



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

PROGRAMA DE POSGRADO EN CIENCIA E INGENIERÍA DE
LA COMPUTACIÓN

SOLUCIÓN A PROBLEMAS
INVERSOS DE ESTIMACIÓN DE
PARÁMETROS MEDIANTE EL USO
DE TÉCNICAS ESTADÍSTICAS Y
MATEMÁTICA COMPUTACIONAL,
APLICADAS A LA
CARACTERIZACIÓN DE
YACIMIENTOS DE AGUA Y
PETRÓLEO

T E S I S

QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS

P R E S E N T A:
MINUTTI MARTÍNEZ CARLOS

DIRECTORA DE TESIS:
DRA. SUSANA GÓMEZ GÓMEZ

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Solución a problemas inversos de estimación de
parámetros mediante el uso de técnicas estadísticas
y matemática computacional, aplicadas a la
caracterización de yacimientos de agua y petróleo

Minutti Martínez Carlos

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas
Universidad Nacional Autónoma de México

cminutti@gmail.com

Enero, 2020

Índice general

Introducción	VII
I Caracterización de yacimientos de agua	1
1. Tomografía hidráulica	2
1.1. Descripción del problema	3
1.2. El algoritmo de inversión GM	7
1.2.1. Parametrización del campo de conductividad	7
1.2.2. Modelo de Mezclas Gaussianas	8
1.2.3. Función objetivo	12
1.2.4. Regularización	13
1.2.5. Información geológica	15
1.2.6. Relación con modelos geológicos y geoestadísticos	16
1.2.7. Implementación	18
1.2.8. Campo de conductividad de Bayes	20
1.3. Metodología y Experimentos	21
1.4. Resultados	24
1.4.1. Caso no Gaussiano / Acuífero fracturado	35
1.4.2. Caso real / Caja de arena	37
1.5. Resumen y conclusiones	40

II Caracterización de yacimientos de petróleo 43

2. Optimización ad-hoc para el ajuste de curvas tipo mediante un modelo de flujo de triple porosidad y geometría fractal 44

2.1. Descripción del problema	45
2.2. El modelo de flujo de triple porosidad y geometría fractal	48
2.2.1. El problema de la identificación de parámetros	49
2.3. Análisis de sensibilidad	52
2.3.1. Sensibilidad: parámetro-tiempo	53
2.3.2. Sensibilidad: parámetro-valor	57
2.3.3. Conclusiones del análisis de sensibilidad	61
2.4. Estimación de los parámetros	63
2.4.1. Error de estimación	64
2.4.2. Resultados	64
2.4.3. Dependencia entre parámetros y el error de estimación	70
2.5. Fallas y almacenamiento variable	72
2.5.1. Análisis de sensibilidad	72
2.5.2. Estimación de los parámetros	76
2.6. Casos reales	81
2.6.1. Pozo F1	81
2.6.2. Pozo F1B	84
2.6.3. Pozo F2	84
2.6.4. Pozo F3	88

3. Reducción del ruido en los datos mediante un enfoque de *machine-learning* 90

3.1. Descripción del problema	91
3.1.1. Metodología	92
3.1.2. Diferencias finitas de varios pasos	93
3.1.3. Splines	94
3.1.4. Mezcla de curvas sub-óptimas	95
3.1.5. Proceso de entrenamiento y comparativo de los métodos	96
3.1.6. Error de los parámetros del modelo del problema directo	97
3.2. Resultados	97
3.3. Conclusiones	104

<i>ÍNDICE GENERAL</i>	I
Conclusiones	105
Bibliografía	116

Índice de figuras

1.1.	(a) Imagen a ser replicada; (b)-(f) Representación de la imagen a través de un modelo de mezcla Gaussiana con diferentes componentes. ($N_k = \{3, 6, 12, 24, 48\}$). MAE y MSE se presentan cómo medidas de precisión.	10
1.2.	Funciones objetivo y gráficos de contorno para los valores de K de las dos capas de material cuando (a) se utiliza el error cuadrático de los valores de la presión y (b) cuando se utiliza el error cuadrático de las derivadas espaciales. La ubicación del valor óptimo está marcada con una línea sólida y una cruz.	14
1.3.	Campo esquemático de $\ln-K$ simulado con el generador de campo aleatorio de K . Las cruces azules marcan la ubicación de los pozos de bombeo para la calibración, las rojas la ubicación de los pozos de bombeo utilizados para la validación y los puntos negros los pozos de observación. [1]-[4] son los puertos de bombeo para experimentos que incluyen de 1 a 4 pruebas de bombeo en el proceso de inversión.	22
1.4.	Los campos K verdaderos y los tomogramas $\ln-K$ obtenidos con VSAFT2 y GM cómo métodos de inversión para los 6 casos con mayor varianza de $\ln-K$, utilizando 4 pruebas de bombeo con 48 valores de altura por prueba.	25
1.5.	Los campos K verdaderos y los tomogramas $\ln-K$ obtenidos con VSAFT2 y GM cómo métodos de inversión para los 6 casos con menor varianza de $\ln-K$, utilizando 4 pruebas de bombeo con 48 valores de altura por prueba.	26

