 8.15: Desviación estándar de los POES para la realizaciones de la gravedad específica del gas

El método más robusto fue (F-We vs. Et), para todos los casos. El método (F/Et vs. We/Et) muestra una dispersión comparable al método anterior para todos los casos. El método $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$ sigue siendo el que posee mayor dispersión entre todos los métodos para los casos evaluados. En general, los valores del RMSE obtenidos para todos los métodos y ambas desviaciones muestran ajustes bastante aceptables, con la excepción para el caso 15 con una desviación de $\sigma = 0,0062$ donde claramente el método (F/Et vs. We/Et) obtuvo ajustes bastante pobres en los datos.

8.1.4. Relación gas petróleo en solución

Caso 1

Método: (F-We vs. Et)	$\sigma = 0,738$ [SCF/STB]	$\sigma = 14,76$ [SCF/STB]
POES (promedio) [MMSTB]	100.04	100.67
POES (mediana) [MMSTB]	100.00	99.99
POES (máximo) [MMSTB]	100.28	105.90
POES (mínimo) [MMSTB]	99.99	99.96
POES (desviación estándar) [MMSTB]	0.06	1.08
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.15

Cuadro 8.16: Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero

El valor de desviación estándar obtenido fue de 0.06 MMSTB, para un $\sigma = 0,738$ SCF/STB. Para un $\sigma = 14,76$ SCF/STB la desviación estándar fue de 1.08 MMSTB. Para este caso se obtuvieron valores del RMSE que indican ajustes excelentes en los datos. El gráfico de error en el POES 8.25, indica que el error esperado para el POES es de 1.25 %. El gráfico 8.26 muestra que el valor del error es estable, entre 1.25 y 1.50 % para todos los valores de realizaciones utilizados (para un $\sigma = 14,76$ SCF/STB).

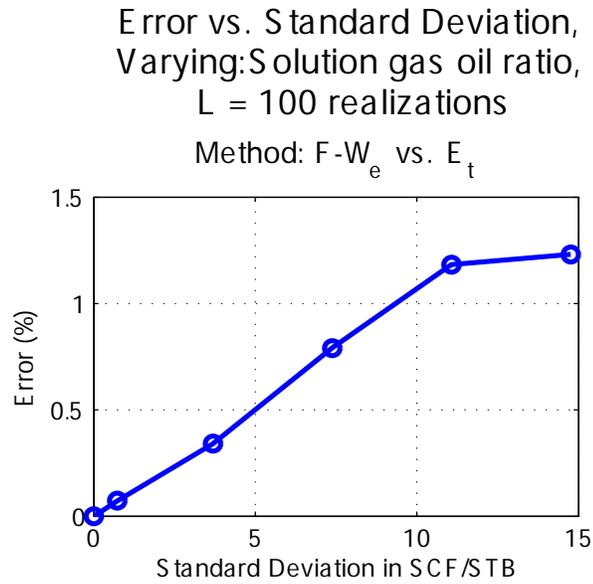


Figura 8.25: Gráfico del error en el POES vs. la desviación estándar de la relación gas petróleo en solución para el caso 1

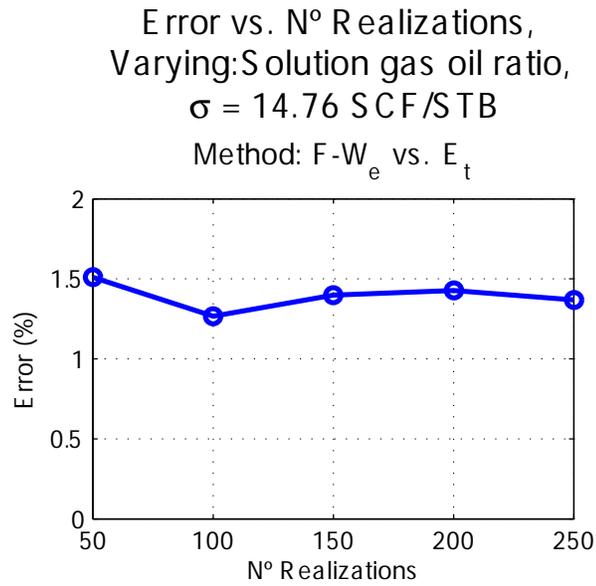


Figura 8.26: Gráfico del error en el POES vs. el número de realizaciones para el caso 1

Caso 5

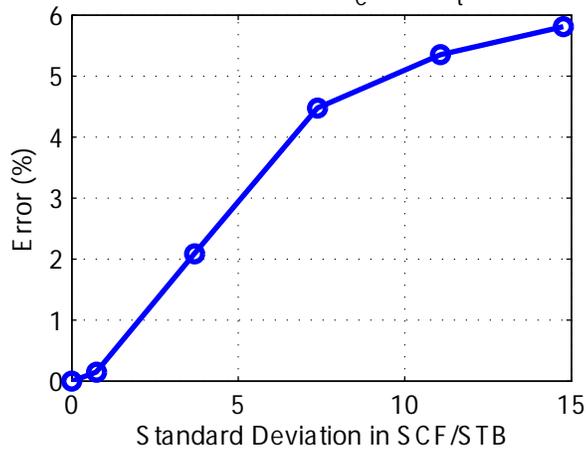
Método: (F-We vs. Et)	$\sigma = 0,738$ [SCF/STB]	$\sigma = 14,76$ [SCF/STB]
POES (promedio) [MMSTB]	100.11	103.94
POES (mediana) [MMSTB]	100.09	103.00
POES (máximo) [MMSTB]	100.38	109.38
POES (mínimo) [MMSTB]	99.95	100.05
POES (desviación estándar) [MMSTB]	0.11	3.28
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.14
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	$\sigma = 0,738$ [SCF/STB]	$\sigma = 14,76$ [SCF/STB]
POES (promedio) [MMSTB]	103.49	134.76
POES (mediana) [MMSTB]	99.03	132.13
POES (máximo) [MMSTB]	138.14	230.47
POES (mínimo) [MMSTB]	45.21	-87.27
POES (desviación estándar) [MMSTB]	15.40	72.79
RMSE mínimo [MMSTB]	0.00	0.03
RMSE máximo [MMSTB]	1.74	6.11

Cuadro 8.17: Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero

Las desviaciones estándar para los POES calculados fueron: 0.11 MMSTB para el método de (F-We vs. Et) y 15.40 MMSTB para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw). Para un $\sigma = 14,76$ SCF/STB la desviación estándar fue 3.28 MMSTB para el método (F-We vs. Et), y de 72.79 MMSTB para el método de la capa de gas. El método (F-We vs. Et) muestra de los valores del RMSE que se realizaron ajustes excelentes en los datos para ambas desviaciones, mientras que el método de capa de gas no muestra un ajuste tan bueno como el método anterior todavía muestra ajustes aceptables en los datos. El gráfico 8.27 señala que para el método (F-We vs. Et), con un $\sigma = 14,76$ SCF/STB, se obtendrá un error promedio en el POES alrededor de 6%, mientras que para el otro método se tendría un error de 90% con la misma desviación en la relación gas petróleo en solución ($\sigma = 14,76$ SCF/STB). El gráfico 8.28 indica que para el primer método (F-We vs. Et) el error permanece estable en aproximadamente 5%, para todos los valores de realizaciones tomados. El método de la capa de gas mostró bastante estabilidad en el error para los valores de realizaciones tomados, el promedio de los errores obtenidos estuvo alrededor de 80%.

Error vs. Standard Deviation,
 Varying: Solution gas oil ratio,
 L = 100 realizations

Method: $F - W_e$ vs. E_t



Method: $(F - W_e)/(E_o + E_{fw})$ vs. $(E_g + E_{fw})/(E_o + E_{fw})$

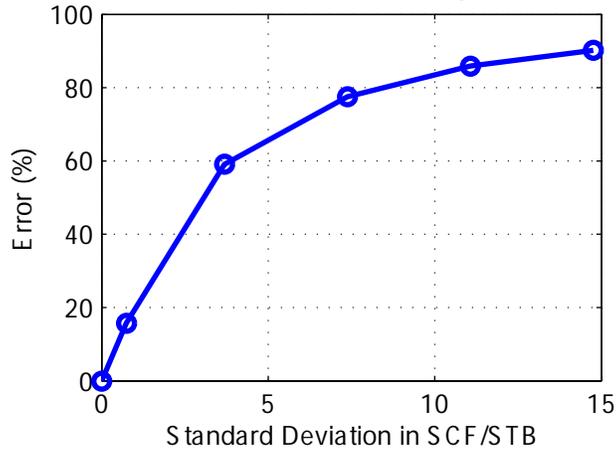


Figura 8.27: Gráfico del error en el POES vs. la desviación estándar de la relación gas petróleo en solución para el caso 5

Error vs. N° Realizations,
 Varying: Solution gas oil ratio,
 $\sigma = 14.76$ SCF/STB
 Method: $F - W_e$ vs. E_t

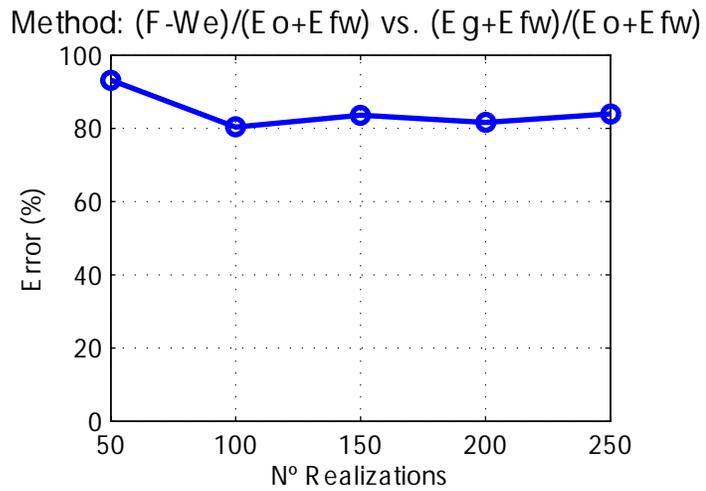
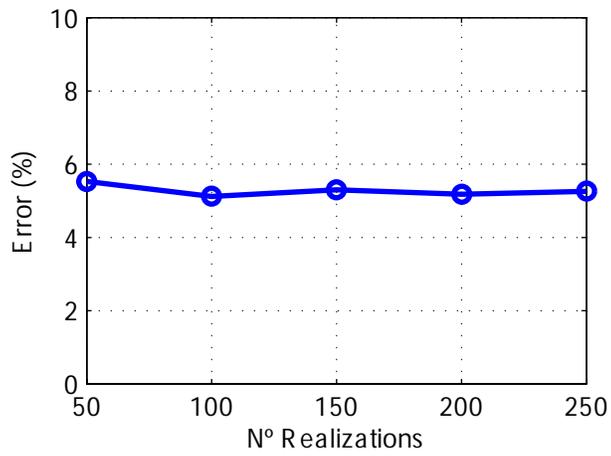


Figura 8.28: Gráfico del error en el POES vs. el número de realizaciones para el caso 5

Caso 11

Método: (F-We vs. Et)	$\sigma = 0,738$ [SCF/STB]	$\sigma = 14,76$ [SCF/STB]
POES (promedio) [MMSTB]	100.01	104.07
POES (mediana) [MMSTB]	100.02	101.51
POES (máximo) [MMSTB]	100.52	121.92
POES (mínimo) [MMSTB]	99.64	99.51
POES (desviación estándar) [MMSTB]	0.17	5.04
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.09
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	$\sigma = 0,738$ [SCF/STB]	$\sigma = 14,76$ [SCF/STB]
POES (promedio) [MMSTB]	103.00	110.45
POES (mediana) [MMSTB]	100.22	105.49
POES (máximo) [MMSTB]	163.11	194.20
POES (mínimo) [MMSTB]	32.38	35.96
POES (desviación estándar) [MMSTB]	31.23	28.04
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	2.55	2.71
Método: (F/Et vs. We/Et)	$\sigma = 0,738$ [SCF/STB]	$\sigma = 14,76$ [SCF/STB]
POES (promedio) [MMSTB]	99.73	104.67
POES (mediana) [MMSTB]	100.01	101.40
POES (máximo) [MMSTB]	100.25	139.71
POES (mínimo) [MMSTB]	98.44	96.97
POES (desviación estándar) [MMSTB]	0.57	7.96
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	1.96	24.08

Cuadro 8.18: Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$

Para un $\sigma = 0,738$ SCF/STB se obtuvo una valor de desviación estándar para el método (F-We vs. Et) de 0.17 MMSTB, 31.23 MMSTB para el método de la capa de gas, y de 0.57 MMSTB para el método (F/Et vs. We/Et). Con un $\sigma = 14,76$ SCF/STB los valores obtenidos de desviaciones estándar fueron: 5.04 MMSTB para el método (F-We vs. Et), 28.04 MMSTB para el método de la capa de gas, y 7.96 MMSTB para el método (F/Et vs. We/Et). De nuevo, para este caso los métodos de balance de materiales muestran ajustes bastantes aceptables para ambas desviaciones, sin embargo cuando se utilizó el método (F/Et vs. We/Et) con una desviación de $\sigma = 14,76$ SCF/STB se obtuvo el mayor valor del RMSE para todas las corridas hechas (para este caso). El gráfico 8.29 muestra que el error promedio para el método (F-We vs. Et) es menos que 7%, para un $\sigma = 14,76$ SCF/STB. Para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) el error estuvo alrededor de 30%, mientras que para el último método el error alcanzó un valor menos de 10% (para un $\sigma = 14,76$ SCF/STB). Del gráfico 8.30 se puede observar que todos los casos tuvieron errores bastante estables. El método (F-We vs. Et) obtuvo un error promedio entre 6 y 7%, (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) con un error promedio de 30%, y (F/Et vs. We/Et) con un error promedio alrededor de 7 a 10%.

Error vs. Standard Deviation, Varying: Solution gas oil ratio, L = 100 realizations

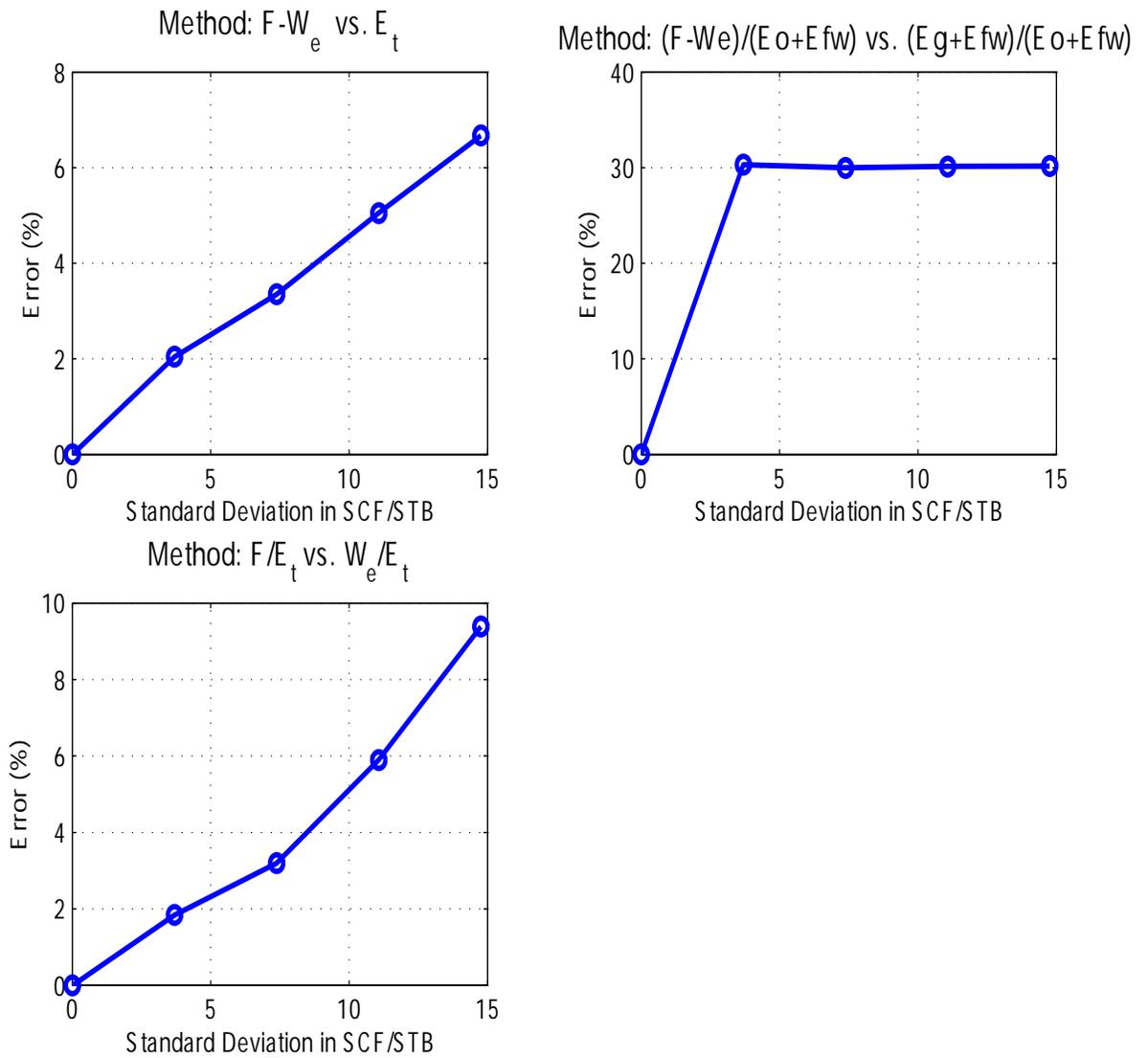


Figura 8.29: Gráfico del error en el POES vs. la desviación estándar de la relación gas petróleo en solución para el caso 11

Error vs. N°Realizations, Varying: Solution gas oil ratio, $\sigma = 14.76$ SCF/STB

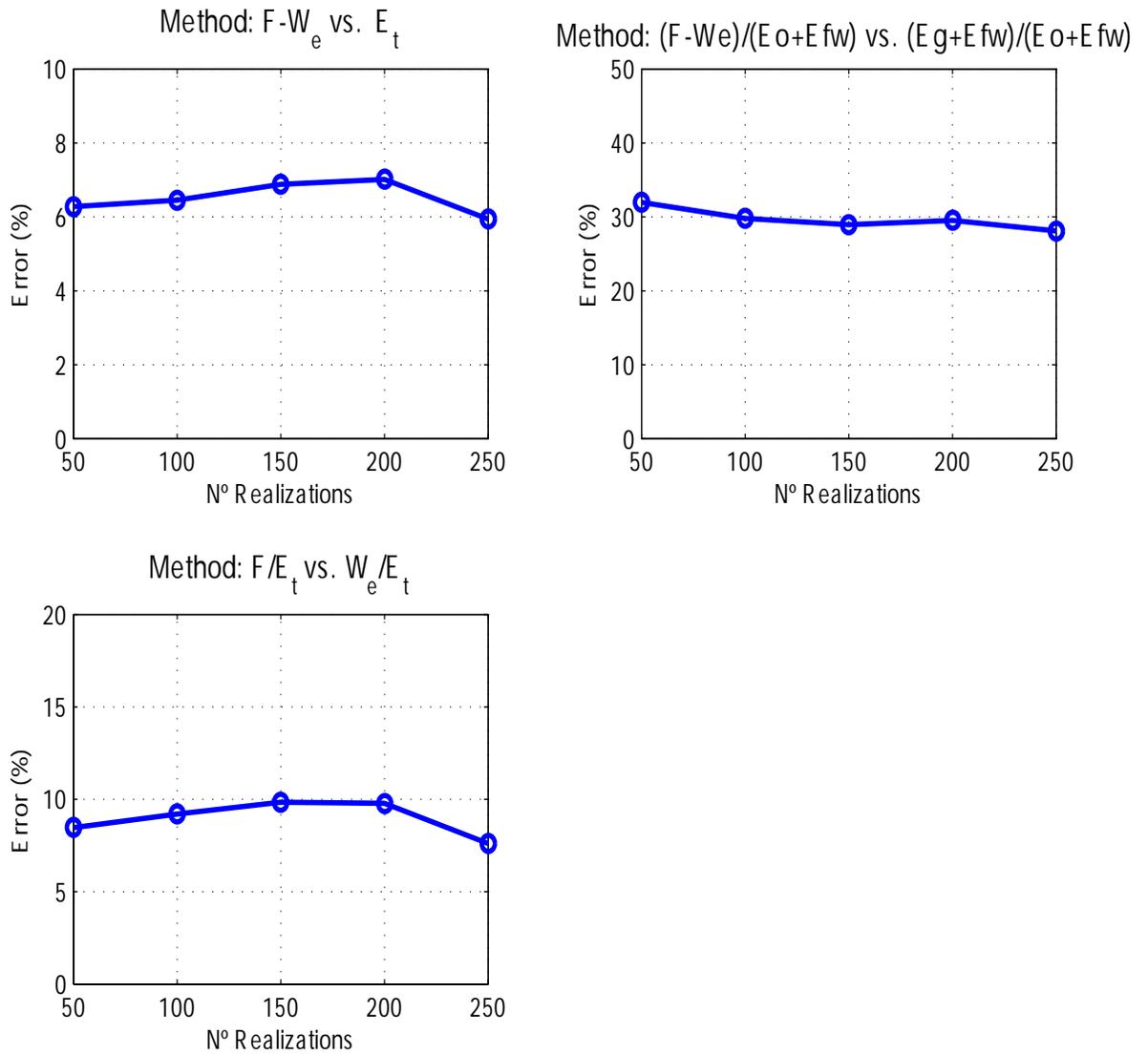


Figura 8.30: Gráfico del error en el POES vs. el número de realizaciones para el caso 11

Caso 15

Método: (F-We vs. Et)	$\sigma = 0,738$ [SCF/STB]	$\sigma = 14,76$ [SCF/STB]
POES (promedio) [MMSTB]	98.03	114.64
POES (mediana) [MMSTB]	99.96	99.77
POES (máximo) [MMSTB]	100.01	265.69
POES (mínimo) [MMSTB]	93.84	94.45
POES (desviación estándar) [MMSTB]	2.54	31.41
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.04
Método: (F/Et vs. We/Et)	$\sigma = 0,738$ [SCF/STB]	$\sigma = 14,76$ [SCF/STB]
POES (promedio) [MMSTB]	97.81	117.23
POES (mediana) [MMSTB]	99.96	99.74
POES (máximo) [MMSTB]	100.01	265.23
POES (mínimo) [MMSTB]	93.20	93.86
POES (desviación estándar) [MMSTB]	2.82	36.28
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	1.85	35.90

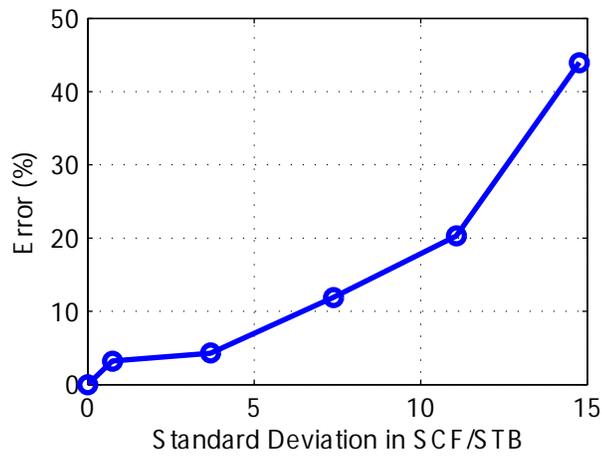
Cuadro 8.19: Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento,

$$r_e D = \infty$$

Con un $\sigma = 0,738$ SCF/STB para la relación gas petróleo en solución, se obtuvo los siguientes valores de desviación estándar para los POES: 2.54 MMSTB para el método (F-We vs. Et), y 2.82 MMSTB para el método (F/Et vs. We/Et). Para un $\sigma = 14,76$ SCF/STB los valores de desviación estándar fueron: 31.41 MMSTB para el método (F-We vs. Et) y de 36.28 para el método (F/Et vs. We/Et). Como para los casos anteriores estos el método (F-We vs. Et) muestra ajustes excelentes para ambas desviaciones, al igual que el método (F/Et vs. We/Et) para la desviación de $\sigma = 0,738$ SCF/STB. Para una desviación de $\sigma = 14,76$ SCF/STB este método arrojó el RMSE más grande para los dos métodos. En el gráfico 8.31 se muestra el error promedio versus la desviación estándar en para la relación gas petróleo en solución. El método (F-We vs. Et) obtuvo un error promedio de 40-45 %, mientras que el método (F/Et vs. We/Et) obtuvo un error de 45-50 %. En el gráfico 8.32 se puede observar que el método (F-We vs. Et) obtuvo un error promedio de 30-40 % para los valores de realizaciones tomados. Para el método (F/Et vs. We/Et) el error promedio varió entre 35 y 45 %.

Error vs. Standard Deviation,
 Varying: Solution gas oil ratio,
 L = 100 realizations

Method: $F - W_e$ vs. E_t



Method: F/E_t vs. W_e/E_t

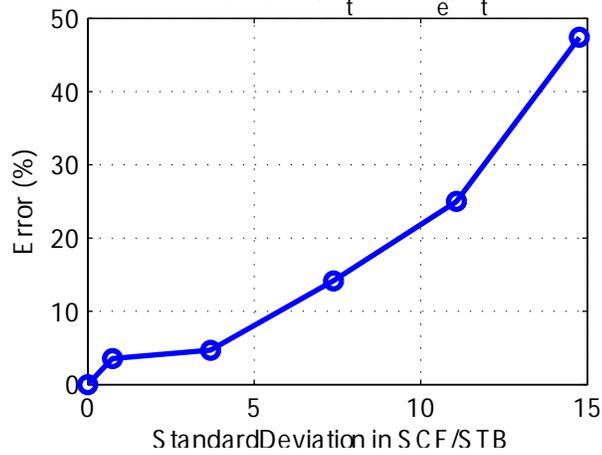


Figura 8.31: Gráfico del error en el POES vs. la desviación estándar de la relación gas petróleo en solución para el caso 15

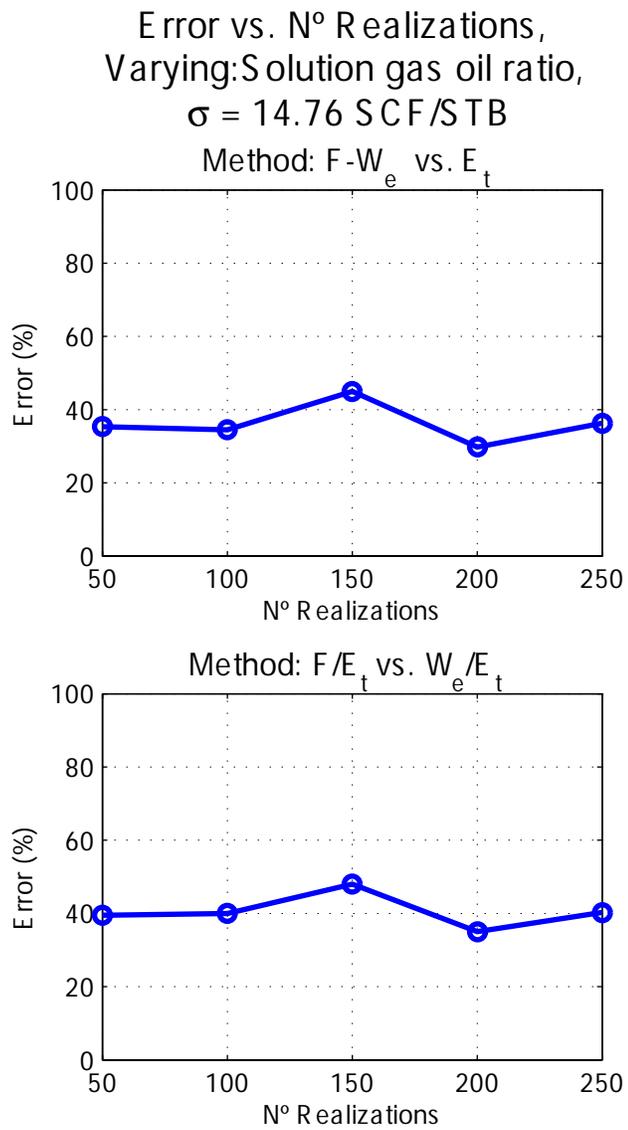


Figura 8.32: Gráfico del error en el POES vs. el número de realizaciones para el caso 15

Análisis final para la relación gas petróleo en solución

Método de balance de materiales	Método 1	Método 2	Método 3
Promedio de la desviación estándar de los POES para un $\sigma = 0,738$ SCF/STB	0.72	9.20	1.70
Promedio de la desviación estándar de los POES para un $\sigma = 14,76$ SCF/STB	20.41	50.42	22.12
Método 1: F-We vs. Et			
Método 2: $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$			
Método 3: F/Et vs. We/Et			

Cuadro 8.20: Desviación estándar de los POES para la realizaciones de la relación gas petróleo en solución

El método más robusto fue (F-We vs. Et), para todos los casos. También para este parámetro, el método (F/Et vs. We/Et) muestra una dispersión comparable al método anterior para todos los casos. El método $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$ sigue siendo el que posee mayor dispersión entre todos los métodos para los casos evaluados (un poco más de 2 veces el error que en los otros métodos para la desviación más alta). En general, todos los métodos de balance materiales muestran valores de RMSE bastante aceptables, específicamente, el mayor error se obtuvo para el caso 15 con una desviación de $\sigma = 14,76$ SCF/STB al utilizar el método (F/Et vs. We/Et).

8.1.5. Temperatura del yacimiento

Caso 1

Método: (F-We vs. Et)	$\sigma = 0,02$ °F	$\sigma = 0,45$ °F
POES (promedio) [MMSTB]	100.00	100.02
POES (mediana) [MMSTB]	100.00	100.04
POES (máximo) [MMSTB]	100.01	100.13
POES (mínimo) [MMSTB]	99.99	99.32
POES (desviación estándar) [MMSTB]	0.01	0.14
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.04

Cuadro 8.21: Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero

Como se ha mencionado anteriormente el único método utilizado fue (F-We vs. Et). Los cálculos de balance de materiales mostraron poca dispersión para $\sigma = 0,02$ °F, alcanzando un valor de desviación estándar de 0.01 MMSTB. Para un $\sigma = 0,45$ °F se observa un aumento en la desviación estándar hasta un valor de 0.14 MMSTB para la desviación estándar. Al igual que para los otros parámetros este método produce ajustes excelentes para ambas desviaciones. El gráfico de error en el POES 8.33, indica que el error esperado para el POES con una desviación estándar de 0.45 °F para la temperatura del yacimiento es alrededor de 0.15 %. A partir del gráfico 8.34 el valor del error se muestra estable para todos los valores de realizaciones utilizados, alcanzando aproximadamente hasta 0.2% (para un $\sigma = 0,45$ °F).

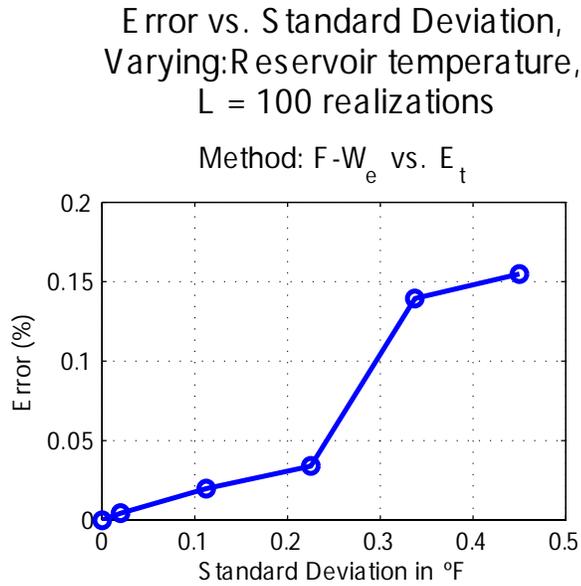


Figura 8.33: Gráfico del error en el POES vs. la desviación estándar de la temperatura del yacimiento para el caso 1

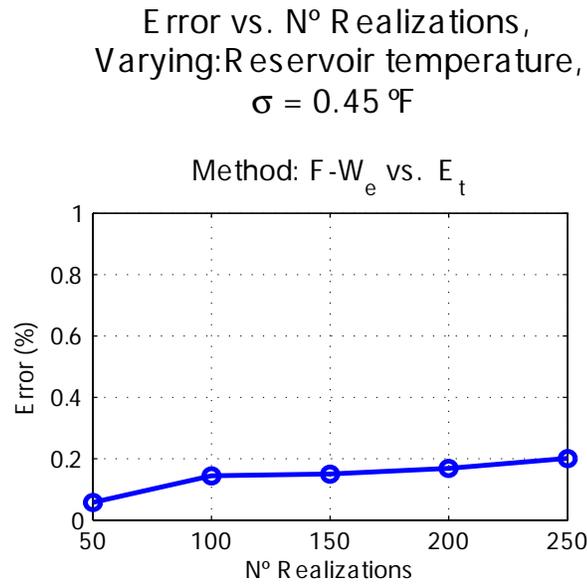


Figura 8.34: Gráfico del error en el POES vs. el número de realizaciones para el caso 1

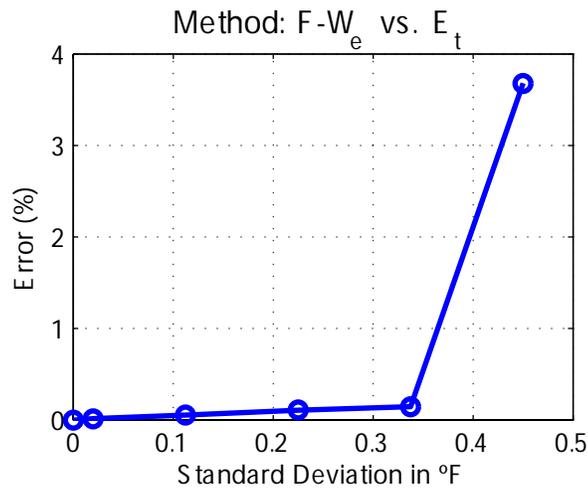
Caso 5

Método: (F-We vs. Et)	$\sigma = 0,02 \text{ } ^\circ\text{F}$	$\sigma = 0,45 \text{ } ^\circ\text{F}$
POES (promedio) [MMSTB]	100.00	100.15
POES (mediana) [MMSTB]	100.00	100.10
POES (máximo) [MMSTB]	100.03	100.75
POES (mínimo) [MMSTB]	99.97	99.94
POES (desviación estándar) [MMSTB]	0.01	0.18
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.17
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	$\sigma = 0,02 \text{ } ^\circ\text{F}$	$\sigma = 0,45 \text{ } ^\circ\text{F}$
POES (promedio) [MMSTB]	99.92	114.58
POES (mediana) [MMSTB]	99.97	99.26
POES (máximo) [MMSTB]	100.37	241.84
POES (mínimo) [MMSTB]	99.41	42.22
POES (desviación estándar) [MMSTB]	0.19	46.80
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	0.02	4.31

Cuadro 8.22: Caso 5. Tamaño de la capa de gas, $m = 1$. No tiene acuífero

Para un $\sigma = 0,02 \text{ } ^\circ\text{F}$ se obtuvo una desviación estándar de 0.01 MMSTB para el método de (F-We vs. Et) y una desviación estándar de 0.19 MMSTB para el método (F-We vs. Et) y (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw). Para un $\sigma = 0,45 \text{ } ^\circ\text{F}$ la desviación estándar fue 0.18 MMSTB para el método (F-We vs. Et), y de 46.80 MMSTB para el método de la capa de gas. Los métodos de balance de materiales producen resultados bastante aceptables, excepto para cuando existe un $\sigma = 0,45 \text{ } ^\circ\text{F}$ y se utiliza el método de la capa de gas. El gráfico 8.35 indica que para el método (F-We vs. Et), con un $\sigma = 0,45 \text{ } ^\circ\text{F}$, se obtendrá un error promedio en el POES de un poco más de 3.5%; mientras que para el otro método se tendría un error de más de 45% con la misma desviación en la temperatura ($\sigma = 0,45 \text{ } ^\circ\text{F}$). El gráfico 8.36 indica que para el primer método (F-We vs. Et) el error permanece estable (tomando en cuenta la magnitud del error), variando a 1-2% para todos los valores de realizaciones tomados. El método de la capa de gas muestra que el error se encuentra entre 50-60% para los valores de realizaciones tomados.

Error vs. Standard Deviation,
 Varying: Reservoir temperature,
 L = 100 realizations



Method: $(F - W_e)/(E_o + E_{fw})$ vs. $(E_g + E_{fw})/(E_o + E_{fw})$

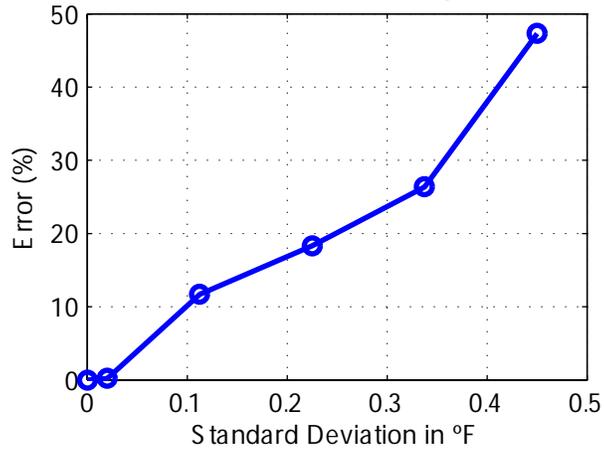
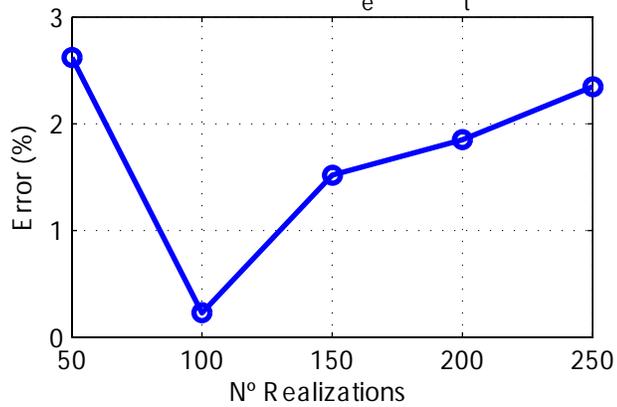


Figura 8.35: Gráfico del error en el POES vs. la desviación estándar de la temperatura del yacimiento para el caso 5

Error vs. N° Realizations,
 Varying: Reservoir temperature,
 $\sigma = 0.45 \text{ } ^\circ\text{F}$
 Method: $F - W_e$ vs. E_t



Method: $(F - W_e)/(E_o + E_{fw})$ vs. $(E_g + E_{fw})/(E_o + E_{fw})$

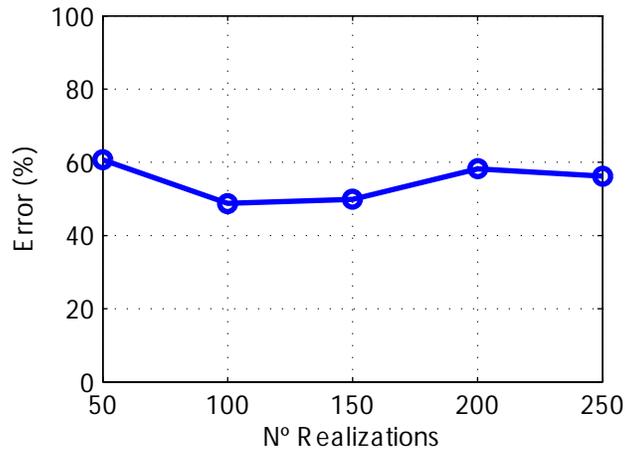


Figura 8.36: Gráfico del error en el POES vs. el número de realizaciones para el caso 5

Caso 11

Método: (F-We vs. Et)	$\sigma = 0,02 \text{ } ^\circ\text{F}$	$\sigma = 0,45 \text{ } ^\circ\text{F}$
POES (promedio) [MMSTB]	100.00	100.29
POES (mediana) [MMSTB]	100.00	100.14
POES (máximo) [MMSTB]	100.03	109.30
POES (mínimo) [MMSTB]	99.97	99.62
POES (desviación estándar) [MMSTB]	0.01	1.31
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.01
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	$\sigma = 0,02 \text{ } ^\circ\text{F}$	$\sigma = 0,45 \text{ } ^\circ\text{F}$
POES (promedio) [MMSTB]	100.00	97.86
POES (mediana) [MMSTB]	100.00	100.43
POES (máximo) [MMSTB]	100.08	161.57
POES (mínimo) [MMSTB]	99.91	35.74
POES (desviación estándar) [MMSTB]	0.03	31.21
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	0.01	2.51
Método: (F/Et vs. We/Et)	$\sigma = 0,02 \text{ } ^\circ\text{F}$	$\sigma = 0,45 \text{ } ^\circ\text{F}$
POES (promedio) [MMSTB]	100.00	100.02
POES (mediana) [MMSTB]	100.00	100.07
POES (máximo) [MMSTB]	100.03	108.95
POES (mínimo) [MMSTB]	99.97	98.44
POES (desviación estándar) [MMSTB]	0.01	1.42
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	0.01	1.92

Cuadro 8.23: Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$

Para un $\sigma = 0,02$ °F se obtuvo un valor de desviación estándar para el método (F-We vs. Et) de 0.01 MMSTB, 0.03 MMSTB para el método de la capa de gas, y de 0.01 MMSTB para el método (F/Et vs. We/Et). Con un $\sigma = 0,45$ °F los valores obtenidos de desviaciones estándar fueron: 1.31 MMSTB para el método (F-We vs. Et), 31.21 MMSTB para el método de la capa de gas, y 1.42 MMSTB para el método (F/Et vs. We/Et). También para este caso los métodos de balance de materiales producen resultados bastante buenos, sin embargo para cuando hay un $\sigma = 0,45$ °F y se utilizan el método de la capa de gas y el método (F/Et vs. We/Et) se obtienen valores del RMSE mucho mayores que para la otra desviación. El gráfico 8.37 muestra que el error promedio para el método (F-We vs. Et) es menos que 1.3%, para un $\sigma = 0,45$ °F. Para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) el error está alrededor de 30%, mientras que para el último método el error alcanzó un valor de 1.3-1.4% (para un $\sigma = 0,45$ °F). El gráfico 8.38 muestra el error vs. el número de realizaciones para este caso. Se puede observar que todos los casos tuvieron errores bastante estables (tomando en cuenta la magnitud de los errores). El método (F-We vs. Et) obtuvo un error promedio de 1.0%, (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) con un error promedio alrededor de 30%, y (F/Et vs. We/Et) con un error promedio alrededor de 1-1.5%.

Error vs. S Standard Deviation, Varying: Reservoir temperature, L = 100 realizations

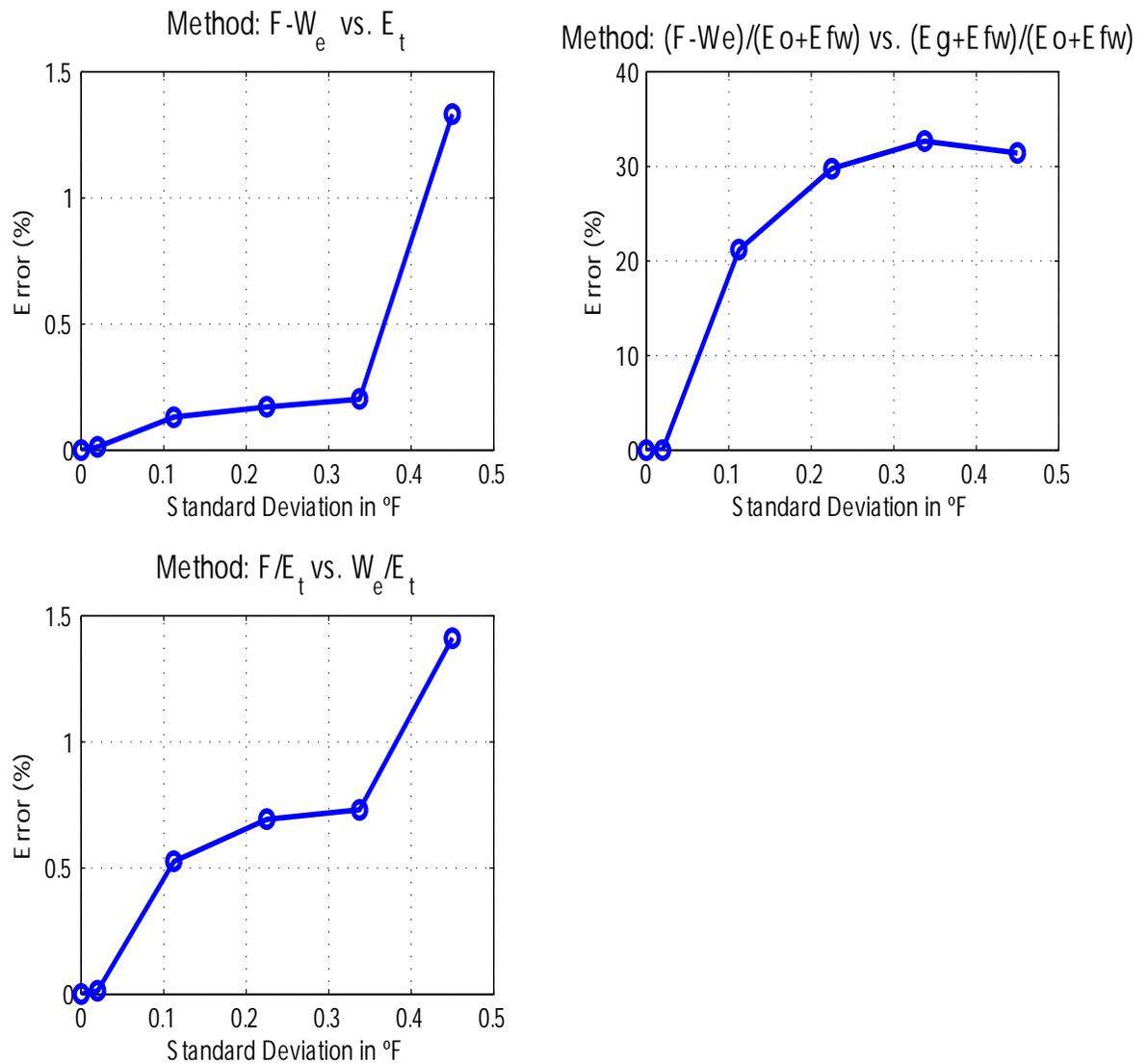


Figura 8.37: Gráfico del error en el POES vs. la desviación estándar de la temperatura del yacimiento para el caso 11

Error vs. N° Realizations, Varying: Reservoir temperature, $\sigma = 0.45^\circ\text{F}$