1.6. Los campos K verdaderos y los tomogramas $\ln-K$ obtenidos con VSAFT2 y GM cómo métodos de inversión de una muestra aleatoria de 6 casos, utilizando 4 pruebas de bombeo con 48 valores de altura por prueba. 27

1.7. Correlograma de los errores de estimación de K (GM_K_Err, VSAFT_K_Err), Errores para los *drawdowns* en el conjunto de calibración (GM_Cal_Err, VSAFT_Cal_Err), Errores para los *drawdowns* en el conjunto de validación (GM_Val_Err, VSAFT_Val_Err), Varianza, Media y Rango de K (Var_K, Mean_K, Range_K), *Drawdown* promedio en el conjunto de calibración (Mean_Drawdown), Número de pruebas de bombeo (N_Tests) y Tiempo de cómputo (GM_Time, VSAFT_Time) de los 1,000 experimentos, para el algoritmo GM y VSAFT2. 28

1.8. Error de estimación promedio de K (E_K) cuando GM y VSAFT2 son usados cómo métodos de inversión, para 1-4 pruebas de bombeo (n_{test}). 29

1.9. Box-plots de los errores estimados de K (E_K) para los 250 campos sintéticos de K , cuando GM y VSAFT2 se utilizan cómo métodos de inversión, para 1-4 pruebas de bombeo (n_{tests}). La línea horizontal discontinua representa el error de un K efectivo cómo campo K estimado. 30

1.10. Eficiencia media del algoritmo (reducción de error por segundo de tiempo de cómputo) cuando GM y VSAFT2 se utilizan cómo métodos de inversión, para 1-4 pruebas de bombeo (n_{tests}). 31

1.11. Box-plots de los errores de *drawdowns* de validación para los 250 campos K sintéticos, cuando GM y VSAFT2 se utilizan cómo métodos de inversión, para 1-4 pruebas de bombeo (n_{tests}). La línea horizontal discontinua representa el error de un K efectivo cómo campo K estimado. 32

1.12. Box-plots de los tiempos de cómputo para los 250 campos sintéticos de K , cuando GM y VSAFT2 se utilizan cómo métodos de inversión, para 1-4 pruebas de bombeo (n_{tests}). 33

1.13. Gráficas de dispersión de los <i>drawdowns</i> simulados vs observados para el conjunto de datos de calibración de los 250 experimentos, cuando el campo K se estima usando (a) el K efectivo, (b) VSAFT2 y (c) GM, con 4 pruebas de bombeo y 48 valores de presión por prueba. La línea de 45 grados representa una coincidencia perfecta. La línea sólida corresponde a una regresión lineal, mostrando el intercepto y el coeficiente de determinación (R^2).	33
1.14. Gráficas de dispersión de los <i>drawdowns</i> simulados vs observados para el conjunto de datos de validación de los 250 experimentos, cuando el campo K se estima usando (a) el K efectivo, (b) VSAFT2 y (c) GM, con 4 pruebas de bombeo y 48 valores de presión por prueba. La línea de 45 grados representa una coincidencia perfecta. La línea sólida corresponde a una regresión lineal, mostrando el intercepto y el coeficiente de determinación (R^2).	34
1.15. (a) Campo K sintético de un acuífero fracturado y los campos K estimados con (b) VSAFT2 y (c) GM cómo métodos de inversión, utilizando 10 pruebas de bombeo con 48 valores de presión por prueba.	36
1.16. Campos K estimados para el experimento de caja de arena, obtenidos con (a) VSAFT2 y (b) GM cómo métodos de inversión, utilizando 8 pruebas de inyección con 47 valores de presión por prueba. Imagen superpuesta de los tomogramas y la fotografía del arenero (g)-(h). La línea de puntos de 45 grados representa una coincidencia perfecta. . .	39
2.1. Curvas tipo: comparación con la modificación de un parámetro (ω_v). . .	50
2.2. Función objetivo sin pesos y con pesos optimizados.	51
2.3. Correlación de ω_f y ω_v con la presión, en diferentes tiempos.	53
2.4. Correlación de λ_{mf} , λ_{mv} y λ_{vf} con la presión, en diferentes tiempos. . .	54
2.5. Correlación de dL y Δ con la presión, en diferentes tiempos.	56
2.6. Correlación de los parámetros de pozo con la presión, en diferentes tiempos.	56
2.7. Correlación de los parámetros de yacimiento con la presión, en diferentes tiempos, cuando no existe almacenamiento y cuando sí existe. . .	58
2.8. Efecto promedio de ω_f y ω_v en la presión.	59
2.9. Efecto promedio de λ_{mf} , λ_{mv} y λ_{vf} en la presión.	60
2.10. Efecto promedio de dL y Δ en la presión.	61