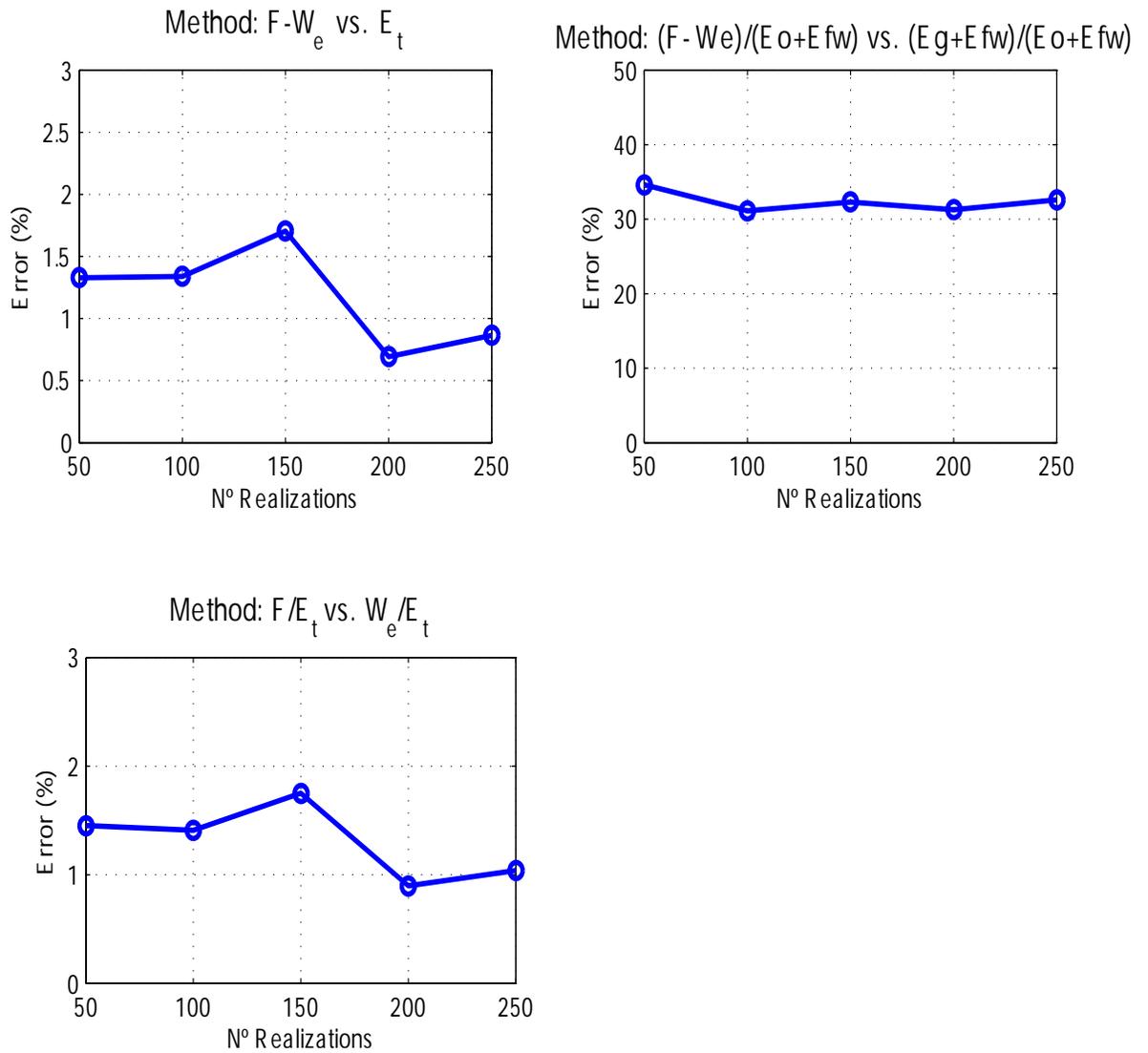


Figura 8.38: Gráfico del error en el POES vs. el número de realizaciones para el caso 11

Caso 15

Método: (F-We vs. Et)	$\sigma = 0,02 \text{ } ^\circ\text{F}$	$\sigma = 0,45 \text{ } ^\circ\text{F}$
POES (promedio) [MMSTB]	99.99	98.48
POES (mediana) [MMSTB]	99.99	100.05
POES (máximo) [MMSTB]	100.04	100.86
POES (mínimo) [MMSTB]	99.95	93.79
POES (desviación estándar) [MMSTB]	0.02	2.50
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.01	0.01
Método: (F/Et vs. We/Et)	$\sigma = 0,02 \text{ } ^\circ\text{F}$	$\sigma = 0,45 \text{ } ^\circ\text{F}$
POES (promedio) [MMSTB]	99.99	98.29
POES (mediana) [MMSTB]	99.99	100.05
POES (máximo) [MMSTB]	100.04	100.76
POES (mínimo) [MMSTB]	99.96	93.17
POES (desviación estándar) [MMSTB]	0.02	2.72
RMSE mínimo [MMSTB]	0.00	0.01
RMSE máximo [MMSTB]	0.01	1.80

Cuadro 8.24: Caso 15. Tamaño de la capa de gas, $m = 0$. Relación radio del acuífero-radio del yacimiento,

$$r_e D = \infty$$

Con un $\sigma = 0,02 \text{ } ^\circ\text{F}$ en la temperatura del yacimiento, se obtuvo los siguientes valores de desviación estándar para los POES: 0.02 MMSTB para el método (F-We vs. Et), y 0.02 MMSTB para el método (F/Et vs. We/Et). Para un $\sigma = 0,45 \text{ } ^\circ\text{F}$ los valores de desviación estándar fueron: 2.50 MMSTB para el método (F-We vs. Et) y de 2.72 para el método (F/Et vs. We/Et). Los dos métodos producen resultados para el RMSE bien admisibles, siendo el mayor valor del RMSE para cuando existe un $\sigma = 0,45 \text{ } ^\circ\text{F}$ y se utiliza el método (F/Et vs. We/Et). En el gráfico 8.39 se muestra el error promedio versus la desviación estándar en la historia de presión. El método (F-We vs. Et) obtuvo un error promedio de 3.5%, mientras que el (F/Et vs. We/Et) obtuvo un error de 3.5-3.8%. El gráfico 8.40 muestra el error promedio en los POES versus el número de realizaciones. El método (F-We vs. Et) obtuvo un error promedio de 3-3.5% para los valores de realizaciones tomados. Para el método (F/Et vs. We/Et) el error promedio estuvo alrededor de 3.5%.

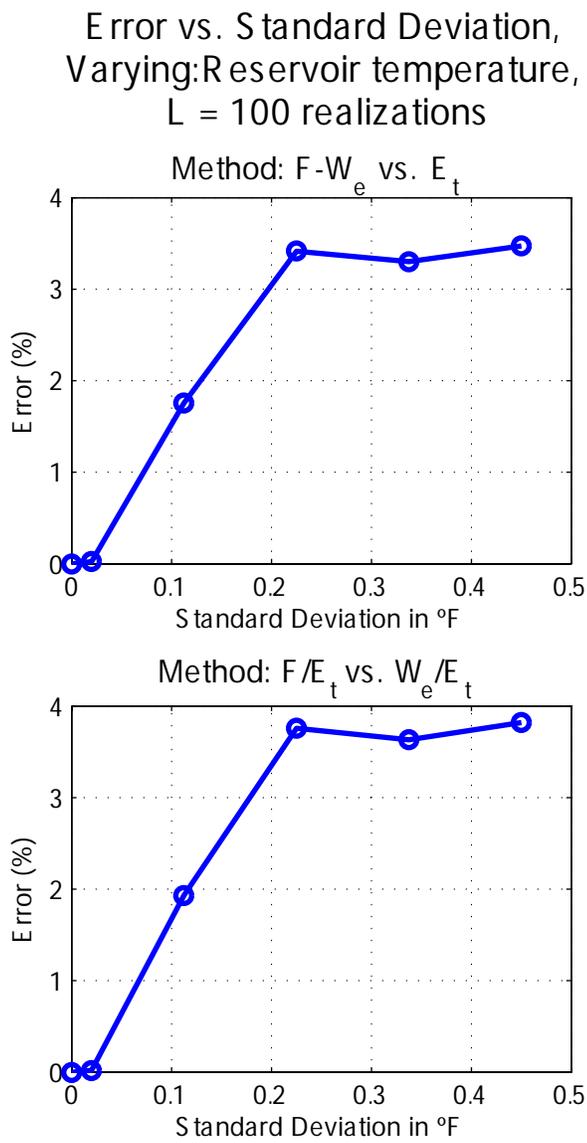


Figura 8.39: Gráfico del error en el POES vs. la desviación estándar de la temperatura del yacimiento para el caso 15

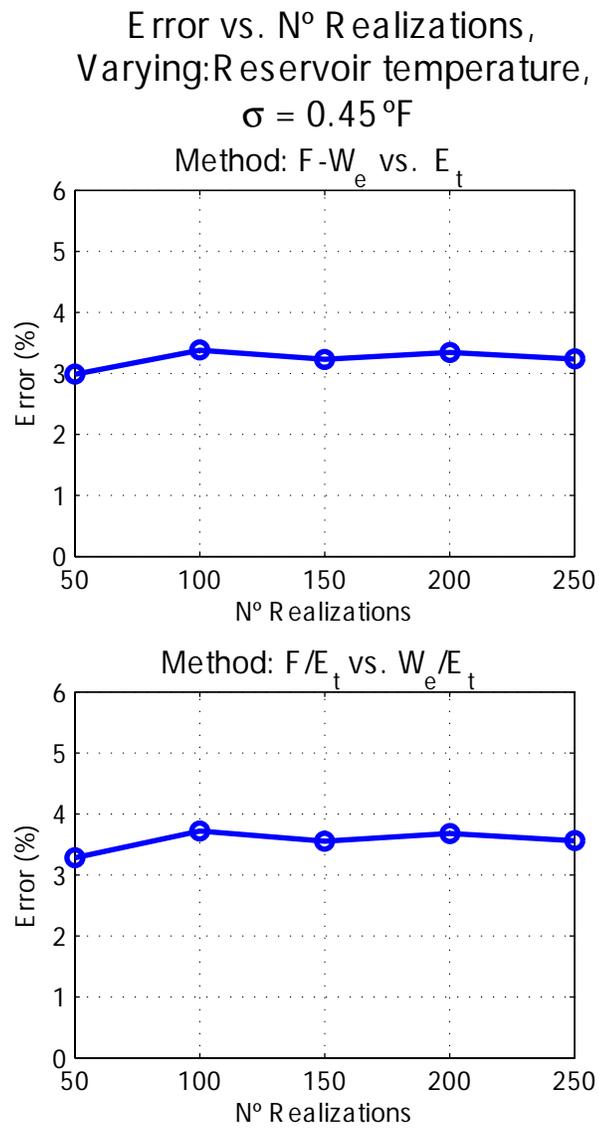


Figura 8.40: Gráfico del error en el POES vs. el número de realizaciones para el caso 15

Análisis final para la temperatura del yacimiento

Método de balance de materiales	Método 1	Método 2	Método 3
Promedio de la desviación estándar de los POES para un $\sigma = 0,02$ °F	0.013	0.11	0.015
Promedio de la desviación estándar de los POES para un $\sigma = 0,45$ °F	1.03	33.75	2.07
Método 1: F-We vs. Et			
Método 2: $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$			
Método 3: F/Et vs. We/Et			

Cuadro 8.25: Desviación estándar de los POES para la realizaciones de la temperatura del yacimiento

Los resultados muestran que los métodos (F-We vs. Et) y (F/Et vs. We/Et) poseen dispersiones comparables para todos los casos. El método $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$ muestra una dispersión en los valores obtenidos mucho mayor a los otros métodos evaluados. Los valores reportados del RMSE muestran que para todos los casos y métodos los ajustes realizados son excelentes. Los mayores valores de RMSE se encuentran cuando la desviación fue de $\sigma = 0,45$ °F y se utilizaron el método de la capa de gas y el método (F/Et vs. We/Et).

8.1.6. Diagramas ternarios

A continuación se presentan los diagramas ternarios para cada uno de los parámetros PVT y la historia de presión. En éstos se podrá apreciar como influye la incertidumbre de un parámetro en las estimaciones de balance de materiales de acuerdo al mecanismo de empuje del yacimiento. De la misma manera, se podrán identificar los mecanismos de empuje con menos sensibilidad a la incertidumbre en los parámetros PVT y de la historia de presión. Se presentará un diagrama ternario por cada método de balance de materiales utilizado. Al igual que para los mecanismos de empuje, se podrá identificar cuál método es el más robusto frente a la incertidumbre en los parámetros.

Historia de presión

Para una desviación en la historia de presión de $\sigma = 1$ psi (gráfico 8.41), los métodos (F-We vs. Et) y (F/Et vs. We/Et), presentan las regiones de error muy similares alcanzando un máximo entre 25 y 27 %. Para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) se alcanza un máximo de casi 2000 % de error. Para una desviación de $\sigma = 100$ psi (gráfico 8.42), el método (F-We vs. Et) presenta regiones de error de mucha menor magnitud que los otros métodos, alcanzando un máximo alrededor de 1600 %. El método (F/Et vs. We/Et) es el siguiente en cuanto a las regiones de error, con un máximo alrededor de 9000 %. Y por último, el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) presenta la región con mayor error al alcanzar un máximo de casi 15000 % de error. También se puede observar que la localización de los puntos en este gráfico difiere del anterior (gráfico 8.41), indicando que el error influye, grandemente, en los cálculos de la contribución relativa de energía al yacimiento. En otras palabras, el error no se refleja sólo en el POES calculado sino también en la capa de gas estimada por este método, y posteriormente en los cálculos de los índices de producción. El comportamiento general en estos gráficos ternarios es que a medida que aumenta la contribución de energía debida al influjo de agua va aumentando el error en los POES calculados. Por tanto el caso más grave es aquel donde el radio del acuífero es infinito. Esto se puede explicar debido a que en los cálculos de influjo de agua se utiliza las realizaciones de la historia de presión. Es por ello que estas desviaciones van a influir directamente en el influjo de agua calculado y posteriormente en el cálculo del POES. En resumen los mecanismos de empuje más robustos frente a la incertidumbre en la historia de presión son aquellos donde no hay influjo de agua, y los más susceptibles a error son aquellos donde el influjo de agua es la principal fuente de energía.

Ternary Plot, Varying: Pressure, $\sigma = 1$ psia, $L = 100$ realizations

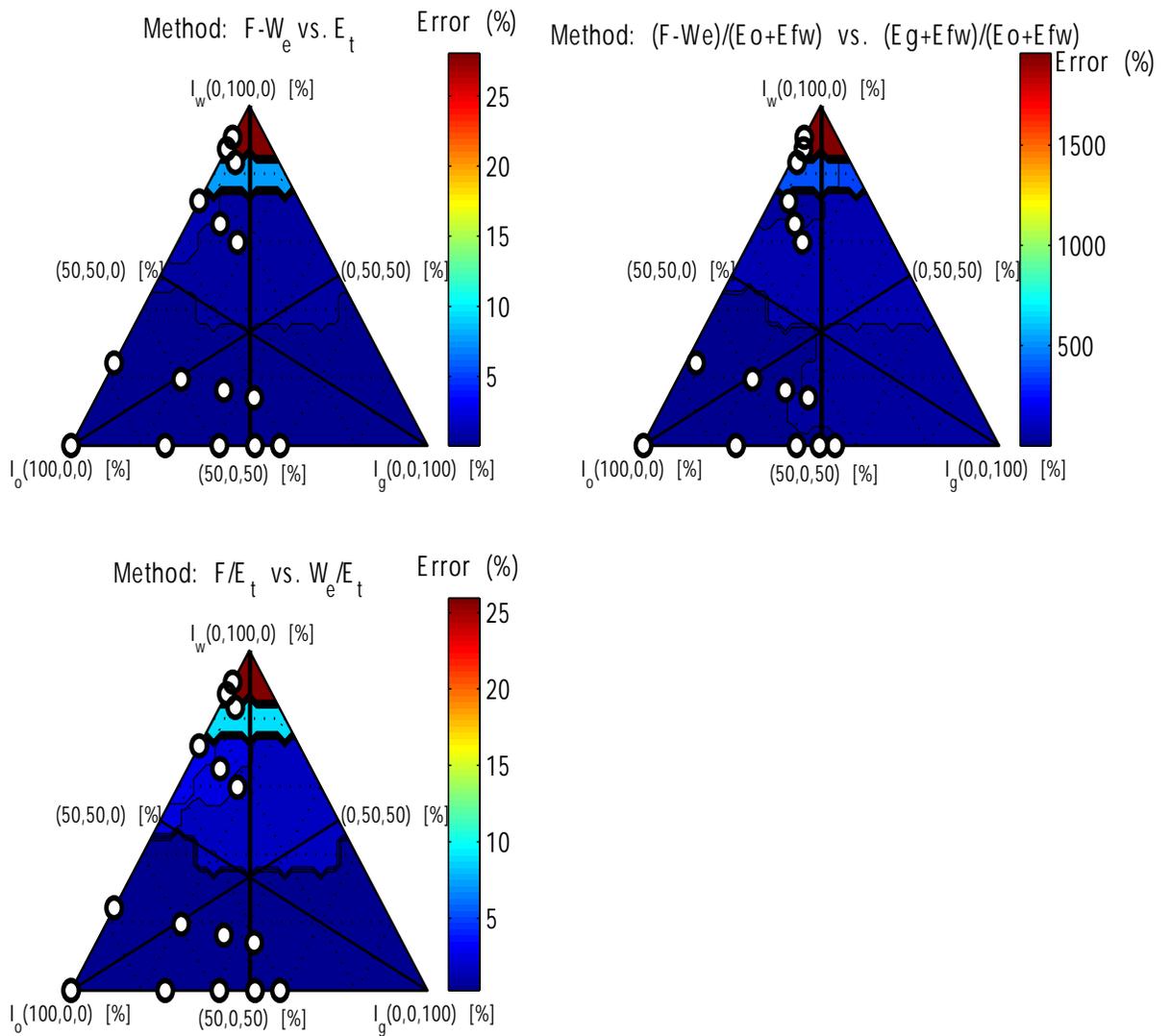


Figura 8.41: Gráfico ternario para la historia de presión. $\sigma = 1$ psi

Ternary Plot, Varying: Pressure, $\sigma = 100$ psia, $L = 100$ realizations

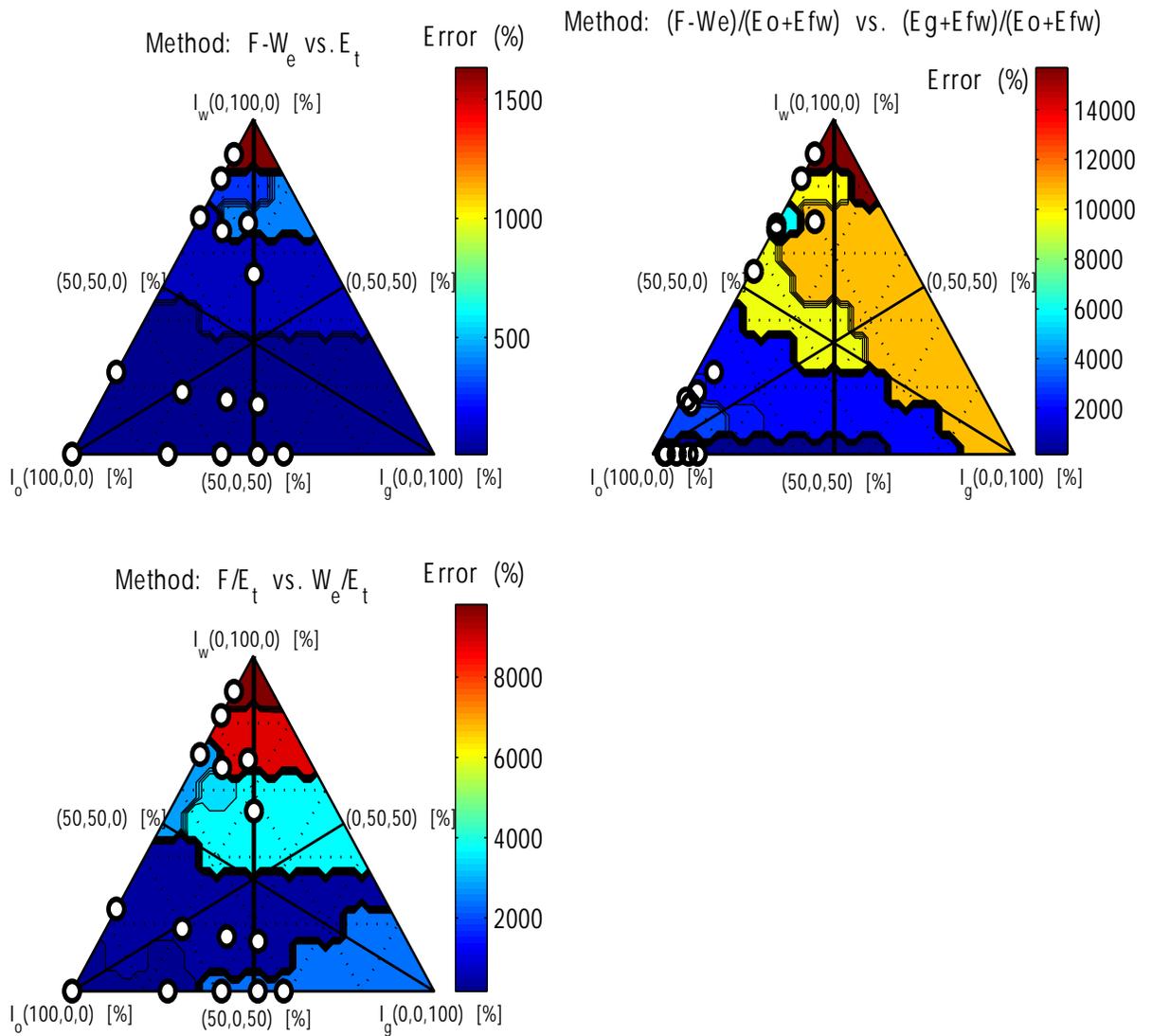


Figura 8.42: Gráfico ternario para la historia de presión. $\sigma = 100$ psi

Gravedad API

Para una desviación en la gravedad API de $\sigma = 0,05$ °API (gráfico 8.43), los métodos (F-We vs. Et) y (F/Et vs. We/Et), al igual que en el caso anterior, presentan regiones de error muy similares alcanzando un máximo entre 3 y 3.3% aproximadamente. Para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) se alcanza un máximo de casi 65% de error. Para una desviación de $\sigma = 0,25$ °API (gráfico 8.44), el método (F-We vs. Et) alcanzó un error máximo alrededor de 24%. El método (F/Et vs. We/Et) es el siguiente en cuanto a las regiones de error, con un máximo alrededor de 26%. El método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) presenta la región con mayor error alcanzando un máximo de casi 180% de error. A diferencia que para los gráficos ternarios de la historia de presión, los puntos en los gráficos se mantienen relativamente constantes, indicando que el error en la gravedad API no influye severamente en los cálculos de los índices de producción. También se observa el comportamiento general que para los gráficos ternarios de la historia de presión, a medida que aumenta la contribución de energía debida al influjo de agua va aumentando el error en los POES calculados. Sin embargo también se observa otra tendencia, a medida que aumenta la capa de gas va aumentando el error en los POES calculados, pero en menor grado que para el influjo de agua. El caso más grave continua siendo aquel donde el radio del acuífero es infinito. Los mecanismos de empuje más robustos frente a la incertidumbre en la gravedad API son aquellos donde no hay influjo de agua o una gran capa de gas, y los más susceptibles a error son aquellos donde existan gran influjo de agua o grandes capas de gas, ó una combinación de ambas.

Ternary Plot, Varying: API degrees, $\sigma = 0.05^\circ$ API, $L = 100$ realizations

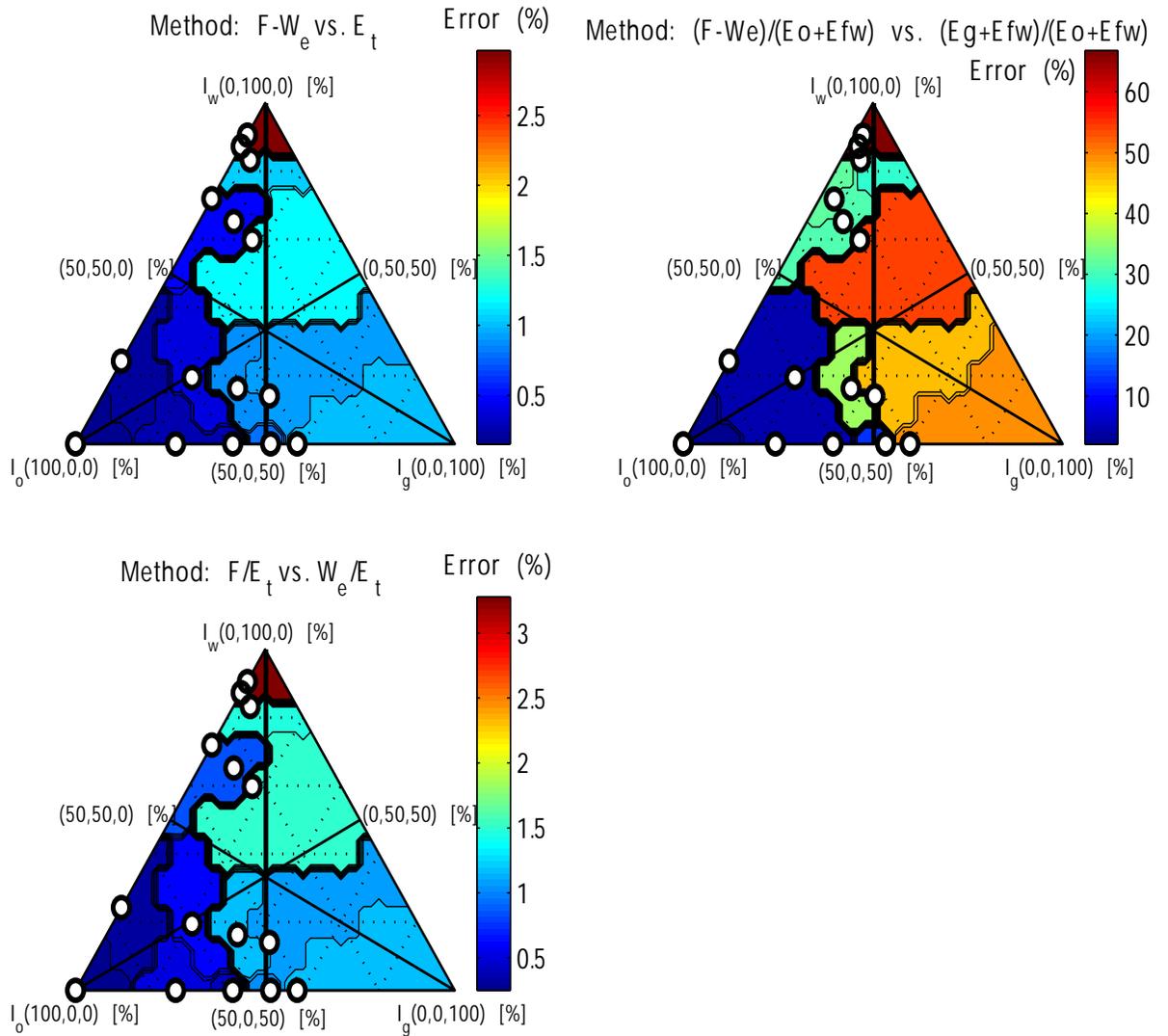


Figura 8.43: Gráfico ternario para la gravedad API. $\sigma = 0,05^\circ$ API

Ternary Plot, Varying: API degrees, $\sigma = 0.25^\circ$ API, $L = 100$ realizations

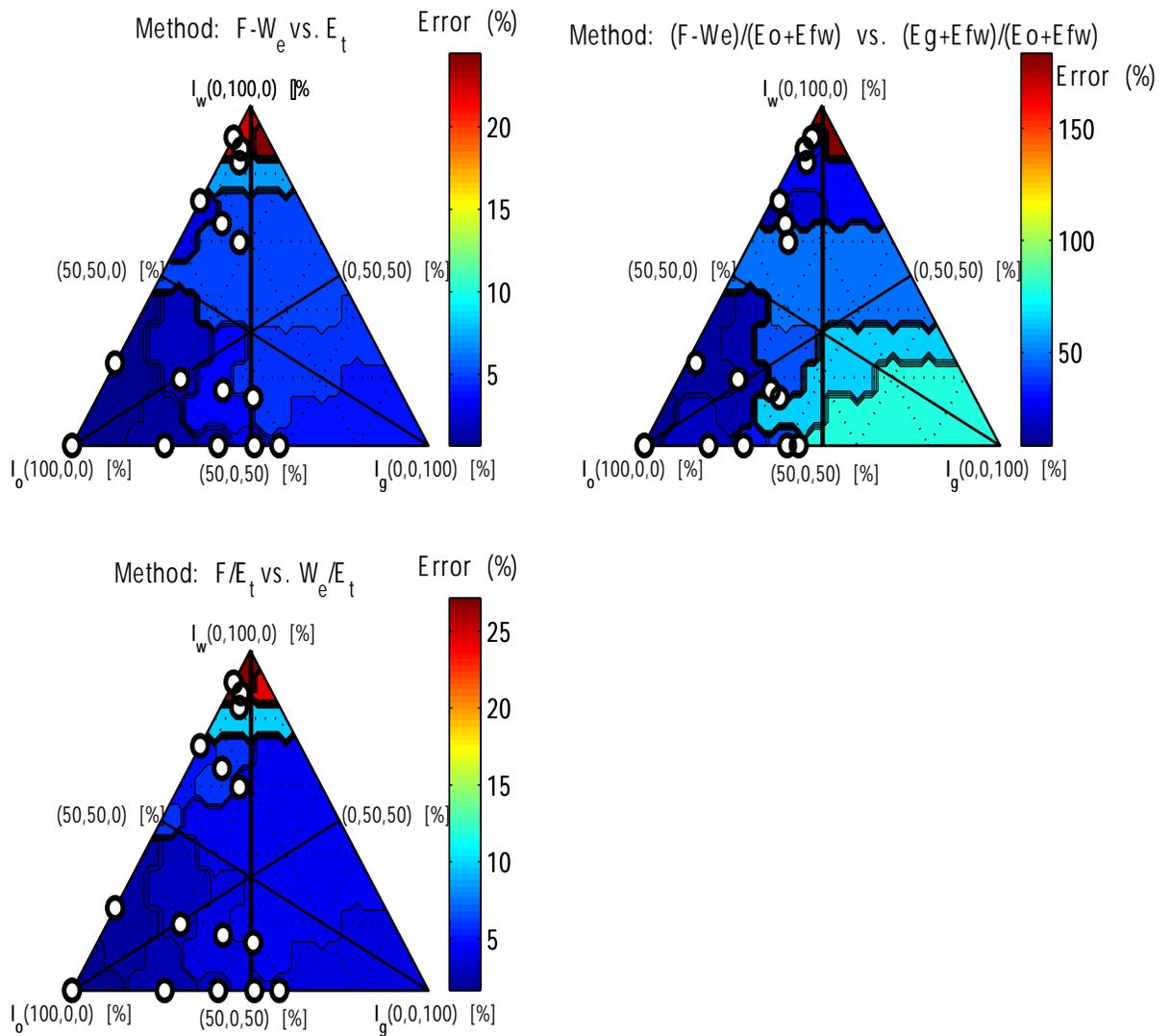


Figura 8.44: Gráfico ternario para la gravedad API. $\sigma = 0,25^\circ$ API

Gravedad específica del gas

Para una desviación en la gravedad API de $\sigma = 6e - 5$ (gráfico 8.45), los métodos (F-We vs. Et) y (F/Et vs. We/Et), al igual que los casos anteriores, presentan regiones de error parecidas alcanzando un máximo entre 0.6 y 0.65 % aproximadamente. Para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) se alcanza un máximo alrededor de 11 % de error. Para una desviación de $\sigma = 6e - 3$ (gráfico 8.46), el método (F-We vs. Et) alcanzó un error máximo entre 8.5 y 9 %. El método siguiente en cuanto a las regiones de error es (F/Et vs. We/Et), con un poco más de 10 %. El método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) presenta la región con mayor error alcanzando hasta 150 % de error. También se observa que los puntos en los gráficos ternarios se mantienen relativamente constantes, indicando que el error en la gravedad específica del gas no influye, en gran medida, en los cálculos de los índices de producción. De la misma manera que para los gráficos ternarios anteriores, a medida que aumenta el influjo de agua va aumentando el error en los POES calculados. Al igual que para el caso anterior de la gravedad API se la tendencia de que a medida que aumenta la capa de gas va aumentando el error en los POES calculados. El caso más grave continua siendo en donde el radio del acuífero es infinito. Los mecanismos de empuje más robustos frente a la incertidumbre en la gravedad específica del gas son aquellos donde la contribución relativa de energía debida al influjo de agua o una capa de gas es poca. Los mecanismos de empuje más susceptibles a error son donde existen grandes influjo de agua y/o grandes capas de gas, al igual que una combinación de ambas.

Ternary Plot, Varying: Gas specific gravity,
 $\sigma = 6e-5$ Specific Gravity Units, L = 100 realizations

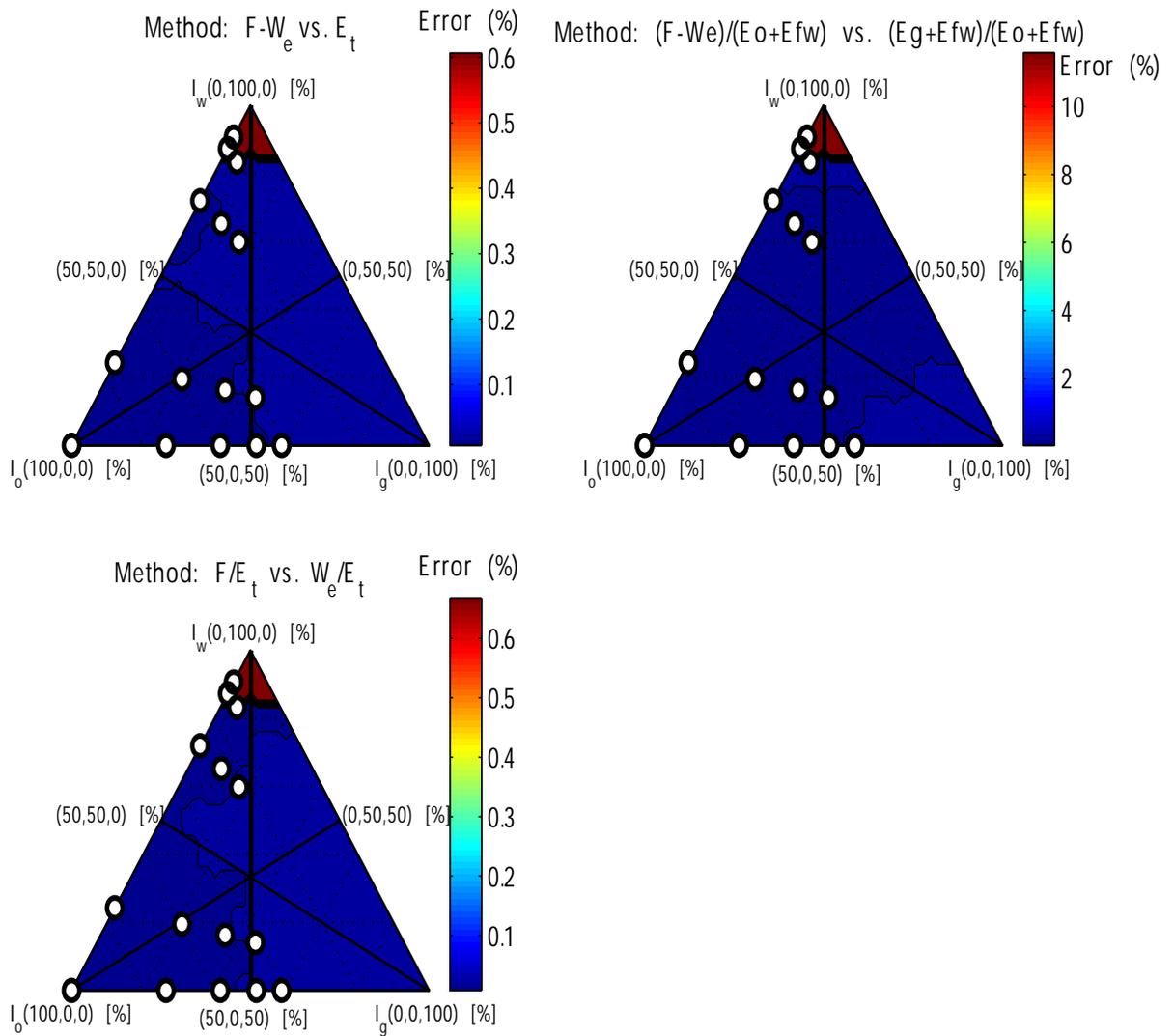


Figura 8.45: Gráfico ternario para la gravedad específica del gas. $\sigma = 6e - 05$

Ternary Plot, Varying: Gas specific gravity,
 $\sigma = 6e-3$ Specific Gravity Units, $L = 100$ realizations

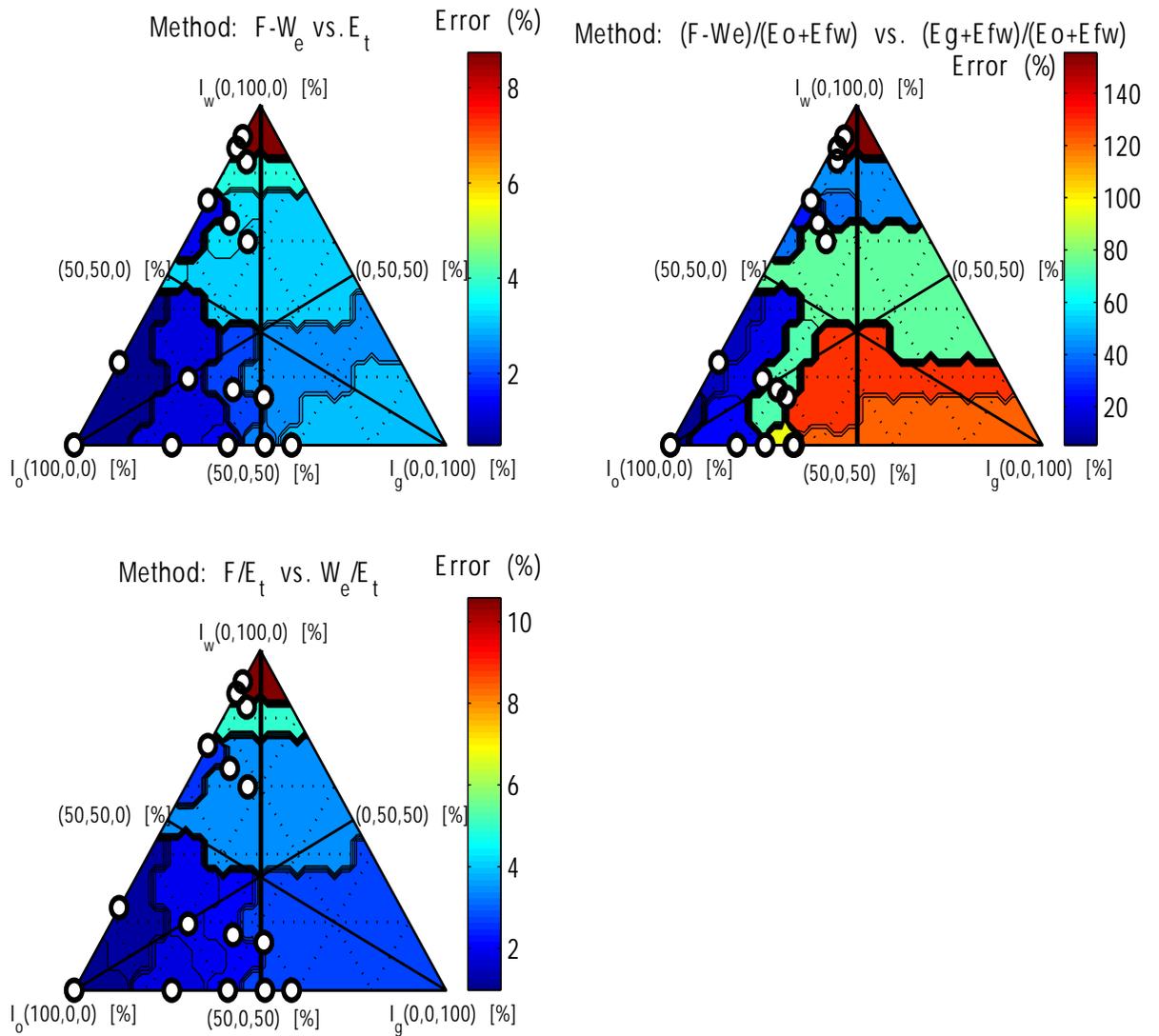


Figura 8.46: Gráfico ternario para la gravedad específica del gas. $\sigma = 0,0062$

Relación gas petróleo en solución

Para una desviación en la relación gas petróleo en solución de $\sigma = 0,738$ SCF/STB (gráfico 8.47), los métodos (F-We vs. Et) y (F/Et vs. We/Et), de igual forma como en los casos anteriores, presentan regiones de error muy similares alcanzando un alrededor de 25-26 % aproximadamente. Para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) el error máximo es de 55 % aproximadamente. Para una desviación de $\sigma = 14,76$ SCF/STB (gráfico 8.48), el método (F-We vs. Et) alcanzó un error máximo de 35 % aproximadamente. El método (F/Et vs. We/Et) obtuvo un error máximo alrededor de 40 %. El método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) presenta la región con mayor error alcanzando hasta 75 % de error. Como en los casos anteriores, se observa que los puntos en los gráficos ternarios se mantienen relativamente constantes, por tanto, el error en la relación gas petróleo en solución no influye en gran magnitud en los cálculos de los índices de producción. También se observa que a medida que aumenta el influjo de agua va aumentando el error en los POES calculados. De igual forma se observa la tendencia de que a medida que aumenta la capa de gas va aumentando el error en los POES calculados. El caso más grave fue aquel donde la capa de gas tuvo un valor de 1 (para una desviación de $\sigma = 14,76$ SCF/STB). En lo referente a la robustez de los mecanismos de empuje se obtuvo los mismos resultados que para los casos anteriores, aquellos donde la contribución relativa de energía debida al influjo de agua o una capa de gas es poca son los más robustos; y aquellos donde existen grandes influjo de agua y/o grandes capas de gas, al igual que una combinación de ambas, son los más susceptibles.

Ternary Plot, Varying: Solution gas oil ratio, $\sigma = 0.738$ SCF/STB, $L = 100$ realizations

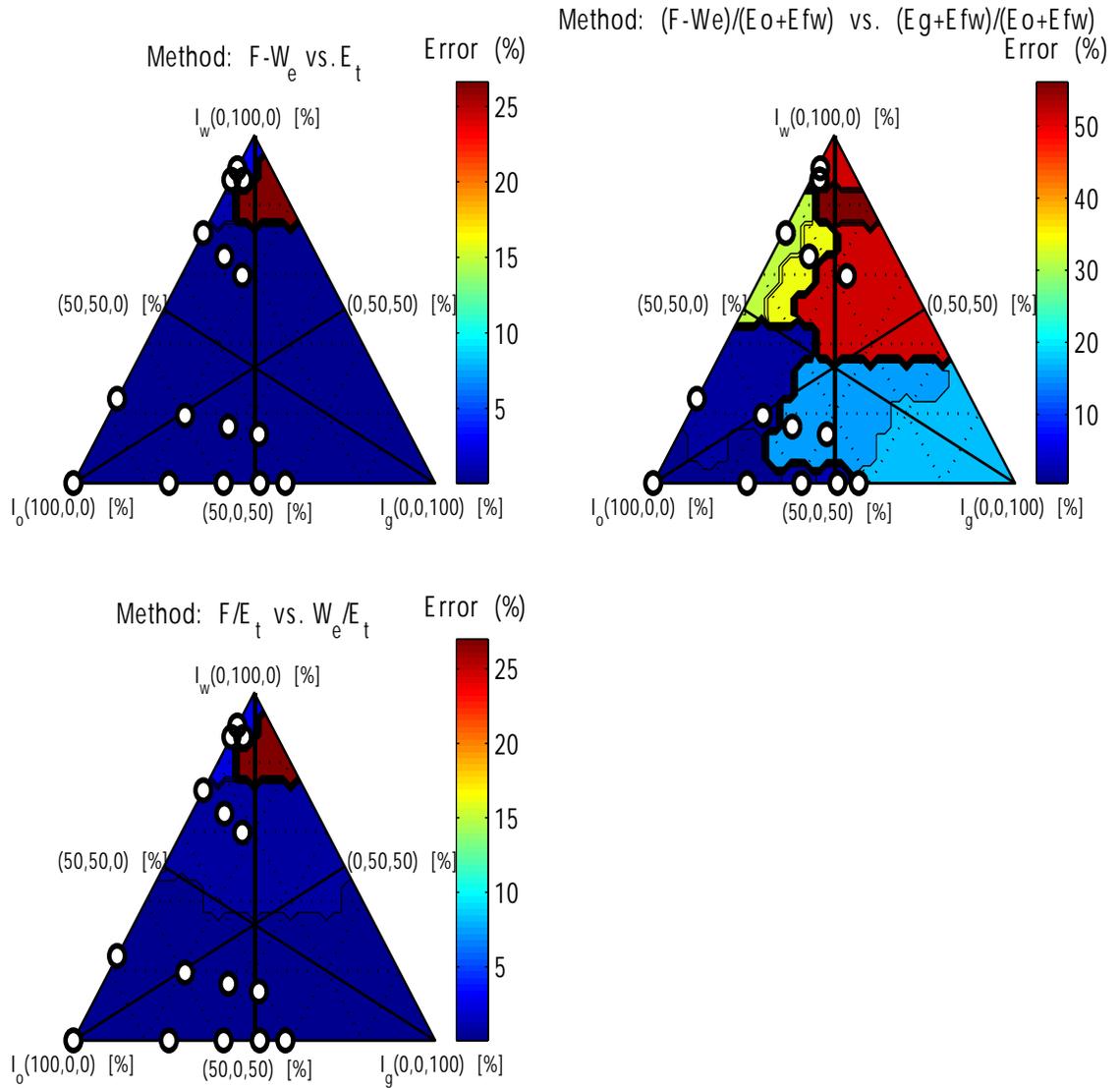


Figura 8.47: Gráfico ternario para la relación gas petróleo en solución. $\sigma = 0,738$ SCF/STB

Ternary Plot, Varying: Solution gas oil ratio, $\sigma = 14.76$ SCF/STB, $L = 100$ realizations