2.11. Efecto promedio de s y C_D en la presión.	62
2.12. Dispersión de los valores a estimar de ω 's vs estimación.	66
2.13. Dispersión de los valores a estimar de λ 's vs estimación.	67
2.14. Dispersión de los valores a estimar de dL y Δ vs estimación.	68
2.15. Dispersión de los valores a estimar de s y C_D vs estimación.	69
2.16. Correlación parcial (de Spearman) entre errores de estimación y valores de los parámetros.	71
2.17. Correlación de L_D con la presión, en diferentes tiempos.	73
2.18. Correlación de $L1_D$, $L2_D$ y n_α con la presión, en diferentes tiempos.	75
2.19. Correlación de C_D , $C_{\phi D}$ y $\alpha_{\phi D}$ con la presión, en diferentes tiempos.	75
2.20. Efecto promedio de L_D en la presión, para una fallas a presión constante y fallas sellantes.	76
2.21. Dispersión de los valores a estimar de Δ y L_D vs estimación.	78
2.22. Dispersión de los valores a estimar de $C_{\phi D}$ y $\alpha_{\phi D}$ vs estimación.	79
2.23. Correlación parcial (de Spearman) entre errores de estimación y valores de los parámetros, incluyendo fallas y almacenamiento variable.	80
2.24. Ajuste para el pozo F1 (datos eliminados debido al tiempo de producción) - Método <i>ad-hoc</i>	82
2.25. Ajuste para el pozo F1 (sin eliminar datos) - Método <i>ad-hoc</i>	85
2.26. Ajuste para el pozo F2 - Método <i>ad-hoc</i>	86
2.27. Ajuste para el pozo F3 - Método <i>ad-hoc</i>	89
3.1. Boxplot de $\log_{10}(Error)$ para los algoritmos entrenados.	99
3.2. Histograma de $\log_{10}(Error)$ para los algoritmos entrenados.	100
3.3. Ejemplo 1: Reducción de ruido con los tres mejores métodos.	101
3.4. Ejemplo 2: Reducción de ruido con los tres mejores métodos.	102
3.5. Ejemplo 3: Reducción de ruido con los tres mejores métodos.	103

Índice de cuadros

1.1. Mediana de los errores de K y de <i>drawdowns</i> para el problema de TH, usando 250 casos sintéticos.	35
1.2. Tiempo de cómputo y diferentes medidas de error para el problema de TH, en un acuífero sintético fracturado ($\sigma_{\ln K}^2 = 0.3271$).	37
1.3. Tiempo de cómputo y diferentes medidas de error para el problema de TH en un experimento de caja de arena.	38
2.1. Media y mediana del error de estimación.	64
2.2. Media y mediana del error de estimación, incluyendo fallas y almacenamiento variable.	77
2.3. Parámetros para el pozo F1 (datos eliminados debido al tiempo de producción) - Método <i>ad-hoc</i>	83
2.4. Parámetros para el pozo F1 (sin eliminar datos) - Método <i>ad-hoc</i>	84
2.5. Parámetros para el pozo F2 - Método <i>ad-hoc</i>	87
2.6. Parámetros para el pozo F3 - Método <i>ad-hoc</i>	88
3.1. Media y mediana de $\log_{10}(Error)$ para los algoritmos probados.	98
3.2. Error medio de estimación de los parámetros del modelo con diferentes métodos de reducción de ruido.	102

Introducción

Los problemas inversos de estimación de parámetros, tienen gran importancia y aplicabilidad en las diversas áreas de ciencias, debido a que permiten obtener información de parámetros de los modelos que simulan la realidad física y que no podemos observar de manera directa. Algunos ejemplos de estos problemas incluyen la tomografía computarizada, mejoramiento de imágenes, diferenciación numérica y reconstrucción de fuentes acústicas. La solución a estos problemas está ligada principalmente al área de cómputo científico y matemáticas aplicadas.

Sin embargo, en muchos problemas inversos la precisión de las estimaciones y el tiempo de cómputo requerido pueden verse afectados de manera importante debido a múltiples factores, cómo son, la existencia de óptimos locales, la escasa sensibilidad de los modelos a los parámetros, ruido, escasez de datos y un alto número de parámetros a ser estimados.

En este trabajo se presenta una metodología que permite abordar problemas inversos de alta complejidad, mediante el uso de diferentes técnicas usadas en estadística y en cómputo científico (ciencia de datos), las cuales permitan entender el problema de forma más eficiente, a escala de datos, para así re-enfocar el proceso de solución del problema.

Para mostrar la eficiencia de esta metodología, se han seleccionado dos problemas inversos complejos, que presentan estas características: un problema inverso del área de caracterización de yacimientos de agua y otro en la caracterización de yacimientos de petróleo. Se dan nuevos enfoques a la solución de estos problemas mediante el uso de diversos métodos estadísticos y computacionales, los cuales son adaptables a otros problemas inversos.

Aunque los problemas propuestos pueden parecer similares, se resuelven usando enfoques diferentes.

El enfoque del primer caso, caracterización de yacimientos de agua, consiste en re-parametrizar el problema, modificando así el espacio de parámetros a estimar, dando un nuevo significado a los parámetros y a su interpretación, por lo que se realiza un cambio de paradigma del problema.

En el segundo caso, caracterización de yacimientos de petróleo, se realiza un estudio de la sensibilidad que tiene la presión (que es la solución del modelo matemático que representa el problema a resolver), a los distintos parámetros, a nivel global e incluso temporal, para así crear un método de optimización *ad-hoc* al problema, el cual no requiere un cambio de paradigma, pero mejora significativamente la precisión en la estimación de los parámetros y los tiempos de cómputo.

En ambos problemas inversos se pone a prueba la metodología desarrollada mediante la solución de casos reales.

Adicionalmente, ya que estos problemas inversos necesitan información medida sobre el fenómeno de estudio, es decir datos reales, se aborda el problema de reducción de ruido, que a su vez, es transformado en un problema inverso, que se resuelve usando técnicas de *machine-learning*. El problema de reducción de ruido es esencial para la adecuada estimación de los parámetros, en especial en el segundo problema, que utiliza la denominada log-derivada de los datos de presión.

Parte I

Caracterización de yacimientos de agua

Capítulo 1

Tomografía hidráulica mediante un algoritmo de mezclas Gaussianas

La Tomografía Hidráulica (TH) se ha convertido en uno de los métodos más robustos para caracterizar la heterogeneidad de los parámetros hidráulicos tales como la conductividad hidráulica. Debido a que se requieren varias pruebas de bombeo/inyección con un gran número de mediciones, la carga computacional para el análisis de TH puede ser grande. En esta sección se presenta un nuevo algoritmo inverso para modelar la heterogeneidad de acuíferos, el cual está basado en un Modelo de Mezclas Gaussianas (MMG) que reduce significativamente el número de parámetros a ser estimados durante el proceso de inversión. Además, se introduce una nueva función objetivo que aumenta la sensibilidad del modelo usado a los parámetros, utilizando las derivadas espaciales de los valores de presión. Adicionalmente se mejora el algoritmo mediante la inclusión de un estimador Bayesiano que aprovecha las múltiples soluciones del problema inverso.

El enfoque desarrollado se prueba a través de múltiples experimentos sintéticos que consisten en 250 conjuntos de datos que simulan aleatoriamente la conductividad hidráulica (K), los cuales representan los diferentes valores de conductividad hidráulica del acuífero (conocido como campo K), lo que da como resultado diferentes niveles de heterogeneidad. Además, usando un número diferente de pruebas de bombeo (1-4 pruebas) dan como resultado un total de 1,000 experimentos.