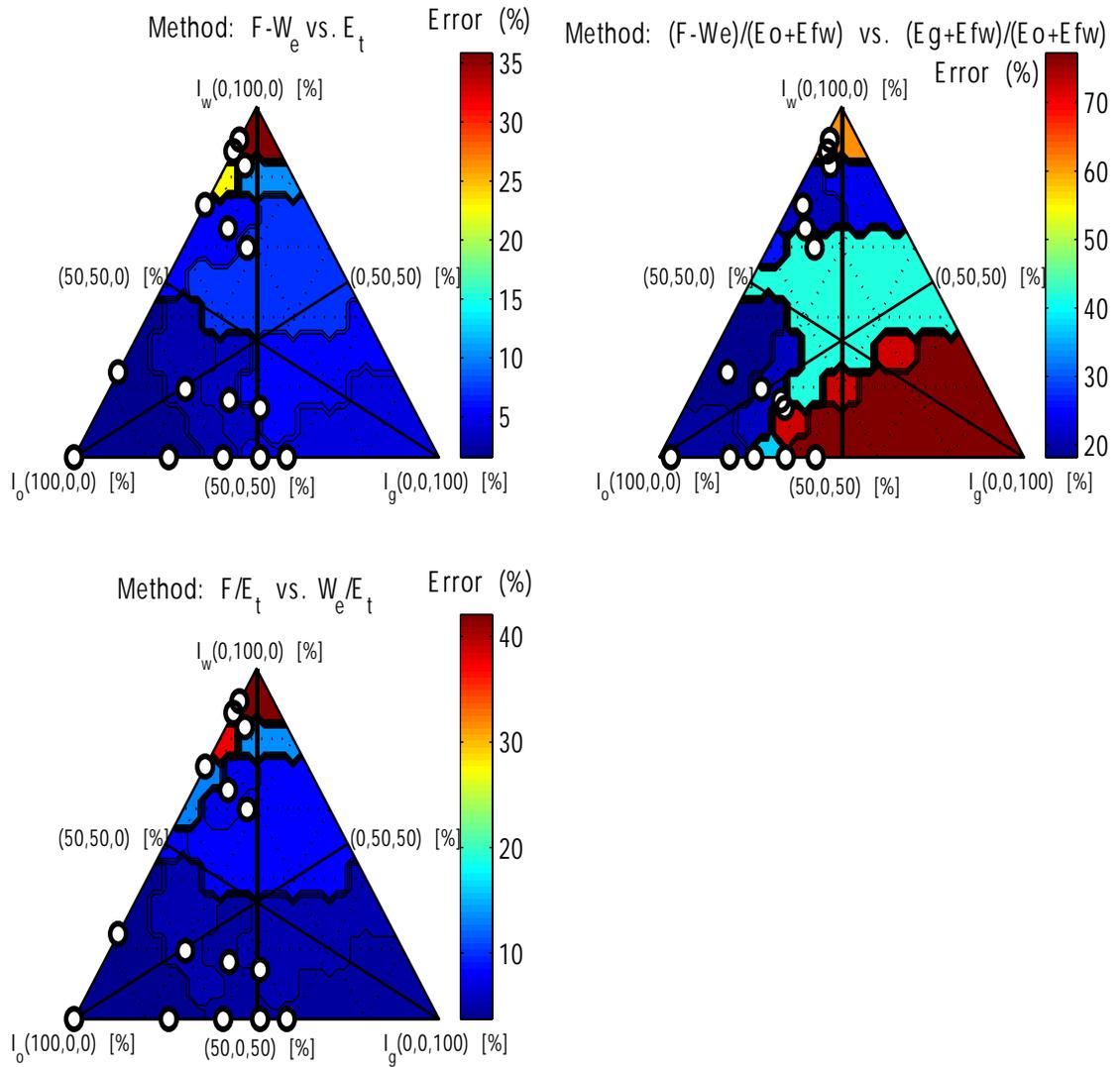


Figura 8.48: Gráfico ternario para la relación gas petróleo en solución. $\sigma = 14,76$ SCF/STB

Temperatura del yacimiento

También para este caso, para una desviación en la temperatura del yacimiento de $\sigma = 0,02$ °F (gráfico 8.49), los métodos (F-We vs. Et) y (F/Et vs. We/Et) presentan regiones de error muy similares alcanzando un alrededor de 0.02 % aproximadamente. Para el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) el error máximo es cerca de 2.5 % aproximadamente. Para una desviación de $\sigma = 0,45$ °F (gráfico 8.50), el método (F-We vs. Et) alcanzó un error máximo cerca de 3.3 % aproximadamente. El método (F/Et vs. We/Et) obtuvo un error máximo alrededor de 3.6 %. El método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) presenta la región con mayor error alcanzando alrededor de 90 % de error. También se observa en este caso que los puntos en los gráficos ternarios se mantienen relativamente constantes, por lo que el error en la temperatura del yacimiento no influye en gran medida en los cálculos de los índices de producción. De igual forma se observa la tendencia general de que a medida que aumenta el influjo de agua va aumentando el error en los POES calculados, y que a medida que aumenta la capa de gas va aumentando el error. El caso más grave, para los casos evaluados, fue aquel donde el acuífero posee un radio $r_e D = \infty$. En lo referente a los mecanismos de empuje también se observó el comportamiento general: para los modelos donde la contribución relativa de energía debida al influjo de agua o una capa de gas es poca los mecanismos de empuje son los más robustos; y aquellos donde existen grandes influjo de agua y/o grandes capas de gas, al igual que una combinación de ambas, son los más susceptibles.

Ternary Plot, Varying: Reservoir temperature, $\sigma = 0.02 \text{ } ^\circ\text{F}$, $L = 100$ realizations

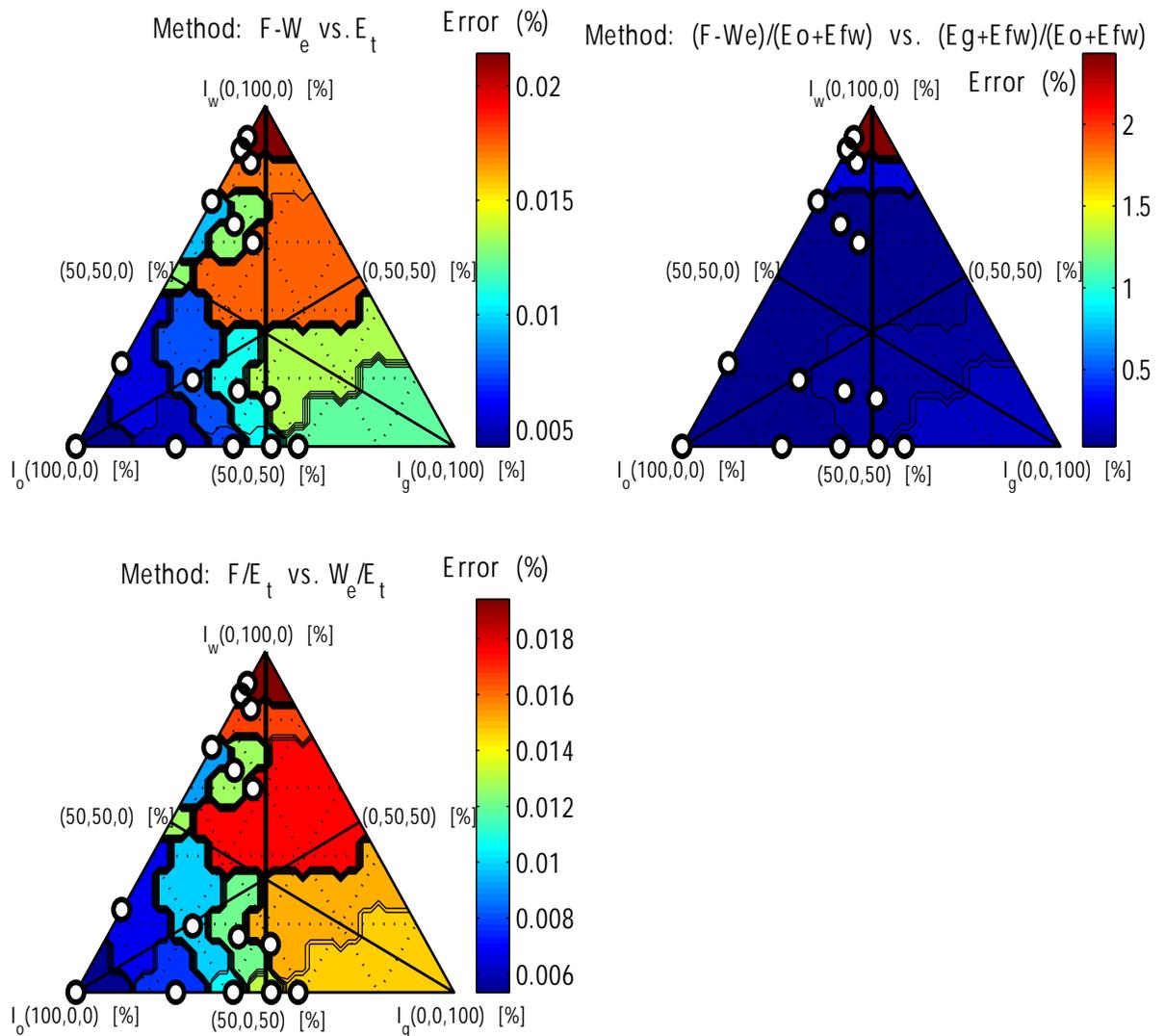


Figura 8.49: Gráfico ternario para la temperatura del yacimiento. $\sigma = 0,02 \text{ } ^\circ\text{F}$

Ternary Plot, Varying: Reservoir temperature, $\sigma = 0.45^\circ\text{F}$, $L = 100$ realizations

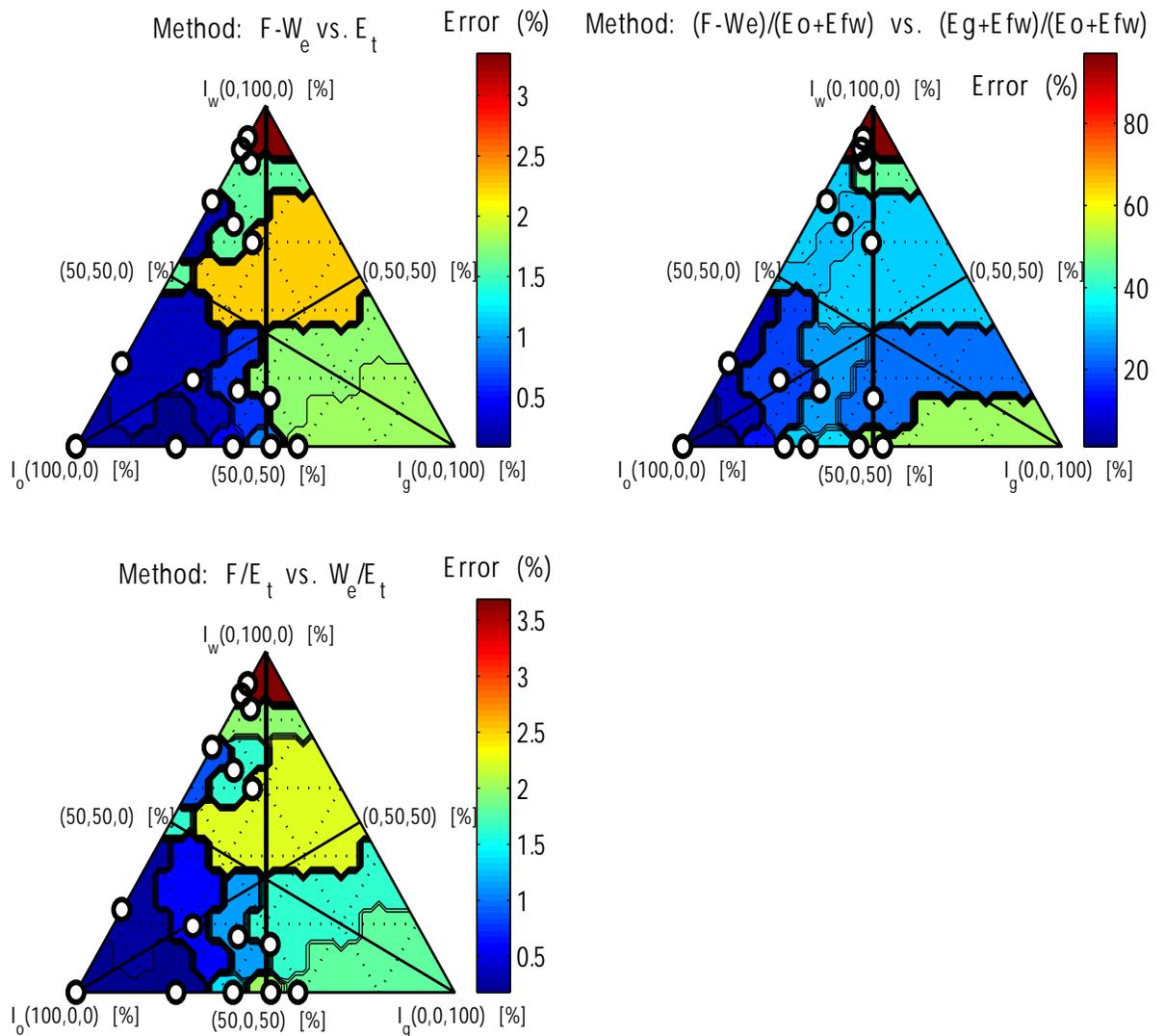


Figura 8.50: Gráfico ternario para la temperatura del yacimiento. $\sigma = 0,45^\circ\text{F}$

8.1.7. Resumen de resultados

En los cuadros de esta sección se utilizó la siguiente notación:

Método 1 = F-We vs. Et

Método 2 = $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$

Método 3 = F/Et vs. We/Et

m = promedio

s = desviación estándar

Realizaciones con desviaciones bajas

Caso	Método N°	Estadística	
		m (MMSTB)	s (MMSTB)
1	1	100,03	0,07
5	1	100,03	0,13
5	2	120,90	55,01
11	1	99,75	1,30
11	2	82,58	95,49
11	3	99,18	2,65
15	1	96,51	26,05
15	3	96,58	24,06

Cuadro 8.26: Historia de presión. $\sigma = 1$ psi

Caso	Método N°	Estadística	
		m (MMSTB)	s (MMSTB)
1	1	100,11	0,08
5	1	100,35	0,30
5	2	76,07	51,46
11	1	100,30	0,39
11	2	99,62	31,05
11	3	100,11	0,82
15	1	99,41	2,28
15	3	99,29	2,50

Cuadro 8.27: Gravedad API. $\sigma = 0.05$ °API

Caso	Método N°	Estadística	
		m (MMSTB)	s (MMSTB)
1	1	100,00	0,01
5	1	100,00	0,02
5	2	99,93	0,23
11	1	100,00	0,02
11	2	100,00	0,03
11	3	100,00	0,01
15	1	100,00	0,02
15	3	100,00	0,02

Cuadro 8.28: Gravedad específica del gas. $\sigma = 5e-05$

Caso	Método N°	Estadística	
		m (MMSTB)	s (MMSTB)
1	1	100,04	0,06
5	1	100,11	0,11
5	2	103,49	15,40
11	1	100,01	0,17
11	2	103,00	31,23
11	3	99,73	0,57
15	1	98,03	2,54
15	3	97,81	2,82

Cuadro 8.29: Relación gas-petróleo en solución. $\sigma = 0.738$ SCF/STB

Caso	Método N°	Estadística	
		m (MMSTB)	s (MMSTB)
1	1	100,00	0,01
5	1	100,00	0,01
5	2	99,92	0,19
11	1	100,00	0,01
11	2	100,00	0,03
11	3	100,00	0,01
15	1	99,99	0,02
15	3	99,99	0,02

Cuadro 8.30: Temperatura del yacimiento. $\sigma = 0.02$ °F

Realizaciones con desviaciones altas

Caso	Método N°	Estadística	
		m (MMSTB)	s (MMSTB)
1	1	100,68	7,34
5	1	100,35	16,6
5	2	178,95	7388,01
11	1	72,63	73,29
11	2	142,44	6806,53
11	3	119,8	1227,72
15	1	-756,54	1216,53
15	3	-59,12	16407,87

Cuadro 8.31: Historia de presión. $\sigma = 100$ psi

Caso	Método N°	Estadística	
		m (MMSTB)	s (MMSTB)
1	1	100,6	0,49
5	1	103,92	3,11
5	2	118,52	92,32
11	1	103,02	2,81
11	2	113,69	24,74
11	3	102,6	3,22
15	1	109,88	21,88
15	3	111,41	25,59

Cuadro 8.32: Gravedad API. $\sigma = 0.25$ °API

Caso	Método N°	Estadística	
		m (MMSTB)	s (MMSTB)
1	1	100,26	0,51
5	1	102,36	2,77
5	2	153,29	135
11	1	101,82	1,99
11	2	118,8	36,67
11	3	101,95	2,19
15	1	1252,27	159,05
15	3	1073,59	125,83

Cuadro 8.33: Gravedad específica del gas. $\sigma = 0.00615$

Caso	Método N°	Estadística	
		m (MMSTB)	s (MMSTB)
1	1	100,67	1,08
5	1	103,94	3,28
5	2	134,76	72,79
11	1	104,07	5,04
11	2	110,45	28,04
11	3	104,67	7,96
15	1	114,64	31,41
15	3	117,23	36,28

Cuadro 8.34: Relación gas-petróleo en solución. $\sigma = 14.76$ SCF/STB

Caso	Método N°	Estadística	
		m (MMSTB)	s (MMSTB)
1	1	100,02	0,14
5	1	100,15	0,18
5	2	114,58	46,8
11	1	100,29	1,31
11	2	97,86	31,21
11	3	100,02	1,42
15	1	98,48	2,5
15	3	98,29	2,72

Cuadro 8.35: Temperatura del yacimiento. $\sigma = 0.45$ °F

8.1.8. Análisis final para las realizaciones

En base a todos los resultados de las realizaciones calculadas se clasificaron los parámetros de acuerdo al error obtenido para cada método de balance de materiales.

Método de balance de materiales	Método 1	Método 2	Método 3
mucha incertidumbre	P	P	P
incertidumbre alta	γ_g	γ_g	γ_g
incertidumbre media	R_s	$^{\circ}\text{API}$	R_s
incertidumbre baja	$^{\circ}\text{API}$	R_s	$^{\circ}\text{API}$
poca incertidumbre	T_y	T_y	T_y

donde:

Método 1: F-We vs. Et
Método 2: $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$
Método 3: F/Et vs. We/Et
P: historia de presión, γ_g : gravedad específica del gas
$^{\circ}\text{API}$: gravedad API, T_y : temperatura del yacimiento
R_s : relación gas petróleo en solución

Cuadro 8.36: Clasificación de los parámetros de acuerdo a la incertidumbre en los POES calculados

En general, el método (F-We vs. Et) resultó ser el método más robusto frente a todos los rangos de incertidumbre y todos los modelos de yacimientos evaluados. El método (F/Et vs. We/Et) también mostró resultados comparables al método anterior excepto para cuando existen desviaciones en la historia de presión, frente a las cuales este método muestra hipersensibilidad. Por último, el método $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$, en comparación con los otros dos, muestra gran sensibilidad a los rangos de incertidumbre evaluados para todos los parámetros y modelos de yacimientos evaluados. Hay que mencionar que mientras a los otros métodos se les proporcionó el valor de la capa de gas, a este método $((F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw))$ se le permitió calcularlo. Esto pudo influir en los resultados obtenidos debido a que el método ya no sólo calculaba una incógnita (POES), sino dos (POES y capa de gas), dando la posibilidad de que existiesen múltiples soluciones. En resumen el método $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$ muestra hipersensibilidad a todos los parámetros evaluados. Todos los métodos de balance de materiales mostraron que el margen de error se incrementa si el modelo de yacimiento tiene una gran capa de gas ó un gran influjo de agua. Por tanto para aquellos modelos de yacimiento donde la capa de gas era pequeña o no existía influjo de agua, los errores en los POES calculados eran mínimos no importando el rango de incertidumbre evaluado. Se obtuvieron

los mayores valores del RMSE para el caso 15 con la desviación más alta para todos los parámetros evaluados, en otras palabras, para este caso los ajustes realizados no fueron tan buenos como para los demás casos con los mismo valores de desviación. Sin embargo, hay que mencionar que la linealidad del ajuste (es decir, el valor del RMSE) no indica que el valor obtenido del POES esté correcto, dado que se presentaron casos donde la tendencia lineal era casi perfecta, y no obstante, el valor del POES obtenido del ajuste no era el correcto (incluso obteniendo valores negativos). Por tanto no se puede utilizar la linealidad de los puntos en los gráficos de los métodos evaluados como una medida de la calidad de los cálculos de balance de materiales.

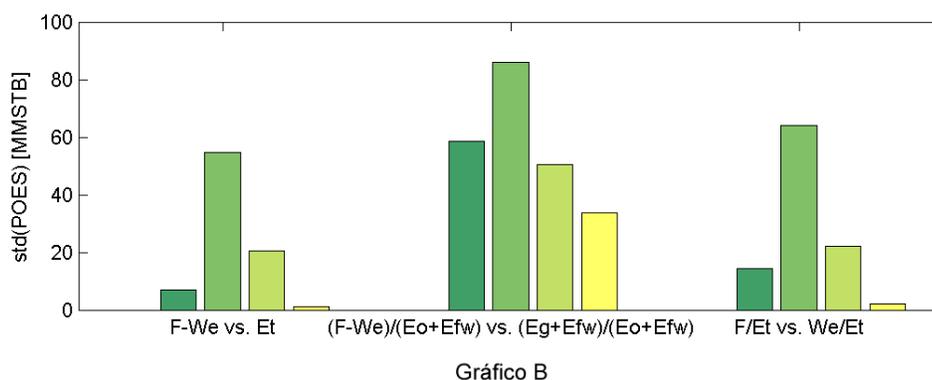
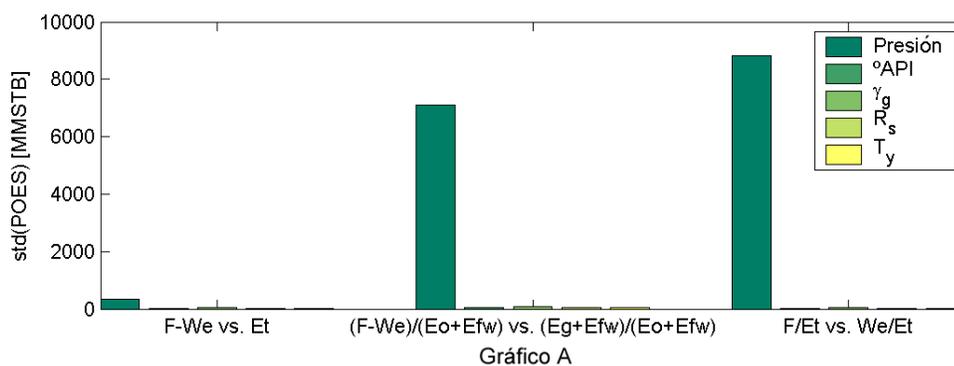


Figura 8.51: Clasificación de los parámetros de acuerdo a la incertidumbre en los POES calculados y los métodos utilizados. Gráfico A, incluyendo la presión. Gráfico B, excluyendo la presión

8.2. Diseño experimental

En las próximas secciones se muestran los resultados para los dos tipos de diseños experimentales: full factorial y Box Behnken. Se mostrarán las matrices de diseño en forma codificada; es decir, con los niveles dados por -1 para el nivel bajo, 0 para el nivel del caso base, y 1 para el nivel alto. También se utilizará la siguiente notación:

$$\begin{array}{c}
 \boxed{\pm 1 \mid 0 \mid 0 \mid 0 \mid 0} = \begin{array}{|c|c|c|c|c|} \hline -1 & 0 & 0 & 0 & 0 \\ \hline +1 & 0 & 0 & 0 & 0 \\ \hline \end{array} \\
 \\
 \boxed{\pm 1 \mid \pm 1 \mid 0 \mid 0 \mid 0} = \begin{array}{|c|c|c|c|c|} \hline -1 & -1 & 0 & 0 & 0 \\ \hline +1 & -1 & 0 & 0 & 0 \\ \hline -1 & +1 & 0 & 0 & 0 \\ \hline +1 & +1 & 0 & 0 & 0 \\ \hline \end{array}
 \end{array}$$

Cuadro 8.37: Notación para las matrices de los diseños experimentales

En el cuadro anterior se mostró la notación para variar 1 y 2 factores, de la misma manera se puede deducir la notación para más de 3 factores. En la página siguiente se muestra la matriz de diseño full factorial para 5 factores utilizada en los cálculos. En total son 243 corridas que se realizaron por matriz. Para la matriz de diseño donde sólo se variaron los parámetros PVT el número de corridas fue de 81 (4 parámetros con 3 niveles cada uno).

$^{\circ}\text{API}$	γ_g	R_s (SCF/STB)	T_y ($^{\circ}\text{F}$)	std(P) (psia)
0	0	0	0	0
0	0	0	0	± 1
0	0	0	± 1	0
0	0	0	± 1	± 1
0	0	± 1	0	0
0	0	± 1	0	± 1
0	0	± 1	± 1	0
0	0	± 1	± 1	± 1
0	± 1	0	0	0
0	± 1	0	0	± 1
0	± 1	0	± 1	0
0	± 1	0	± 1	± 1
0	± 1	± 1	0	0
0	± 1	± 1	0	± 1
0	± 1	± 1	± 1	0
0	± 1	± 1	± 1	± 1
± 1	0	0	0	0
± 1	0	0	0	± 1
± 1	0	0	± 1	0
± 1	0	0	± 1	± 1
± 1	0	± 1	0	0
± 1	0	± 1	0	± 1
± 1	0	± 1	± 1	0
± 1	0	± 1	± 1	± 1
± 1	± 1	0	0	0
± 1	± 1	0	0	± 1
± 1	± 1	0	± 1	0
± 1	± 1	0	± 1	± 1
± 1	± 1	± 1	0	0
± 1	± 1	± 1	0	± 1
± 1	± 1	± 1	± 1	0
± 1	± 1	± 1	± 1	± 1

Cuadro 8.38: Matriz de diseño full factorial para todos los parámetros (5 factores)

$^{\circ}\text{API}$	γ_g	R_s (SCF/STB)	T_y ($^{\circ}\text{F}$)
0	0	0	0
0	0	0	± 1
0	0	± 1	0
0	0	± 1	± 1
0	± 1	0	0
0	± 1	0	± 1
0	± 1	± 1	0
0	± 1	± 1	± 1
± 1	0	0	0
± 1	0	0	± 1
± 1	0	± 1	0
± 1	0	± 1	± 1
± 1	± 1	0	0
± 1	± 1	0	± 1
± 1	± 1	± 1	0
± 1	± 1	± 1	± 1

Cuadro 8.39: Matriz de diseño full factorial para los parámetros PVT

$^{\circ}\text{API}$	γ_g	R_s (SCF/STB)	T_y ($^{\circ}\text{F}$)	std(P) (psia)
± 1	± 1	0	0	0
0	0	± 1	± 1	0
0	± 1	0	0	± 1
± 1	0	± 1	0	0
0	0	0	± 1	± 1
0	± 1	± 1	0	0
± 1	0	0	± 1	0
0	0	± 1	0	± 1
± 1	0	0	0	± 1
0	± 1	0	± 1	0
0	0	0	0	0

Cuadro 8.40: Matriz de diseño Box Behnken para todos los parámetros (5 factores)

El cuadro 8.40 muestra la matriz de diseño para 5 factores Box Behnken en forma codificada. El número total de corridas es de tan sólo de 41 por matriz. El cuadro 8.41 muestra la matriz de diseño para 4 factores Box Behnken en forma codificada (matriz para los parámetros PVT). El número total de corridas es apenas 25 por matriz.

$^{\circ}\text{API}$	γ_g	R_s (SCF/STB)	T_y ($^{\circ}\text{F}$)
± 1	± 1	0	0
0	0	± 1	± 1
± 1	0	± 1	0
0	± 1	± 1	0
± 1	0	0	± 1
0	± 1	0	± 1
0	0	0	0

Cuadro 8.41: Matriz de diseño Box Behnken para los parámetros PVT

Recordando que en el capítulo de la metodología se especificaron los niveles para dos matrices de diseños, entonces, se generaron 8 matrices de diseño en total, es decir, 4 matrices donde se variaron todos los parámetros los cuales consisten de dos matrices de diseño full factorial con desviaciones bajas y altas; y dos matrices de diseño Box Behnken con desviaciones bajas y altas. Las otras 4 matrices corresponden al diseño donde sólo se varían los parámetros PVT (4 factores). En la siguiente sección se presentan los resultados de los diseños full factorial. En la sección 8.2.2 se muestran los resultados de los diseños Box Behnken, y por último se realiza un análisis en conjunto de los dos tipos de diseños en la sección 8.2.4.

8.2.1. Diseños full factorial

A continuación se presentan para cada uno de los casos mencionados en la metodología los resultados de la matriz de diseño N°1 (con desviaciones bajas) y de la matriz de diseño N°2 (con desviaciones altas), junto al análisis ANOVA al final de la tabla.

Caso 1

Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	101.88	77.41
POES (mediana) [MMSTB]	100.81	90.78
POES (máximo) [MMSTB]	109.68	137.39
POES (mínimo) [MMSTB]	97.45	9.67
POES (desviación estándar) [MMSTB]	2.61	33.35
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.24	5.44
N-way ANOVA	(valor p)	
Gravedad API, °API	0.5675	0.9828
Gravedad específica del gas, γ_g	0.3871	0.3315
Relación gas petróleo en solución, R_s	0.0185	0
Temperatura del yacimiento, T_y	0	0
Historia de presión	0	0

Cuadro 8.42: Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero

Para la matriz de diseño N°1 (con desviaciones bajas) los resultados son bastante concluyentes, existe poca dispersión en los POES calculados. Muestra de ello es que se obtuvo una desviación estándar de apenas 2.61 MMSTB entre todas las corridas de la matriz de diseño. El análisis ANOVA para esta matriz indica que los parámetros altamente significativos son: la relación gas petróleo en solución, la temperatura del yacimiento, y la historia de presión; ésto debido a que sus valores p son menores a 0.05 (el nivel de significancia establecido para la prueba). Para la matriz de diseño N°2 (con desviaciones altas) los resultados muestran un aumento en la dispersión del POES, con una desviación estándar de 33.35 MMSTB. El análisis ANOVA arrojó que los parámetros altamente significativos son: la relación gas petróleo en solución, la temperatura del yacimiento, y la historia de presión. De igual forma el análisis ANOVA indica que la gravedad API, y la gravedad específica del gas son poco significativas en los resultados. Los valores del RMSE para este caso indican que se obtuvieron buenos ajustes para ambas matrices de diseño.

Parámetros PVT. Caso 1		
Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	100.62	101.63
POES (mediana) [MMSTB]	100.59	101.47
POES (máximo) [MMSTB]	102.26	107.45
POES (mínimo) [MMSTB]	99.51	97.63
POES (desviación estándar) [MMSTB]	0.67	2.31
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.063	0.243
N-way ANOVA	(valor p)	
Gravedad API, °API	0.039	0.3441
Gravedad específica del gas, γ_g	0.132	0
Relación gas petróleo en solución, R_s	0	0
Temperatura del yacimiento, T_y	0.9986	0.112

Cuadro 8.43: Resultados de la matriz de diseño de los parámetros PVT para el caso 1

Como se mencionó en la metodología aparte de realizar un diseño experimental con todos los parámetros PVT y la historia de presión, se realizó de manera independiente un diseño experimental donde actuaban sólo los parámetros PVT. Para la matriz de diseño N°1 de los parámetros PVT los resultados muestran que la desviación estándar es 0.67 MMSTB. El análisis ANOVA señala que los parámetros altamente significativos son: la gravedad API, y la relación gas petróleo en solución. Los demás parámetros son poco significativos en los resultados. Para la matriz de diseño N°2 se obtuvo una desviación estándar de 2.31 MMSTB. El análisis ANOVA mostró que los parámetros altamente significativos son: la gravedad específica del gas y la relación gas petróleo en solución; también indica que la gravedad API, y la temperatura del yacimiento son poco significativas en los resultados. Para ambas matrices de diseño se tiene que los valores del RMSE indican que se realizaron ajustes excelentes en los datos.

Caso 5

Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	98.73	52.43
POES (mediana) [MMSTB]	100.11	62.74
POES (máximo) [MMSTB]	113.20	225.41
POES (mínimo) [MMSTB]	81.79	-30.57
POES (desviación estándar) [MMSTB]	6.97	56.36
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	1.111	5.627
N-way ANOVA	(valor p)	
Gravedad API, °API	0.0146	0.7587
Gravedad específica del gas, γ_g	0.8757	0.8697
Relación gas petróleo en solución, R_s	0.0001	0
Temperatura del yacimiento, T_y	0.4132	0
Historia de presión	0	0
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	510.52	216.27
POES (mediana) [MMSTB]	229.77	229.94
POES (máximo) [MMSTB]	9087.20	7412.62
POES (mínimo) [MMSTB]	-2171.27	-17919.43
POES (desviación estándar) [MMSTB]	969.01	1293.99
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	922.531	20123.919
N-way ANOVA	(valor p)	
Gravedad API, °API	0.3528	0.8213
Gravedad específica del gas, γ_g	0.3056	0.2697
Relación gas petróleo en solución, R_s	0.4515	0.2184
Temperatura del yacimiento, T_y	0.0259	0.6452
Historia de presión (valor p)	0	0.8642

Cuadro 8.44: Caso 5. Tamaño de la capa de gas, $m = 1$.
No tiene acuífero

Este caso representa un yacimiento con una gran capa de gas ($m=1.0$) y sin acuífero. Es por ello que se aplicó el método de capa de gas, además del método (F-We vs. Et) para los cálculos del POES.

Método (F-We vs. Et) Para la matriz N°1 la desviación estándar obtenida fue de 6.97 MMSTB entre todas las corridas de la matriz de diseño. El análisis ANOVA para esta matriz indica que los parámetros más altamente significativos son: la gravedad API, la relación gas petróleo en solución, y la historia de presión. El análisis ANOVA también indica que la gravedad específica del gas, y la temperatura del yacimiento son poco significativos en los resultados. Para la matriz N°2 se obtuvo una desviación estándar de 56.36 MMSTB. El análisis ANOVA muestra que los parámetros altamente significativos son: la relación gas petróleo en solución, la temperatura del yacimiento, y la historia de presión. Además indica que la gravedad API, y la gravedad específica del gas son poco significativas en los resultados. Para ambas matrices los valores del RMSE indican buenos ajustes en los datos, resultados bastantes lineales.

Método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) Para la matriz N°1 los resultados muestran una gran dispersión en los POES calculados con una desviación estándar de 969.01 MMSTB. Para la matriz N°2 se obtuvo una desviación estándar de 1293.99 MMSTB. El análisis ANOVA para la matriz N°1 indica que los parámetros más altamente significativos fueron la temperatura del yacimiento, y la historia de presión. Para la matriz N°2 el análisis ANOVA muestra que ninguno de los parámetros resultó altamente significativo. Es de hacer notar que los valores del POES mínimo calculado entre todas las corridas para las matrices N°1 y N°2 son negativos (resultados claramente no físicos) y de gran magnitud en comparación al magnitud real del POES (100 MMSTB). Específicamente, para la matriz N°1 los resultados muestran un POES de -2171.27 MMSTB. Para la matriz N°2 se obtuvo POES de -17919.43 MMSTB. Otra característica importante de los resultados obtenidos es que para ambas matrices se observan que los datos se encuentran sesgados hacia valores por encima del POES real. Para la matriz N°1 el promedio obtenido fue de 510.52 MMSTB, mientras que la mediana fue de 229.77 MMSTB. Para la matriz N°2 el promedio fue de 216.27 MMSTB, y la mediana de 229.94 MMSTB. Se observa que los valores del RMSE son bastante altos indicando por tanto que se realizaron ajustes pobres en los datos.

Parámetros PVT. Caso 5		
Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	103.56	104.79
POES (mediana) [MMSTB]	102.35	105.24
POES (máximo) [MMSTB]	109.61	114.72
POES (mínimo) [MMSTB]	99.65	93.70
POES (desviación estándar) [MMSTB]	3.33	4.75
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.146	0.338
N-way ANOVA	(valor p)	
Gravedad API, °API	0.0079	0.4113
Gravedad específica del gas, γ_g	0.9381	0.0005
Relación gas petróleo en solución, R_s	0.0001	0.0014
Temperatura del yacimiento, T_y	0.8581	0.1242
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	130.68	172.06
POES (mediana) [MMSTB]	100.00	185.28
POES (máximo) [MMSTB]	232.24	573.92
POES (mínimo) [MMSTB]	-156.30	-260.64
POES (desviación estándar) [MMSTB]	72.82	140.40
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	6.469	16.513
N-way ANOVA	(valor p)	
Gravedad API, °API	0.4166	0.0007
Gravedad específica del gas, γ_g	0.7836	0.0796
Relación gas petróleo en solución, R_s	0.6133	0.0001
Temperatura del yacimiento, T_y	0.6347	0.8119

Cuadro 8.45: Resultados de la matriz de diseño de los parámetros PVT para el caso 5

Método (F-We vs. Et) Para la matriz N°1 se obtuvo una desviación estándar de 3.33 MMSTB. El análisis ANOVA indicó que los parámetros más altamente significativos fueron: la gravedad API, y la relación gas petróleo en solución. Para la matriz N°2 la desviación estándar obtenida fue de 4.75 MMSTB. El análisis ANOVA mostró que los parámetros altamente significativos fueron: la relación gas petróleo en solución, y la gravedad específica del gas. Todos los demás parámetros resultaron ser poco significativas en los resultados. Como para los casos anteriores los valores del RMSE muestran que todos los ajustes realizados son bastante buenos.

Método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) Para la matriz N°1 los resultados indican que la desviación estándar fue de 72.82 MMSTB. Para la matriz N°2 se obtuvo una desviación estándar de 140.40 MMSTB. El análisis ANOVA para la matriz N°1 indicó que ningún parámetro fue altamente significativo en los resultados. Para la matriz N°2 el análisis ANOVA señaló que la gravedad API, y la relación gas petróleo en solución fueron los parámetros más altamente significativos. También se observa en estas matrices de diseño que los valores del POES mínimo son negativos. Específicamente, para la matriz N°1 los resultados muestran un POES de -156.30 MMSTB. Para la matriz N°2 se obtuvo POES de -260.64 MMSTB. Para la matriz N°2 se nota que los datos se encuentran sesgados por encima del valor real del POES; esto es, el promedio es de 172.06 MMSTB, y la mediana de 185.28 MMSTB. Los máximos valores del RMSE indican que los ajustes realizados son aceptables, sin embargo no tienen la calidad que los ajustes del método anterior.

Caso 11

Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	98.25	-136.18
POES (mediana) [MMSTB]	100.97	-219.30
POES (máximo) [MMSTB]	176.05	137.02
POES (mínimo) [MMSTB]	27.00	-437.19
POES (desviación estándar) [MMSTB]	33.74	205.81
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	1.518	9.346
N-way ANOVA	(valor p)	
Gravedad API, °API	0.7444	0.6295
Gravedad específica del gas, γ_g	0.9247	0.2169
Relación gas petróleo en solución, R_s	0.3942	0.0003
Temperatura del yacimiento, T_y	0.4484	0
Historia de presión (valor p)	0	0
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	-115.74	107.34
POES (mediana) [MMSTB]	103.22	105.86
POES (máximo) [MMSTB]	6072.19	36959.77
POES (mínimo) [MMSTB]	-9345.69	-2872.48
POES (desviación estándar) [MMSTB]	1528.90	2444.49
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	1094.999	72341.574
N-way ANOVA	(valor p)	
Gravedad API, °API	0.1504	0.4874
Gravedad específica del gas, γ_g	0.2393	0.4064
Relación gas petróleo en solución, R_s	0.0418	0.4363
Temperatura del yacimiento, T_y	0.4715	0.3134
Historia de presión (valor p)	0	0.9794
Caso 11. $m=0.25$, $r_e D = 5$. Continúa ...		

Caso 11. $m=0.25$, $r_e D = 5$		
Método: (F/Et vs. We/Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	82.80	-37.40
POES (mediana) [MMSTB]	98.86	-51.97
POES (máximo) [MMSTB]	498.49	12757.35
POES (mínimo) [MMSTB]	-716.56	-6035.67
POES (desviación estándar) [MMSTB]	95.12	1227.13
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	2692.576	37789.809
N-way ANOVA		
Gravedad API, °API	0.8373	0.2131
Gravedad específica del gas, γ_g	0.8001	0.1452
Relación gas petróleo en solución, R_s	0.3593	0.4904
Temperatura del yacimiento, T_y	0.0308	0.3221
Historia de presión (valor p)	0.0001	0.2513

Cuadro 8.46: Caso 11. Tamaño de la capa de gas, $m = 0.25$. Relación radio del acuífero-radio del yacimiento, $r_e D = 5$

Este caso representa un yacimiento con un capa de gas moderada ($m=0.25$) y un acuífero con un $r_e D = 5$. Por esta razón se aplicaron los métodos: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw), (F/Et vs. We/Et), y el método (F-We vs. Et) para los cálculos del POES.

Método (F-We vs. Et) Para la matriz N°1 los resultados muestran una desviación estándar de 33.74 MMSTB. El análisis ANOVA señala que el parámetro más altamente significativo es la historia de presión. De igual forma, indica que la gravedad API, la gravedad específica del gas, la relación gas petróleo en solución, y la temperatura del yacimiento son poco significativas en los resultados. Para la matriz N°2 se obtuvo una desviación estándar de 205.81 MMSTB. El análisis ANOVA señala que los parámetros más altamente significativos son: la relación gas petróleo en solución, la temperatura del yacimiento, y la historia de presión. Al mismo tiempo indica que la gravedad API, la gravedad específica del gas son poco significativas en los resultados. Los valores del RMSE indican que los ajustes realizados en los datos se encuentran en un nivel aceptable.

Método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) Para la matriz de diseño N°1 la desviación estándar obtenida fue de 1528.90 MMSTB. Es de hacer notar que aunque el promedio fue negativo (-115.74 MMSTB), la mediana obtuvo un valor de 103.22 MMSTB, indicando la existencia de algún valor extremo negativo, sin embargo la distribución está centrada en el valor real del POES. El análisis ANOVA señala que los parámetros altamente significativos en los resultados fueron la relación gas petróleo en solución y la historia de presión. Para la matriz de diseño N°2 se obtuvo una desviación estándar de 2444.49 MMSTB. El análisis ANOVA indicó que ningún parámetro fue altamente significativo. También para estas estos diseños se observa que los valores del POES mínimo calculado entre todas las corridas son negativos, específicamente los resultados muestran un POES de -9245.69 MMSTB, y de -2872.48 MMSTB, para los diseños N°1 y N°2 respectivamente. Para este método los valores del RMSE son exorbitantes si se compara al método anterior, por tanto se tiene que los ajustes resultaron ser pobres.

Método (F/Et vs. We/Et) Para la matriz N°1 los resultados muestran una desviación estándar de 95.12 MMSTB. El análisis ANOVA señala que los parámetros más altamente significativos son: la temperatura del yacimiento y la historia de presión. Además muestra que la gravedad API, la gravedad específica del gas, y la relación gas petróleo en solución son poco significativos en los resultados. Para la matriz N°2 se obtuvo una desviación estándar de 1227.13 MMSTB. El análisis ANOVA señala que ninguno de los parámetros PVT resultaron altamente significativos en los resultados. Este método también presenta el mismo comportamiento del método anterior, es decir, los valores del RMSE son bastante grandes indicando que el ajuste resultante fue muy pobre.

Parámetros PVT. Caso 11		
Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	103.04	107.88
POES (mediana) [MMSTB]	101.32	105.56
POES (máximo) [MMSTB]	109.39	137.02
POES (mínimo) [MMSTB]	99.12	94.36
POES (desviación estándar) [MMSTB]	3.39	9.31
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.041	0.182
N-way ANOVA	(valor p)	
Gravedad API, °API	0	0.0002
Gravedad específica del gas, γ_g	0.992	0.0003
Relación gas petróleo en solución, R_s	0	0
Temperatura del yacimiento, T_y	0.9297	0.23
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	111.36	99.21
POES (mediana) [MMSTB]	105.69	107.10
POES (máximo) [MMSTB]	205.20	283.37
POES (mínimo) [MMSTB]	41.41	-158.21
POES (desviación estándar) [MMSTB]	22.48	70.32
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	2.198	3.759
N-way ANOVA	(valor p)	
Gravedad API, °API	0.5563	0
Gravedad específica del gas, γ_g	0.3434	0.1767
Relación gas petróleo en solución, R_s	0.0649	0
Temperatura del yacimiento, T_y	0.9922	0.6704
Caso 11. $m=0.25$, $r_e D = 5$. Continúa ...		

Parámetros PVT. Caso 11		
Método: (F/Et vs. We/Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	102.50	112.61
POES (mediana) [MMSTB]	101.27	105.31
POES (máximo) [MMSTB]	111.31	172.92
POES (mínimo) [MMSTB]	97.31	94.72
POES (desviación estándar) [MMSTB]	3.63	18.48
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	4.044	37.154
N-way ANOVA		
Gravedad API, °API	0	0
Gravedad específica del gas, γ_g	0.8496	0
Relación gas petróleo en solución, R_s	0	0
Temperatura del yacimiento, T_y	0.9967	0.3794

Cuadro 8.47: Resultados de la matriz de diseño de los parámetros PVT para el caso 11

Método (F-We vs. Et) Para la matriz N°1 se obtuvo una desviación estándar de 3.39 MMSTB. El análisis ANOVA indicó que la gravedad API, y la relación gas petróleo en solución fueron altamente significativos en los resultados. Para la matriz N°2 se obtuvo una desviación estándar de 9.31 MMSTB. El análisis ANOVA señala que los parámetros más altamente significativos fueron: la gravedad API, la gravedad específica del gas, y la relación gas petróleo en solución. Para este método los valores del RMSE muestran que los ajustes hechos en los datos son excelentes, siguen un comportamiento bastante lineal.

Método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) Para la matriz de diseño N°1 los resultados muestran una desviación estándar de 22.48 MMSTB. El análisis ANOVA señala que ningún parámetro fue altamente significativo. Para la matriz de diseño N°2 se obtuvo una desviación estándar de 70.32 MMSTB. En este caso también se observa que el valor del POES mínimo calculado entre todas las corridas para la matriz N°2 es negativo, específicamente de -158.21 MMSTB. El análisis ANOVA indica que la gravedad API, y la relación gas petróleo en solución fueron los parámetros más altamente significativos en los resultados. Los valores del RMSE indican que los datos poseen una tendencia bastante lineal de acuerdo al método aplicado.

Método (F/Et vs. We/Et) Para la matriz N°1 la desviación estándar obtenida fue de 3.63 MMSTB. El análisis ANOVA señaló que los parámetros más altamente significativos fueron: la gravedad API, y la relación gas petróleo en solución. De igual forma, también indicó que la gravedad específica del gas, la temperatura del yacimiento y la historia de presión son poco significativos en los resultados. Para la matriz N°2 se obtuvo una desviación estándar de 18.48 MMSTB. El análisis ANOVA mostró que el único parámetro PVT que no resultó altamente significativo en los resultados fue la temperatura del yacimiento. Este método también presenta resultados aceptables de acuerdo a los valores del RMSE obtenidos.

Caso 15

Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	-504.68	-1415.76
POES (mediana) [MMSTB]	98.24	-1791.24
POES (máximo) [MMSTB]	1678.04	1625.76
POES (mínimo) [MMSTB]	-3743.74	-5419.15
POES (desviación estándar) [MMSTB]	1247.95	1318.75
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	11.745	67.988
N-way ANOVA	(valor p)	
Gravedad API, °API	0.0094	0.5961
Gravedad específica del gas, γ_g	0.9173	0.4108
Relación gas petróleo en solución, R_s	0.7256	0.5135
Temperatura del yacimiento, T_y	0	0
Historia de presión (valor p)	0	0
Método: (F/Et vs. We/Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	-122.05	-7385.58
POES (mediana) [MMSTB]	-160.05	-3428.42
POES (máximo) [MMSTB]	91189.93	84743.44
POES (mínimo) [MMSTB]	-25469.47	-427427.60
POES (desviación estándar) [MMSTB]	8361.14	36376.27
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	265921.817	1278746.075
N-way ANOVA		
Gravedad API, °API	0.1744	0.9951
Gravedad específica del gas, γ_g	0.1948	0.6211
Relación gas petróleo en solución, R_s	0.9827	0.877
Temperatura del yacimiento, T_y	0.7767	0.7291
Historia de presión (valor p)	0.1092	0.0308

Cuadro 8.48: Caso 15. Tamaño de la capa de gas, $m = 0$.