Se considera un gran número de casos para asegurar que los hallazgos y conclusiones sean robustos. Los resultados revelaron mejoras significativas en las estimaciones

de K , el tiempo de cómputo y las predicciones de pruebas realizadas de forma independiente (que no se utilizaron durante la estimación de los parámetros en el problema inverso). Los resultados son comparados con un método geoestadístico altamente usado en la actualidad. En general, los resultados revelan que el enfoque de inversión propuesto (el algoritmo de inversión GM), es capaz de alcanzar niveles de precisión similares o superiores utilizando la mitad de pruebas de bombeo y el 20% del tiempo de cómputo, en comparación con un enfoque de inversión geoestadístico.

Los análisis y resultados presentados en esta sección fueron publicados en [Minutti et al. \(2019\)](#).

1.1. Descripción del problema

El agua de un acuífero típicamente se encuentra almacenada en materiales porosos, cómo pueden ser arenas e incluso roca, y que se encuentran agrupados en diferentes capas. Las diferentes porosidades de los materiales producen diferentes conductividades del fluido, haciendo más fácil o difícil el flujo del mismo, afectando a su vez las direcciones de flujo. El conjunto de conductividades hidráulicas (K) que se encuentran en un acuífero se le conoce cómo **campo K** , y su caracterización es de gran importancia para otras áreas de estudio cómo son la predicción del flujo o transporte de contaminantes. Una de las técnicas más exitosas hasta ahora para modelar la heterogeneidad de K es la tomografía hidráulica (TH).

Durante un estudio de tomografía hidráulica, el agua es bombeada (o inyectada) en un acuífero en diferentes zonas del mismo. Durante cada evento de bombeo/inyección, se monitorean las respuestas de presión en los pozos del acuífero, los cuales están presentes a diferentes intervalos, obteniendo un conjunto de datos de aumento o disminución de presión (*drawdowns*). Mediante el bombeo/inyección secuencial de agua en un intervalo y el monitoreo de las respuestas de la presión en estado estacionario en otros, se obtienen múltiples conjuntos de datos de cambios de presión ([Yeh and Liu, 2000](#)).

Estos valores de presión pueden compararse con los resultados dados por la ecuación de flujo de agua subterránea (el modelo del problema directo) que describe la respuesta de la presión, teniendo en cuenta los valores de conductividad hidráulica (K), la tasa de bombeo/inyección y las condiciones iniciales y de contorno. El objetivo de TH es encontrar los valores de K en todo el acuífero que minimicen la diferencia

entre los valores de presión observados y los simulados, donde a este proceso también se conoce cómo el problema inverso.

Con el fin de estimar los valores de K , se han desarrollado muchos enfoques. Por ejemplo, [Hoeksema and Kitanidis \(1984\)](#) usaron el método de *cokriging* para modelar la heterogeneidad de K en el acuífero, pero este método se basa en la disponibilidad de suficientes mediciones tanto de presión como de K . Además, debido a que el *cokriging* es un estimador lineal, su aplicación y precisión es limitada debido a la naturaleza no lineal del problema.

[Gottlieb and Dietrich \(1995\)](#) minimizaron el error cuadrado medio entre valores observados y los simulados usando la regularización de Tikhonov para penalizar cuánto se desvía la solución de una conjetura geológica *a priori*, donde el valor de K de cada elemento en un método de elementos finitos se optimiza. La suavidad y la precisión de la solución pueden verse afectadas por la cantidad de datos disponibles y la calidad de la suposición previa. Además, la optimización de cada valor de K en todo el dominio de simulación es computacionalmente costosa.

[Yeh et al. \(1996\)](#) desarrollaron un enfoque de estimación estocástica iterativa llamado Estimador Lineal Sucesivo (*Successive Linear Estimator*, SLE) donde el campo de valores del $\ln-K$ estimado mediante *cokriging* se actualiza en cada iteración con nuevas covarianzas y estimaciones de covarianza cruzada, con el fin de intentar capturar la relación no lineal entre los valores de presión y K . [Yeh and Liu \(2000\)](#) ampliaron el enfoque SLE incorporando secuencialmente conjuntos de datos de múltiples pruebas de bombeo/inyección en el modelo inverso, lo que dio como resultado el Estimador Lineal Sucesivo Secuencial (SSLE, *Sequential Successive Linear Estimator*).