Relación radio del acuífero-radio del yacimiento,

$$r_e D = \infty$$

En este caso se representa un yacimiento sin capa de gas ($m=0.0$) y un acuífero con un $r_e D = \infty$. Por esta razón se aplicó el método (F/Et vs. We/Et), junto con el método (F-We vs. Et) para los cálculos del POES.

Método (F-We vs. Et) Para la matriz N°1 la desviación estándar fue de 1247.95 MMSTB. El análisis ANOVA indicó que los parámetros más altamente significativos fueron: la gravedad API, la temperatura del yacimiento, y la historia de presión. Para la matriz N°2 se obtuvo una desviación estándar de 1318.75 MMSTB. El análisis ANOVA mostró que los parámetros altamente significativos fueron: la temperatura del yacimiento y la historia de presión, mientras que los demás parámetros resultaron poco significativos. Para la matriz N°2 los resultados se encuentran sesgados hacia valores negativos (el promedio es igual a -1415.76 MMSTB y la mediana fue de -1791.24 MMSTB). El valor del RMSE para la matriz N° 2 es bastante alto en comparación a los otros casos donde se utilizó este método.

Método (F/Et vs. We/Et) Para la matriz N°1 los resultados muestran una desviación estándar de 8361.14 MMSTB. El análisis ANOVA indicó que ninguno de los parámetros fue altamente significativo. Para la matriz N°2 se obtuvo una desviación estándar de 36376.27 MMSTB. El análisis ANOVA dió como resultado que ninguno de los parámetros fue altamente significativos en los resultados. Se observa la tendencia de que los datos se encuentran sesgados hacia valores negativos, específicamente, se obtuvo un promedio de -122.05 MMSTB, y una mediana de -160.05 MMSTB para la matriz N°1; y un promedio de -7385.58 MMSTB, y una mediana de -3428.42 MMSTB para la matriz N°2. Al igual que para ambos diseños se obtuvieron valores mínimos negativos de gran magnitud. Los valores del RMSE obtenidos muestran que los ajustes lineales del método fueron pobres.

Parámetros PVT. Caso 15		
Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	106.92	143.93
POES (mediana) [MMSTB]	101.72	108.56
POES (máximo) [MMSTB]	147.17	264.13
POES (mínimo) [MMSTB]	92.88	93.55
POES (desviación estándar) [MMSTB]	13.11	62.38
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.026	0.036
N-way ANOVA	(valor p)	
Gravedad API, °API	0	0
Gravedad específica del gas, γ_g	0.5206	0.0053
Relación gas petróleo en solución, R_s	0	0
Temperatura del yacimiento, T_y	0.9802	0.8812
Método: (F/Et vs. We/Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	108.40	145.93
POES (mediana) [MMSTB]	101.58	107.67
POES (máximo) [MMSTB]	161.74	263.54
POES (mínimo) [MMSTB]	92.39	94.10
POES (desviación estándar) [MMSTB]	16.86	63.17
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	35.018	39.077
N-way ANOVA		
Gravedad API, °API	0	0
Gravedad específica del gas, γ_g	0.4233	0.0017
Relación gas petróleo en solución, R_s	0	0
Temperatura del yacimiento, T_y	0.9835	0.852

Cuadro 8.49: Resultados de la matriz de diseño de los parámetros PVT para el caso 15

Método (F-We vs. Et) Para la matriz N°1 la desviación estándar obtenida fue de 13.11 MMSTB. El análisis ANOVA señala que los parámetros más altamente significativos son: la gravedad API, y la relación gas petróleo en solución, mientras que los demás parámetros son poco significativos. Para la matriz N°2 se obtuvo una desviación estándar de 62.38 MMSTB. El análisis ANOVA muestra que los parámetros altamente significativos son: la gravedad API, la gravedad específica del gas, y la relación gas petróleo en solución. De la misma manera indica que la temperatura del yacimiento es poco significativa en los resultados. Los valores reportados del RMSE muestran que los ajustes realizados son bastante satisfactorios.

Método (F/Et vs. We/Et) Para la matriz N°1 los resultados muestran una desviación estándar de 16.86 MMSTB. El análisis ANOVA indica que los parámetros más altamente significativos son: la gravedad API, y la relación gas petróleo en solución, aparte que los demás parámetros son poco significativos. Para la matriz N°2 se obtuvo una desviación estándar de 63.17 MMSTB. El análisis ANOVA dió como resultado que los parámetros altamente significativos fueron: la gravedad API, la gravedad específica del gas, y la relación gas petróleo en solución. Los valores del RMSE son de mucha mayor magnitud con respecto al caso anterior, sin embargo todavía son resultados aceptables.

8.2.2. Diseños Box Behnken

En esta sección se presentan para cada uno de los casos mencionados en la metodología los resultados de la matriz de diseño N°1 (con desviaciones bajas) y de la matriz de diseño N°2 (con desviaciones altas) implementando el tipo de diseño Box Behnken, junto al análisis ANOVA al final de la tabla.

Caso 1

Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	100.51	99.21
POES (mediana) [MMSTB]	100.49	100.01
POES (máximo) [MMSTB]	102.24	108.51
POES (mínimo) [MMSTB]	99.49	85.97
POES (desviación estándar) [MMSTB]	0.69	5.54
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.177	2.126
N-way ANOVA	(valor p)	
Gravedad API, °API	0.8766	0.2136
Gravedad específica del gas, γ_g	0.7302	0.3823
Relación gas petróleo en solución, R_s	0.0128	0.0166
Temperatura del yacimiento, T_y	0.2649	0.3581
Historia de presión (valor p)	0.8759	0.0515

Cuadro 8.50: Caso 1. Tamaño de la capa de gas, $m = 0$. No tiene acuífero

Para la matriz N°1 se obtuvo una desviación estándar de apenas 0.69 MMSTB. Para la matriz N°2 los resultados muestran una desviación estándar de 5.54 MMSTB. El análisis ANOVA para la matriz N°1 indica que el parámetro altamente significativo fue: la relación gas petróleo en solución; todos los demás parámetros resultaron ser poco significativos en los resultados. Para la matriz N°2 el análisis ANOVA arrojó que el único parámetro altamente significativo fue: la relación gas petróleo en solución. Para este caso los valores del RMSE muestran que los ajustes realizados en los datos son excelentes.

Parámetros PVT. Caso 1		
Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	100.53	101.49
POES (mediana) [MMSTB]	100.57	101.55
POES (máximo) [MMSTB]	101.83	105.68
POES (mínimo) [MMSTB]	99.60	98.23
POES (desviación estándar) [MMSTB]	0.56	1.91
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.053	0.195
N-way ANOVA	(valor p)	
Gravedad API, °API	0.138	0.4117
Gravedad específica del gas, γ_g	0.3691	0.044
Relación gas petróleo en solución, R_s	0.0003	0
Temperatura del yacimiento, T_y	0.9797	0.4345

Cuadro 8.51: Resultados de la matriz de diseño de los parámetros PVT para el caso 1

Para la matriz N°1 los resultados muestran una desviación estándar de 0.56 MMSTB. Para la matriz N°2 la desviación estándar obtenida fue de 1.91 MMSTB. El único parámetro altamente significativo fue, para la matriz N°1, la relación gas petróleo en solución; los demás parámetros fueron poco significativos en los resultados. Para la matriz N°2 el análisis ANOVA indicó que los parámetros altamente significativos fueron: la gravedad específica del gas, y la relación gas petróleo en solución. Los ajustes obtenidos para el caso de los parámetros PVT muestran un mucho mejor ajuste que para el caso general pudiendo decirse que son casi perfectos.

Caso 5

Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	103.96	109.07
POES (mediana) [MMSTB]	102.35	107.30
POES (máximo) [MMSTB]	113.03	130.85
POES (mínimo) [MMSTB]	96.57	93.81
POES (desviación estándar) [MMSTB]	4.00	8.45
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.386	2.754
N-way ANOVA	(valor p)	
Gravedad API, °API	0.0048	0.9691
Gravedad específica del gas, γ_g	0.8644	0.5581
Relación gas petróleo en solución, R_s	0.0007	0.0343
Temperatura del yacimiento, T_y	0.9848	0.742
Historia de presión (valor p)	0.3701	0.0118
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	186.65	-327.06
POES (mediana) [MMSTB]	103.10	217.72
POES (máximo) [MMSTB]	1141.47	5328.69
POES (mínimo) [MMSTB]	-377.49	-29423.40
POES (desviación estándar) [MMSTB]	217.79	4734.64
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	30.974	9303.257
N-way ANOVA	(valor p)	
Gravedad API, °API	0.6489	0.9858
Gravedad específica del gas, γ_g	0.8065	0.9662
Relación gas petróleo en solución, R_s	0.8276	0.9844
Temperatura del yacimiento, T_y	0.8456	0.2088
Historia de presión (valor p)	0.0086	0.1669

Cuadro 8.52: Caso 5. Tamaño de la capa de gas, $m = 1$.
No tiene acuífero

Método (F-We vs. Et) Para la matriz N°1 la desviación estándar obtenida fue de 4.00 MMSTB entre todas las corridas de la matriz. El análisis ANOVA para esta matriz de diseño indicó que los parámetros más altamente significativos fueron: la gravedad API, y la relación gas petróleo en solución; y que los parámetros con poca significancia fueron: la gravedad específica del gas, la temperatura del yacimiento y la historia de presión. Para la matriz N°2 se obtuvo una desviación estándar de 8.45 MMSTB. El análisis ANOVA demostró que los parámetros más altamente significativos fueron: la relación gas petróleo en solución, y la historia de presión. Los ajustes obtenidos para este método muestran valores del RMSE bastante bajos por lo que se consideran que los ajustes fueron muy buenos.

Método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) Para la matriz N°1 los resultados mostraron una gran dispersión con una desviación estándar de 217.79 MMSTB. El análisis ANOVA mostró que sólo un parámetro resultó altamente significativo, la historia de presión. Para la matriz N°2 la desviación estándar fue de 4734.64 MMSTB, y el análisis ANOVA indicó que ningún parámetro fue altamente significativo en los resultados. Para la matriz N°2 se observa claramente que los datos están sesgados debido a que el promedio fue de -327.06 MMSTB, y la mediana fue de 217.72 MMSTB, indicando que los datos están sesgados por encima del valor real del POES; además de que existen valores negativos de gran magnitud. Para estos diseños también se observan valores mínimos negativos. Para este método los ajustes realizados estuvieron pobres debido a que se obtuvieron valores grandes del RMSE.

Parámetros PVT. Caso 5		
Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	103.85	105.32
POES (mediana) [MMSTB]	102.35	105.74
POES (máximo) [MMSTB]	109.52	111.26
POES (mínimo) [MMSTB]	99.79	93.81
POES (desviación estándar) [MMSTB]	3.59	4.07
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.144	0.288
N-way ANOVA	(valor p)	
Gravedad API, °API	0.0338	0.2083
Gravedad específica del gas, γ_g	0.6919	0.6178
Relación gas petróleo en solución, R_s	0.0055	0.0466
Temperatura del yacimiento, T_y	0.7443	0.6834
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	133.54	177.20
POES (mediana) [MMSTB]	98.65	181.98
POES (máximo) [MMSTB]	231.98	573.92
POES (mínimo) [MMSTB]	46.69	-50.23
POES (desviación estándar) [MMSTB]	57.87	114.41
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	5.022	12.977
N-way ANOVA	(valor p)	
Gravedad API, °API	0.2815	0.1672
Gravedad específica del gas, γ_g	0.8961	0.3901
Relación gas petróleo en solución, R_s	0.0809	0.271
Temperatura del yacimiento, T_y	0.9731	0.1291

Cuadro 8.53: Resultados de la matriz de diseño de los parámetros PVT para el caso 5

Método (F-We vs. Et) Para la matriz N°1 se obtuvo una desviación estándar de 3.59 MMSTB. El análisis ANOVA para esta matriz de diseño indicó que los parámetros más altamente significativos fueron: la gravedad API, y la relación gas petróleo en solución; y que los parámetros con poca significancia fueron: la gravedad específica del gas, y la temperatura del yacimiento. Para la matriz N°2 se obtuvo una desviación estándar de 4.07 MMSTB. El análisis ANOVA mostró que el parámetro más altamente significativo fue: la relación gas petróleo en solución. También para el caso de los parámetros PVT los ajustes obtenidos aplicando este método resultaron ser excelentes.

Método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) Para la matriz N°1 los resultados mostraron una desviación estándar de 57.87 MMSTB. El análisis ANOVA mostró que ninguno de los parámetros resultó altamente significativos en los resultados. Para la matriz N°2 la desviación estándar fue de 114.41 MMSTB. También para este diseño el análisis ANOVA indicó que ningún parámetro fue altamente significativo en los resultados. Para la matriz N°2 se puede observar que los datos se encuentran sesgados por encima del valor real del POES; esto es, el promedio es de 177.20 MMSTB, y la mediana de 181.98 MMSTB. Para este método se obtuvieron valores del RMSE aceptables para los ajustes realizados.

Caso 11

Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	99.65	75.45
POES (mediana) [MMSTB]	100.64	102.96
POES (máximo) [MMSTB]	124.51	193.29
POES (mínimo) [MMSTB]	79.95	-77.76
POES (desviación estándar) [MMSTB]	8.60	68.29
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.346	3.344
N-way ANOVA	(valor p)	
Gravedad API, °API	0.8339	0.7148
Gravedad específica del gas, γ_g	0.7819	0.295
Relación gas petróleo en solución, R_s	0.0014	0.6114
Temperatura del yacimiento, T_y	0.736	0.8575
Historia de presión (valor p)	0.1167	0.0241
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	146.27	310.73
POES (mediana) [MMSTB]	105.94	111.30
POES (máximo) [MMSTB]	1292.66	22936.06
POES (mínimo) [MMSTB]	-1204.22	-9802.78
POES (desviación estándar) [MMSTB]	377.40	4235.67
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	87.070	11426.591
N-way ANOVA	(valor p)	
Gravedad API, °API	0.3517	0.9976
Gravedad específica del gas, γ_g	0.9678	0.9875
Relación gas petróleo en solución, R_s	0.7748	0.9923
Temperatura del yacimiento, T_y	0.1987	0.4383
Historia de presión (valor p)	0.7034	0.0391
Caso 11. $m=0.25$, $r_e D = 5$. Continúa ...		

Caso 11. $m=0.25$, $r_e D = 5$		
Método: (F/Et vs. We/Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	97.43	68.58
POES (mediana) [MMSTB]	100.99	101.75
POES (máximo) [MMSTB]	127.39	2703.83
POES (mínimo) [MMSTB]	58.47	-2810.71
POES (desviación estándar) [MMSTB]	13.29	632.24
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	61.508	9438.929
N-way ANOVA		
Gravedad API, °API	0.7455	0.9965
Gravedad específica del gas, γ_g	0.7203	0.0753
Relación gas petróleo en solución, R_s	0.0355	0.9705
Temperatura del yacimiento, T_y	0.719	0.9937
Historia de presión (valor p)	0.1363	0.9633

Cuadro 8.54: Caso 11. Tamaño de la capa de gas, $m=0.25$.
Relación radio del acuífero-radio del yacimiento, $r_e D = 5$

Método (F-We vs. Et) Para la matriz N°1 la desviación estándar fue de 8.60 MMSTB. El análisis ANOVA mostró que la relación gas petróleo en solución fue el único parámetro altamente significativo. Para la matriz N°2 se obtuvo una desviación estándar de 68.29 MMSTB. De la misma forma, el análisis ANOVA indicó que la historia de presión fue el parámetro más altamente significativo en los resultados. Los ajustes realizados para este método fueron bastantes buenos debido a sus valores bajos del RMSE.

Método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) Para la matriz N°1 los resultados mostraron una desviación estándar de 377.40 MMSTB. El análisis ANOVA indicó que ningún parámetro resultó ser altamente significativo. Para la matriz N°2 la desviación estándar fue de 4235.67 MMSTB. El análisis ANOVA mostró que la historia de presión fue el único parámetro altamente significativo. Claramente se observa que los valores del RMSE obtenidos fueron muy altos indicando ajustes pobres en los datos.

Método (F/Et vs. We/Et) Para la matriz N°1 los resultados arrojaron una desviación estándar de 13.29 MMSTB. Para esta matriz el análisis ANOVA señaló que la relación gas petróleo en solución fue el parámetro más altamente significativo en los resultados. Para la matriz N°2 se obtuvo una desviación estándar de 632.24 MMSTB. El análisis ANOVA mostró que ninguno de los parámetros PVT resultaron ser altamente significativos en los resultados. Los resultados en lo referente al RMSE fueron semejantes que para el método de la capa de gas.

Parámetros PVT. Caso 11		
Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	103.23	107.50
POES (mediana) [MMSTB]	101.78	107.05
POES (máximo) [MMSTB]	108.95	126.99
POES (mínimo) [MMSTB]	99.77	95.40
POES (desviación estándar) [MMSTB]	3.24	7.42
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.040	0.132
N-way ANOVA	(valor p)	
Gravedad API, °API	0.0028	0.0373
Gravedad específica del gas, γ_g	0.6533	0.2498
Relación gas petróleo en solución, R_s	0.0003	0.0001
Temperatura del yacimiento, T_y	0.6786	0.7975
Método: (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	106.13	108.22
POES (mediana) [MMSTB]	104.90	107.90
POES (máximo) [MMSTB]	134.93	283.37
POES (mínimo) [MMSTB]	47.75	-17.12
POES (desviación estándar) [MMSTB]	18.63	53.42
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	1.048	3.759
N-way ANOVA	(valor p)	
Gravedad API, °API	0.7109	0.105
Gravedad específica del gas, γ_g	0.2616	0.8712
Relación gas petróleo en solución, R_s	0.366	0.1136
Temperatura del yacimiento, T_y	0.9123	0.4357
Caso 11. $m=0.25$, $r_e D = 5$		

Caso 11. Parámetros PVT		
Método: (F/Et vs. We/Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	102.55	111.06
POES (mediana) [MMSTB]	101.73	105.42
POES (máximo) [MMSTB]	110.01	153.24
POES (mínimo) [MMSTB]	97.89	95.51
POES (desviación estándar) [MMSTB]	2.89	14.66
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	3.367	31.249
N-way ANOVA		
Gravedad API, °API	0.0022	0.014
Gravedad específica del gas, γ_g	0.7419	0.0552
Relación gas petróleo en solución, R_s	0.0003	0.0004
Temperatura del yacimiento, T_y	0.7813	0.6257

Cuadro 8.55: Resultados de la matriz de diseño de los parámetros PVT para el caso 11

Método (F-We vs. Et) Para la matriz N°1 la desviación estándar obtenida fue de 3.24 MMSTB. El análisis ANOVA mostró que los parámetros más altamente significativos fueron: la gravedad API, y la relación gas petróleo en solución. Para la matriz N°2 se obtuvo una desviación estándar de 7.42 MMSTB. Además, el análisis ANOVA indicó que los parámetros altamente significativos fueron: la gravedad API, y la relación gas petróleo en solución. Para el caso de los parámetros PVT los ajustes realizados resultaron ser mejores que para el caso general.

Método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw) Para la matriz N°1 los resultados mostraron una desviación estándar de 18.63 MMSTB. El análisis ANOVA indicó que ningún parámetro resultó ser altamente significativo. Para la matriz N°2 la desviación estándar fue de 53.42 MMSTB. Para este diseño el análisis ANOVA también mostró que ninguno de los parámetros fue altamente significativo. Los valores tan bajos del RMSE indican ajustes aceptables en los datos.

Método (F/Et vs. We/Et) Para la matriz N°1 los resultados arrojaron una desviación estándar de 2.89 MMSTB. Para esta matriz el análisis ANOVA señaló que los parámetros más altamente significativos fueron: la gravedad API, y la relación gas petróleo en solución. Para la matriz N°2 se obtuvo una desviación estándar de 14.66 MMSTB. El análisis ANOVA mostró que los parámetros PVT más altamente significativos fueron la gravedad API, y la relación gas petróleo en solución. Igual que para el método anterior los valores bajos del RMSE indican buenos ajustes en los datos.

Caso 15

Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	-58.72	-428.88
POES (mediana) [MMSTB]	100.00	100.00
POES (máximo) [MMSTB]	341.77	262.02
POES (mínimo) [MMSTB]	-997.80	-3015.55
POES (desviación estándar) [MMSTB]	306.75	939.77
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	1.681	11.342
N-way ANOVA	(valor p)	
Gravedad API, °API	0.3726	0.7748
Gravedad específica del gas, γ_g	0.9382	0.3057
Relación gas petróleo en solución, R_s	0.4607	0.9853
Temperatura del yacimiento, T_y	0.7124	0.6623
Historia de presión (valor p)	0.0031	0.0004
Método: (F/Et vs. We/Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	-56.38	-845.33
POES (mediana) [MMSTB]	100.00	104.21
POES (máximo) [MMSTB]	336.55	4918.27
POES (mínimo) [MMSTB]	-1132.33	-19866.96
POES (desviación estándar) [MMSTB]	314.92	4694.02
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	604.566	55603.821
N-way ANOVA		
Gravedad API, °API	0.3254	0.9178
Gravedad específica del gas, γ_g	0.9353	0.8688
Relación gas petróleo en solución, R_s	0.499	0.4461
Temperatura del yacimiento, T_y	0.7292	0.8235
Historia de presión (valor p)	0.0053	0.4525

Cuadro 8.56: Caso 15. Tamaño de la capa de gas, $m = 0$.
 Relación radio del acuífero-radio del yacimiento,
 $r_e D = \infty$

Método (F-We vs. Et) Para la matriz N°1 la desviación estándar fue de 306.75 MMSTB. El análisis ANOVA señaló que el parámetro más altamente significativo fue la historia de presión. Para la matriz N°2 se obtuvo una desviación estándar de 939.77 MMSTB. Al igual que para el diseño N°1 el análisis ANOVA para este diseño indicó que el parámetro más altamente significativo fue la historia de presión. Los valores del RMSE indican ajustes aceptables en los datos.

Método (F/Et vs. We/Et) Para la matriz N°1 los resultados mostraron una desviación estándar de 314.92 MMSTB. La historia de presión fue el único parámetro altamente significativo en los resultados. Para la matriz N°2 la desviación estándar fue de 4694.02 MMSTB. De la misma manera, el análisis ANOVA para esta matriz N°2 mostró que ninguno de los parámetros fue altamente significativo en los resultados. Para este método se tiene que los valores tan altos en el RMSE indican ajustes pobres en los datos.

Parámetros PVT. Caso 15		
Método: (F-We vs. Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	105.12	141.04
POES (mediana) [MMSTB]	101.84	108.48
POES (máximo) [MMSTB]	139.71	262.02
POES (mínimo) [MMSTB]	95.52	94.33
POES (desviación estándar) [MMSTB]	9.33	58.90
RMSE (mínimo) [MMSTB]	0.00	0.00
RMSE (máximo) [MMSTB]	0.026	0.035
N-way ANOVA	(valor p)	
Gravedad API, °API	0.0054	0.0102
Gravedad específica del gas, γ_g	0.9796	0.1029
Relación gas petróleo en solución, R_s	0.0054	0.0002
Temperatura del yacimiento, T_y	0.9929	0.5333
Método: (F/Et vs. We/Et)	Matriz 1	Matriz 2
POES (promedio) [MMSTB]	106.01	144.02
POES (mediana) [MMSTB]	101.64	107.45
POES (máximo) [MMSTB]	154.20	259.02
POES (mínimo) [MMSTB]	95.05	94.79
POES (desviación estándar) [MMSTB]	12.07	60.84
RMSE mínimo [MMSTB]	0.00	0.00
RMSE máximo [MMSTB]	34.674	37.443
N-way ANOVA		
Gravedad API, °API	0.0083	0.0044
Gravedad específica del gas, γ_g	0.9712	0.0774
Relación gas petróleo en solución, R_s	0.0112	0.0001
Temperatura del yacimiento, T_y	0.9869	0.4287

Cuadro 8.57: Resultados de la matriz de diseño de los parámetros PVT para el caso 15

Método (F-We vs. Et) Para la matriz N°1 la desviación estándar fue de 9.33 MMSTB. Para la matriz N°2 se obtuvo una desviación estándar de 58.90 MMSTB. Para los dos diseños el análisis ANOVA señaló que los parámetros altamente significativos fueron: la gravedad API, y la relación gas petróleo en solución; además también señaló que la gravedad específica del gas, y la temperatura del yacimiento son poco significativas en los resultados. Los resultados en cuanto al RMSE muestran que los ajustes realizados en los datos fueron casi perfectos para todas las corridas.

Método (F/Et vs. We/Et) Para la matriz N°1 los resultados mostraron una desviación estándar de 12.07 MMSTB. Para la matriz N°2 la desviación estándar fue de 60.84 MMSTB. Al igual que para el método anterior el análisis ANOVA señaló que para los dos diseños los parámetros altamente significativos fueron: la gravedad API, y la relación gas petróleo en solución. También para este caso se muestra una mejoría en el ajuste en los datos en comparación al caso general.

8.2.3. Resumen de resultados

En los cuadros de esta sección se utilizó la siguiente notación:

- Método 1 = F-We vs. Et
- Método 2 = (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw)
- Método 3 = F/Et vs. We/Et
- m = promedio
- s = desviación estándar
- √ = significativo
- × = no significativo
- FF = diseño full factorial
- BB = diseño Box Behnken

Diseños con la matriz de desviaciones bajas

Método Nº	Tipo de diseño	Estadística		N-Way ANOVA				
		m (MMSTB)	s (MMSTB)	API	GravGas	Rs	Ty	p
1	FF	101,88	2,61	×	×	√	√	√
1	BB	100,51	0,69	×	×	√	×	×
1	FF	100,62	0,67	√	×	√	×	-
1	BB	100,53	0,56	×	×	√	×	-

Cuadro 8.58: Caso 1

Método Nº	Tipo de diseño	Estadística		N-Way ANOVA				
		m (MMSTB)	s (MMSTB)	API	GravGas	Rs	Ty	p
1	FF	98,73	6,97	√	×	√	×	√
1	BB	103,96	4,00	√	×	√	×	×
1	FF	103,56	3,33	√	×	√	×	-
1	BB	103,85	3,59	√	×	√	×	-
2	FF	510,52	969,01	×	×	×	√	√
2	BB	186,65	217,79	×	×	×	×	√
2	FF	130,68	72,82	×	×	×	×	-
2	BB	133,54	57,87	×	×	×	×	-

Cuadro 8.59: Caso 5

Método N°	Tipo de diseño	Estadística		N-Way ANOVA				
		m (MMSTB)	s (MMSTB)	API	GravGas	Rs	Ty	p
1	FF	98,25	33,74	×	×	×	×	√
1	BB	99,65	8,60	×	×	√	×	×
1	FF	103,04	3,39	√	×	√	×	-
1	BB	103,23	3,24	√	×	√	×	-
2	FF	-115,74	1528,90	×	×	√	×	√
2	BB	146,27	377,40	×	×	×	×	×
2	FF	111,36	22,48	×	×	√	×	-
2	BB	106,13	18,63	×	×	×	×	-
3	FF	82,80	95,12	×	×	×	√	√
3	BB	97,43	13,29	×	×	√	×	×
3	FF	102,50	3,63	√	×	√	×	-
3	BB	102,55	2,89	√	×	√	×	-

Cuadro 8.60: Caso 11

Método N°	Tipo de diseño	Estadística		N-Way ANOVA				
		m (MMSTB)	s (MMSTB)	API	GravGas	Rs	Ty	p
1	FF	-504,68	1247,95	√	×	×	√	√
1	BB	-58,72	306,75	×	×	×	×	√
1	FF	106,92	13,11	√	×	√	×	-
1	BB	105,12	9,33	√	×	√	×	-
3	FF	-122,05	8361,14	×	×	×	×	×
3	BB	-56,38	314,92	×	×	×	×	√
3	FF	108,40	16,86	√	×	√	×	-
3	BB	106,01	12,07	√	×	√	×	-

Cuadro 8.61: Caso 15

Diseños con la matriz de desviaciones altas

Método Nº	Tipo de diseño	Estadística		N-Way ANOVA				
		m (MMSTB)	s (MMSTB)	API	GravGas	Rs	Ty	p
1	FF	77,41	33,35	×	×	✓	✓	✓
1	BB	101,63	2,31	×	×	✓	×	×
1	FF	99,21	5,54	×	✓	✓	×	-
1	BB	101,49	1,91	×	✓	✓	×	-

Cuadro 8.62: Caso 1

Método Nº	Tipo de diseño	Estadística		N-Way ANOVA				
		m (MMSTB)	s (MMSTB)	API	GravGas	Rs	Ty	p
1	FF	52,43	56,36	×	×	✓	✓	✓
1	BB	109,07	8,45	×	×	✓	×	✓
1	FF	104,79	4,75	×	✓	✓	×	-
1	BB	105,32	4,07	×	×	✓	×	-
2	FF	216,27	1293,99	×	×	×	×	×
2	BB	-327,06	4737,64	×	×	×	×	×
2	FF	172,06	140,4	✓	×	✓	×	-
2	BB	177,2	114,41	×	×	×	×	-

Cuadro 8.63: Caso 5

Método N°	Tipo de diseño	Estadística		N-Way ANOVA				
		m (MMSTB)	s (MMSTB)	API	GravGas	Rs	Ty	p
1	FF	-136,18	205,81	×	×	√	√	√
1	BB	75,45	68,29	×	×	×	×	√
1	FF	107,88	9,31	√	√	√	×	-
1	BB	107,5	7,42	√	×	√	×	-
2	FF	-136,18	205,81	×	×	×	×	√
2	BB	310,73	4235,67	×	×	×	×	√
2	FF	99,21	70,32	√	×	√	×	-
2	BB	108,22	53,42	×	×	×	×	-
3	FF	-37,4	1227,13	×	×	×	×	×
3	BB	68,58	632,24	×	×	×	×	×
3	FF	112,61	18,48	√	√	√	×	-
3	BB	111,06	14,66	√	×	√	×	-

Cuadro 8.64: Caso 11

Método N°	Tipo de diseño	Estadística		N-Way ANOVA				
		m (MMSTB)	s (MMSTB)	API	GravGas	Rs	Ty	p
1	FF	-1415,76	1318,75	×	×	×	√	√
1	BB	-428,88	939,77	×	×	×	×	√
1	FF	143,93	62,38	√	√	√	×	-
1	BB	141,04	58,9	√	×	√	×	-
3	FF	-7385,58	36376,27	×	×	×	×	√
3	BB	-7385,08	4694,02	×	×	×	×	×
3	FF	145,93	63,17	√	√	√	×	-
3	BB	144,02	60,84	√	×	√	×	-

Cuadro 8.65: Caso 15

8.2.4. Comparación de los diseños full factorial y Box Behnken

En general los dos diseños dieron resultados bastantes similares (en lo referente a las medidas de dispersión y tendencia central), a excepción de algunas diferencias (en cuanto al análisis ANOVA). Estas diferencias se deben a la estructura de la matriz de diseño de Box Behnken que no toma en cuenta algunas interacciones entre los parámetros. Sin embargo, este tipo de diseño logra capturar gran cantidad de información con un número de corridas 6 veces menor que el diseño full factorial (sólo para este diseño de 5 factores). En resumen el diseño Box Behnken resulta ser más “óptimo” a la hora de obtener información con el menor número de corridas posibles.

En lo referente a la tendencia general de los análisis ANOVA a continuación se clasifican los resultados de los diseños donde se variaron todos los parámetros PVT y la historia de presión, y posteriormente aquellos diseños donde sólo se variaron los parámetros PVT. La importancia de los parámetros PVT y la historia de presión, tomando en cuenta todos los casos y métodos es la siguiente:

1. historia de presión (P vs. T)
2. relación gas petróleo en solución (R_s)
3. temperatura del yacimiento (T_y)
4. gravedad API ($^{\circ}$ API)
5. gravedad específica del gas (γ_g).

Sin embargo, cuando se hace la clasificación tomando aquellos diseños donde sólo actuaron los parámetros PVT, se obtuvo el orden mostrado a continuación. Esto se puede deber a que los diseños Box Behnken omiten algunas corridas, al parecer de gran importancia, por lo tanto fallan para tomar en cuenta las desviaciones de ciertos parámetros en la respuesta, a diferencia del diseño full factorial.

1. relación gas petróleo en solución (R_s)
2. gravedad API ($^{\circ}$ API)
3. gravedad específica del gas (γ_g)
4. temperatura del yacimiento (T_y).

Realizando un análisis en conjunto de todos los resultados se tiene que los parámetros más significativos fueron, en primer lugar la historia de presión del yacimiento, tomando en cuenta desviaciones máximas de 200 psi con respecto al valor real de presión (recordemos que $\sigma = 100$ psi para los cálculos de las realizaciones). El siguiente parámetro más significativo fue la relación gas petróleo en solución con una desviación máxima de 29.52 SCF/STB, apenas un 2% del valor de la relación gas petróleo en solución (1476 SCF/STB). Para los diseños de los parámetros PVT el siguiente parámetro importante (debido a que resultó ser altamente significativos para la mayoría casos analizados) es la gravedad API, con desviaciones máximas de 0.5 grados API. Aunque en los diseños donde se toma en cuenta la historia de presión, la temperatura del yacimiento parece ser más importante que la gravedad API (debido a que aparece en un mayor número de casos como altamente significativo), sin embargo, para los diseños de los parámetros PVT, la temperatura del yacimiento no resultó ser significativo para ningún caso. La desviación máxima para la temperatura con respecto al valor real fue de 0.9 °F. En cuanto a la gravedad específica del gas fue poco significativa para ambos tipos de diseños (tomando en cuenta la historia de presión y sólo los parámetros PVT), con desviaciones máximas de 0.0123 unidades de gravedad específica.

En cuanto a los métodos de balance de materiales se observó la tendencia general de que la desviación estándar de los POES calculados aumentaban de acuerdo al siguiente orden de los modelos de yacimientos utilizados:

1. Sin capa de gas, ni influjo de agua (menor error).
2. Tamaño grande de la capa de gas ($m = 1$), sin influjo de agua.
3. Empuje combinado de una capa de gas moderada ($m = 0.25$), y un influjo de agua moderado ($r_e D = 5$).
4. Sin capa de gas y con una gran influjo de agua, $r_e D = \infty$ (mayor error).

En el cuadro 8.66 aparecen los promedios de la desviación estándar para todos los casos con el mismo rango de incertidumbre, es decir, se promedió la desviación estándar resultado del diseño full factorial, con la desviación estándar resultado del diseño Box Behnken para la matriz con desviaciones bajas y luego se promedió las desviaciones estándar de las matrices con desviaciones altas. Esto respecto al tipo de diseño que toma en cuenta la historia de presión.

En base a este resultado se puede decir que el método (F-We vs. Et) es el más robusto frente a la incertidumbre en los datos, luego le sigue el método (F-We)/(Eo+Efw) vs. (Eg+Efw)/(Eo+Efw), y por último se tiene el método (F/Et vs. We/Et).

Ahora se presenta un cuadro similar al anterior, sólo que éste se obtuvo en base a los diseños que involucraron sólo los parámetros PVT.

Método de balance de materiales	Método 1	Método 2	Método 3
Promedio de la std(POES) para la matriz con desviaciones bajas (MMSTB)	201.41	773.28	2196.12
Promedio de la std(POES) para la matriz con desviaciones altas (MMSTB)	329.67	3177.20	10732.42
Método 1: F-We vs. Et			
Método 2: $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$			
Método 3: F/Et vs. We/Et			

Cuadro 8.66: Valores de la desviación estándar del POES para los métodos de balance de materiales. Los diseños toman en cuenta la historia de presión

De acuerdo a estos resultados el método más robusto fue (F-We vs. Et). El siguiente método resultó ser (F/Et vs. We/Et), y por último se tiene el método $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$. En resumen, el método (F-We vs. Et) es el más robusto frente a incertidumbres en los datos y a todos los modelos de yacimientos evaluados. El método $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$ muestra hipersensibilidad a todos los parámetros. El método (F/Et vs. We/Et) muestra hipersensibilidad sobre todo a la historia de presión, frente a incertidumbre en otros parámetros se muestra más exacto en los resultados. Si se hace una comparación entre los cuadros 8.66 y 8.67 se aprecia la importancia de la incertidumbre en la presión, y porqué es el parámetro con más influencia en los resultados. En lo referente a la calidad del ajuste se tiene que los mayores valores del RMSE se obtuvieron para el caso 15 con la matriz de diseño con las desviaciones más altas para todos los parámetros evaluados. Al igual que para el análisis de las realizaciones, se tuvo que la linealidad del ajuste no fue un parámetro mediante el cual se pueda evaluar la veracidad del POES obtenido, dado que se presentaron casos donde la tendencia lineal era casi perfecta, y no obstante, el valor del POES obtenido del ajuste no era el correcto. Por tanto, y al igual que para el análisis de las realizaciones, no se puede utilizar la linealidad de los puntos en los gráficos de los métodos de balance de materiales evaluados como una medida de la calidad de los cálculos de balance de materiales.

Método de balance de materiales	Método 1	Método 2	Método 3
Promedio de la std(POES) para la matriz con desviaciones bajas (MMSTB)	4.65	42.95	8.86
Promedio de la std(POES) para la matriz con desviaciones altas (MMSTB)	18.88	94.64	39.29
Método 1: F-We vs. Et			
Método 2: $(F-We)/(Eo+Efw)$ vs. $(Eg+Efw)/(Eo+Efw)$			
Método 3: F/Et vs. We/Et			

Cuadro 8.67: Valores de la desviación estándar del POES para los métodos de balance de materiales. Los diseños sólo involucran los parámetros PVT

Conclusión

En base a los resultados obtenidos podemos llegar a las siguientes conclusiones:

- El análisis estadístico de la incertidumbre en los datos de presión y PVT indicó que las estimaciones de los hidrocarburos en sitio son sensibles a: el método de balance materiales utilizado, errores en los datos de presión y PVT, y al mecanismo de empuje que posea el yacimiento.
- El método $(F-W_e \text{ vs. } E_t)$ demostró ser el más robusto frente a la incertidumbre en los datos de presión y PVT.
- El método $(F-W_e)/(E_o+E_{fw}) \text{ vs. } (E_g+E_{fw})/(E_o+E_{fw})$ mostró hipersensibilidad a todos los rangos de incertidumbre en los datos de presión y PVT.
- El método $(F/E_t \text{ vs. } W_e/E_t)$ demostró hipersensibilidad a la incertidumbre en los datos de presión. Frente a la incertidumbre en los datos PVT mostró una sensibilidad moderada.
- Los errores en la presión se identificaron como la fuente principal de incertidumbre en las estimaciones de balance de materiales.
- Los errores en la relación gas petróleo en solución se identificaron como la principal fuente de incertidumbre proveniente de los datos PVT, en las estimaciones de balance de materiales.
- La gravedad API del crudo y la temperatura del yacimiento son fuentes importantes de incertidumbre en las estimaciones de balance de materiales.
- La gravedad específica del gas resultó ser el parámetro con menos influencia en las estimaciones de balance de materiales.
- El modelo de yacimiento donde el principal mecanismo de empuje fue expansión del petróleo más su gas en solución presentó las estimaciones más exactas de balance de materiales .

- El modelo de yacimiento donde el principal mecanismo de empuje fue expansión de la capa de gas presentó una exactitud moderada en las estimaciones de balance de materiales.
- El modelo de yacimiento donde el principal mecanismo de empuje fue por influjo de agua presentó el caso con menos exactitud en las estimaciones de balance de materiales.
- Para el caso en que se tenga un modelo de yacimiento donde el principal mecanismo de empuje sea la expansión del petróleo más su gas disuelto, exista poca incertidumbre en los datos de presión y PVT, y además se utilice el método $(F-W_e$ vs. $E_t)$ para los cálculos de balance de materiales, representa el caso con mayor exactitud esperada en las estimaciones de balance de materiales.
- Para el caso en que se tenga un modelo de yacimiento donde el principal mecanismo de empuje sea por influjo de agua, exista una gran incertidumbre en los datos de presión, y se utilice el método $(F-W_e)/(E_o+E_{fw})$ vs. $(E_g+E_{fw})/(E_o+E_{fw})$ ó el método $(F/E_t$ vs. $W_e/E_t)$ para los cálculos de balance de materiales, representa el caso donde se espera la menor exactitud en las estimaciones de balance de materiales.
- La metodología de diseño experimental en conjunto con el análisis ANOVA permite identificar eficazmente las principales fuentes de incertidumbre en las estimaciones de balance de materiales.
- Los diseños Box Behnken logran reducir el esfuerzo computacional proporcionando una exactitud comparable al diseño full factorial, debido a que obtienen la mayor cantidad de información con el menor número de corridas posibles.
- La implementación del programa fue requerida para la evaluación y análisis estadístico de los rangos de incertidumbre de la presión y PVT en los cálculos de balance de materiales.

Recomendaciones

- Evaluar los rangos de incertidumbre de presión y PVT en otros métodos de balance de materiales (como la regresión planar en la ecuación de balance de materiales, ó el método de Tehrani) para la determinación de los hidrocarburos en sitio.
- Evaluar la incertidumbre proveniente de los volúmenes de fluidos producidos (petróleo, agua y gas) en las estimaciones de balance de materiales.
- Evaluar modelos de yacimiento donde existan grandes capas de gas junto con grandes volúmenes de influjo de agua.
- Adaptar el programa para llevar a cabo las recomendaciones mencionadas anteriormente.

Bibliografía

- [1] C. R. McEwen. *Material balance calculations with water influx in the presence of uncertainty in pressures*. 32nd Annual California Regional Meeting of SPE. SPE 225., Páginas 120–128, 1961.
- [2] D. H. Tehrani. *An analysis of volumetric balance equation for calculation of oil-in-place and water influx*. Journal of Petroleum Technology. SPE 5990., Páginas 1–12, 1976.
- [3] B. Wang y R. R. Hwan. *Influence of reservoir drive mechanism on uncertainties of material balance calculations*. 1997 SPE Annual Technical Conference and Exhibition held in San Antonio, Texas. SPE 38918., Páginas 767–778, 1997.
- [4] Mark P. Walsh. *Effect of pressure uncertainty on material balance plots*. 1999 SPE Annual Technical Conference and Exhibition held in Houston, Texas. SPE 56691., Páginas 1–11, 1999.
- [5] R. O. Baker; C. Regier; y R. Sinclair. *PVT error analysis for material balance calculations*. The Petroleum Society's Canadian International Petroleum Conference. Paper 2003-203., 2003.
- [6] Omole O.; Ojo K. P. *A new method for estimating oil in place and gas cap size using the material-balance equation*. Journal of Petroleum Technology. SPE 026266., Páginas 6–11, 1993.
- [7] D. Havlena y A.S. Odeh. *The material balance as an equation of a straight line*. SPE Production Research Symposium, Norman, Okla. SPE 559., 1963.
- [8] L.P. Dake. *Fundamentals of Reservoir Engineering*. Elsevier Science B.V., The Netherlands, 1978. pp. 73-77.
- [9] M. Essenfeld y E. Barberii. *Yacimientos de Hidrocarburos*. FONCIED Fondo Editorial del Centro Internacional de Educación y Desarrollo, Caracas, 2001. pp. 141-148, 171-176.

- [10] Phillip L. Moses. *Engineering applications of phase behavior of crude oil and condensate systems*. Journal of Petroleum Technology. SPE 15835., Páginas 715–723, July 1986.
- [11] I. S. Agbon; G. J. Aldana; J. C. Araque; A. A. Mendoza; M. E. Ramirez. *Resolving uncertainties in historical data and the redevelopment of mature fields*. SPE Latin America and Caribbean Petroleum Engineering Conference held in Port-of-Spain, Trinidad, West Indies. SPE 81101., Páginas 1–6, 2003.
- [12] Juan Jones-Parra. *Elementos de Ingeniería de Yacimientos*. EDIT Ediciones Innovación Tecnológica, Caracas, 1989. pp. 3.2-3.4.
- [13] M. Sheldon Ross. *Probabilidad Y Estadística Para Ingenieros*. Segunda edición. Mc Graw-Hill Interamericana Editores, S. A. de C. V. Mexico, 2000.
- [14] Richard A. Johnson. *Probabilidad Y Estadística Para Ingenieros de Miller Y Freud*. Quinta edición. Prentice-Hall Hispano-americana, Mexico, 1997.
- [15] William W. Hines y Douglas C. Montgomery. *Probabilidad Y Estadística Para Ingeniería*. CECSA Continental, Mexico D.F., Mexico, 1998.
- [16] Larousse. *Diccionario Enciclopédico*. Larousse, S. A., 1999. pp. 418.
- [17] Statsoft. *Experimental design (industrial DoE)*. <http://www.statsoftinc.com/textbook/stanman.html>.
- [18] <http://www.itl.nist.gov/div898/handbook/pri/section3/>.
- [19] Statsoft. *ANOVA/MANOVA*. <http://www.statsoftinc.com/textbook/stexdes.html>.
- [20] The Mathworks Inc. *Statistics Toolbox, User's Guide.*, version 4.0 edition, 2002. pp. 10-2,10-10.
- [21] W. B. Fair Jr. *A statistical approach to material balance methods*. SPE 69th Annual Technical Conference and Exhibition held in New Orleans, LA, USA. SPE 28629., 1994.
- [22] M. R. Carlson. *Tips, tricks and traps for oil material balance calculations*. 46th Annual Technical Meeting of The Petroleum Society of CIM in Banff, Alberta, Canada. Paper 95-07., 1995.
- [23] The MathWorks Inc. *Data Acquisition Toolbox User's Guide*, version 2.2 edition, 2002.
- [24] Baker Hughes. *Inflatable systems, electronic gauges*. <http://www.bakerhughes.com>.

- [25] Halliburton. *Reservoir description tool (RDTTM)*. <http://is.myhalliburton.com>.
- [26] Halliburton. *Sequential formation test tool (SFTTTM)*. <http://is.myhalliburton.com>.
- [27] The American Society for Testing y Materials. *Standard test method for api gravity of crude petroleum and petroleum products*. Annual Book of ASTM Standards. Designation D 287 - 82., Páginas 125 – 127, 1987.
- [28] Solartron Mobrey. *Solartron, gas density and specific gravity products*, 2004. Data sheet B1253. <http://www.solartronusa.com>.
- [29] Yokogawa. *EXA GD402 gas density analyzer*, 2001. Bulletin 11T3E1-E-A. <http://www.yca.com>.
- [30] Stephen Prensky. *Temperature measurements in boreholes: An overview of engineering and scientific applications*. <http://www.sprensky.com/index.html>.
- [31] *Unexpected increases in water production*. <http://www.pttc.org>.
- [32] Baker Hughes. *Measurement-while-drilling (MWD)*. <http://www.bakerhughes.com>.
- [33] Muhammad A. Al-Marhoun. *Adjustment of differential liberation data to separator conditions*. 2001 SPE Middle East Oil Show held in Bahrain. SPE 68234., Páginas 1–7, 2001.
- [34] William D. McCain, Jr. *Analysis of black oil pvt reports revisited*. SPE Annual Technical Conference and Exhibition held in San Antonio, Texas. SPE 77386., Páginas 1–5, 2002.
- [35] Tor Bu y Elvind Damsleth. *Errors and uncertainties in reservoir performance predictions*. 1995 SPE Annual Technical Conference and Exhibition held in Dallas. SPE 30604., Páginas 194–200, 1996.
- [36] Schlumberger. *Well evaluation conference - venezuela*, 1997.
- [37] Elementary Concepts in Statistics. *Are all test statistics normally distributed?* <http://www.statsoft.com/textbook/esc.html>.
- [38] Elementary Concepts in Statistics. *Why the normal distribution is important*. <http://www.statsoft.com/textbook/esc.html>.
- [39] Basic Statistics. *Significance of correlations*. <http://www.statsoft.com/textbook/stbasic.html>.