Desde el trabajo pionero de [Yeh and Liu \(2000\)](#), se han desarrollado una serie de algoritmos de tomografía hidráulica basados en geoestadística (*e.g.*, [Zhu and Yeh \(2005\)](#); [Cardiff et al. \(2009\)](#); [Castagna and Bellin \(2009\)](#); [Mao et al. \(2013\)](#); [Zha et al. \(2017\)](#)), steady shape analysis ([Bohling et al., 2002](#)), y travel time inversion ([Brauchler et al., 2003](#); [Qiu et al., 2019](#)).

Se han publicado varios estudios sobre TH que demuestran el sólido desempeño de SSLE y SimSLE. En particular, [Illman et al. \(2007, 2008, 2010\)](#), y [Berg and Illman \(2011\)](#) han validado la tomografía hidráulica de estado estacionario (SSHT) basada en SSLE con datos recogidos de experimentos de laboratorio. Sin embargo, [Illman et al. \(2008\)](#) reportó que el orden en que los datos se incluyen en el algoritmo SSLE

tuvo un efecto significativo en la calidad de las estimaciones de K . Posteriormente, basado en SSLE, [Xiang et al. \(2009\)](#) desarrollaron el Estimador Lineal Sucesivo Simultáneo (SimSLE) que utiliza todos los conjuntos de datos simultáneamente en el algoritmo de inversión para superar los problemas asociados con la inversión secuencial de datos. Es así que, reconociendo que el algoritmo inverso basado en geoestadística produce estimaciones de parámetros hidráulicos suaves, [Zhao et al. \(2016\)](#) mejoró los resultados de [Illman et al. \(2007, 2010\)](#), y [Berg and Illman \(2011\)](#) al incluir en SimSLE datos geológicos como una distribución *a priori* de K .

En campo, [Berg and Illman \(2015\)](#) compararon los diferentes métodos para modelar la heterogeneidad del acuífero en un sitio de campo altamente heterogéneo a través de datos recolectados en el Sitio de Investigación del Campus Norte (NCRS) en el campus de la Universidad de Waterloo, y mostró que el análisis de tomografía hidráulica transitoria (THT) con SSLE ([Zhu and Yeh, 2005](#)) dio el mejor desempeño. Posteriormente, [Zhao and Illman \(2017, 2018\)](#) realizaron análisis de SSHT y THT con pruebas de bombeo en el NCRS y mejoró los resultados de [Berg and Illman \(2015\)](#) al incluir pruebas de bombeo adicionales y datos geológicos como la distribución inicial de K en el proceso de inversión. Compararon estos resultados con los métodos de zonificación geológica y de parámetros efectivos, demostrando que el TH con SimSLE se desempeñó de la mejor manera en términos de calibración y validación de modelos. La inclusión de datos geológicos fue útil para producir distribuciones de parámetros hidráulicos menos suaves y también mostró cómo la distribución inicial de K produce diferentes soluciones del campo K , sugiriendo la necesidad de tomar en cuenta el problema de las soluciones óptimas locales.

Independientemente de los buenos resultados en TH que se han logrado con enfoques de inversión geoestadística como SSLE y SimSLE, el tiempo de cómputo podría ser un problema en el caso de modelos grandes, tridimensionales y transitorios. Además, los enfoques geoestadísticos inversos pueden depender significativamente de la distribución *a priori* de K , y los algoritmos pueden sufrir de una lenta convergencia, especialmente para acuíferos altamente heterogéneos, ya que los parámetros hidráulicos se actualizan continuamente para capturar la dependencia no lineal entre K y la presión.

Una parametrización del campo K que reduce el número de parámetros a estimar es un componente clave para obtener una solución robusta al problema inverso de TH. Una buena parametrización permite obtener buenas estimaciones del campo K con

un tiempo de cómputo más corto. Además, la parametrización también puede afectar el número de óptimos locales o globales en la solución del problema inverso y también tiene que ver con cuán realista es la solución. Varios estudios han demostrado ser prometedores en la reducción del esfuerzo computacional para la TH basado en los métodos geoestadísticos inversos (Liu et al., 2013; Zha et al., 2018).

Con el fin de modelar la distribución espacial de K , en este capítulo se introduce un nuevo algoritmo llamado algoritmo de inversión de Mezclas Gaussianas (o algoritmo GM), para resolver el problema inverso para aplicaciones de TH. Este algoritmo se basa en el concepto estadístico de un modelo de mezclas, específicamente un Modelo de Mezclas Gaussianas (MMG), el cual se utiliza para parametrizar el campo K . Los resultados del algoritmo GM se compararon con los de SSLE utilizando múltiples experimentos sintéticos así como datos de un acuífero de laboratorio (caja de arena).

En el algoritmo de inversión de GM se introduce una nueva función objetivo, que se basa en las derivadas espaciales de los valores de presión. Esta función permite que el algoritmo de inversión aumente la precisión de las estimaciones de K . También tiene un término de regularización para abordar la naturaleza mal planteada del problema que requiere de estimar más parámetros que el número de observaciones disponibles. La regularización se aplica al campo estimado de K y no a la parametrización, lo cual genera una interpretación más intuitiva del parámetro de regularización, ya que se relaciona con la suavidad del campo K generado. Además, el nuevo algoritmo incluye un término de penalización que permite incluir y controlar el uso de información geológica para mejorar el proceso de inversión.

Para este estudio, se añade un nuevo estimador Bayesiano al proceso de inversión, que aprovecha las múltiples soluciones que se pueden encontrar en el proceso de inversión y produce una nueva solución que es un “consenso” de las diferentes soluciones, aumentando la precisión del campo estimado de K , lo que conduce a un mejor rendimiento de las predicciones de *drawdowns* (disminuciones de presión en pruebas de bombeo).

Los resultados del algoritmo de inversión GM se comparan con los del enfoque de inversión geoestadística, SSLE, utilizando un conjunto de datos de 250 campos K sintéticos, con 1 a 4 pruebas de bombeo en el proceso de inversión, lo que resulta en 1,000 experimentos. Los campos sintéticos K son creados aleatoriamente incluyendo diferentes capas de materiales, lo que resulta en una gran variedad de acuíferos y heterogeneidad, lo que nos permite probar y explorar en profundidad los algoritmos

de inversión para diferentes escenarios. Los resultados y las mediciones de error también se comparan con el caso de un acuífero homogéneo con un valor efectivo de K (valor medio de K en el acuífero), con el fin de mostrar las mejoras en los resultados a través de la caracterización de la heterogeneidad.

1.2. El algoritmo de inversión GM

1.2.1. Parametrización del campo de conductividad

La parametrización del campo K se refiere al proceso en el que un conjunto de valores, es decir, los parámetros, modela el campo K del acuífero, representando cada campo K con un conjunto diferente de valores de los parámetros.

Aunque el valor de estos parámetros no se conoce *a priori*, es posible estimarlos comparando los diferentes valores de presión (altura del agua en el pozo) generados mediante el modelo del problema directo (la ecuación del flujo de agua subterránea), con los datos observados medidos durante las pruebas de bombeo/inyección, es decir, mediante el proceso de minimización del error de una función objetivo:

$$\begin{aligned} \text{mín } F(\boldsymbol{\theta}) = \text{mín } ||H(G(\boldsymbol{\theta})) - \text{data}|| \\ \text{s.a. } L \leq \boldsymbol{\theta} \leq U \end{aligned} \quad (1.1)$$

donde $\boldsymbol{\theta}$ es el vector de parámetros utilizado por una función de parametrización G , que genera un campo K , que será utilizado por la función H para simular los valores de presión con la ecuación del flujo de agua subterránea. L y U se refieren a los límites inferior y superior de $\boldsymbol{\theta}$.

La selección usual y más simple para la parametrización del campo K es $G(\boldsymbol{\theta}) = \mathbf{K}$, es decir, la parametrización es exactamente \mathbf{K} , consistente en un vector de estimaciones locales de K . Sin embargo, optimizar cada valor de K para el campo K puede resultar costoso desde el punto de vista computacional, y en muchos casos resultará en un problema mal planteado, por lo que se propone utilizar una función de parametrización G , basada en un Modelo de Mezclas Gaussianas.

El algoritmo de optimización trabaja sobre la parametrización seleccionada para encontrar los valores de los parámetros que mejor reproducen los valores de presión

observados. Por lo tanto, la selección correcta de la parametrización es un componente clave del algoritmo de inversión.

Mostraremos que la parametrización del MMG tiene un gran impacto en el proceso de optimización así cómo en las características del campo K generado.

1.2.2. Modelo de Mezclas Gaussianas

En estadística, un modelo de mezclas es un modelo probabilístico usado para representar diferentes sub-poblaciones dentro de una población más grande, a través de la combinación de muchas funciones de densidad de probabilidad, donde cada componente de la función podría ser interpretada cómo una “probabilidad” o cómo la verosimilitud de la existencia de cada sub-población en cada coordenada espacial.

En este caso, la población principal sería el tipo de material (medio poroso) en el acuífero que tiene un mayor volumen o el que está más disperso y las sub-poblaciones serían cualquier otro material encontrado dentro de la población principal.

Aunque una mezcla describe una distribución de probabilidad, su valor puede ser re-escalado para describir diferentes atributos, cómo la densidad de valores de K . En el enfoque aquí presentado, la mezcla describe el logaritmo de K ($\ln-K$).

Si se utilizan distribuciones Gaussianas para el modelo de mezcla, esto se conoce cómo el Modelo de Mezclas Gaussianas (MMG), que está representado por:

$$G(\boldsymbol{\theta}) = G(X, \mu_i, \Sigma_i) = \sum_{i=1}^{N_k} \omega_i N(X, \mu_i, \Sigma_i) \quad (1.2)$$

donde ω_i es un parámetro de peso o importancia para la distribución Gaussiana i -ésima, con media μ_i y matriz de covarianza Σ_i , N_k es el número de componentes en la mezcla, y X es el vector de coordenadas espaciales.

En el análisis de conglomerados (*cluster analysis*) y en *machine-learning*, las mezclas Gaussianas también se utilizan cómo método de clasificación estadística, donde la tarea consiste en estimar la probabilidad de que cada observación forme parte de una clase o de un grupo. Debido a que el medio poroso del acuífero puede ser visto cómo una población o conglomerado (zonas) de diferentes materiales con diferentes valores de K , desde un punto de vista estadístico, un MMG parece ser una selección adecuada para describir la heterogeneidad de K dentro de un acuífero.

Si el campo K tiene una distribución espacial compleja que no se asemeja a los

núcleos Gaussianos, esta se puede aproximar aumentando el número de componentes en la mezcla: cuanto mayor sea el número de componentes, mayor será la complejidad del campo K que se puede modelar.

Con el fin de mostrar la efectividad de un MMG para modelar un grupo de diversos materiales, la Figura 1.1a presenta una imagen compuesta de formas no Gaussianas 2D, basada en la Figura 3 de Berg and Illman (2011), que es una discretización del campo K de un experimento de caja de arena, realizado usando una cuadrícula de 741 elementos (39 por 19), es decir, el número de parámetros que posee la imagen.

Debido a que el objetivo es reconstruir la imagen en la Figura 1.1a, en este experimento no hay datos de presión involucrados. Los parámetros del MMG fueron optimizados para aproximarse a la imagen minimizando el Error Cuadrado Medio (o por sus siglas en inglés MSE) entre los valores de 1.1a y los valores del MMG, utilizando diferentes números de componentes: $N_k = 3, 6, 12, 24, 48$. En este caso, cada componente de la mezcla consta de 6 parámetros (un peso, la media, la varianza de las direcciones espaciales x y z , y su covarianza). Los resultados se presentan en la Figura 1.1b-1.1f. Para cada representación se calculan el error absoluto medio (o por sus siglas en inglés MAE) y el MSE.

Se puede ver que con $N_k = 48$ la representación es muy similar a la Figura 1.1a, siendo necesario estimar 288 valores de parámetros (6 por 48) del MMG, en lugar de los 741 valores que tiene la imagen. Sin embargo, si el objetivo del análisis es detectar las áreas principales de la Figura 1.1a, entonces $N_k = 12$ podría ser suficiente para ese propósito, lo que implica la estimación de 72 valores, que es 10 veces menos que el número original de parámetros en la imagen.

Para modelar la distribución espacial de K con un MMG, se propone aproximar el verdadero campo K , \mathbf{K} , en el algoritmo GM mediante:

$$\ln(\mathbf{K}) \approx G(\boldsymbol{\theta}^*) - \ln(\bar{K}) \quad (1.3)$$

donde \bar{K} es el valor efectivo o promedio de K , en el acuífero y $G(\boldsymbol{\theta})$ es el MMG que minimiza la función objetivo propuesta.

Si la Ecuación 1.3 se utiliza para aproximar la distribución espacial de un campo bidimensional de K , se requiere un conjunto de distribuciones Gaussianas bi-variadas para generar el MMG. Para el MMG, el parámetro N_k es el más importante, ya que es el número de componentes/materiales en la mezcla y cuanto mayor sea su valor, mayor

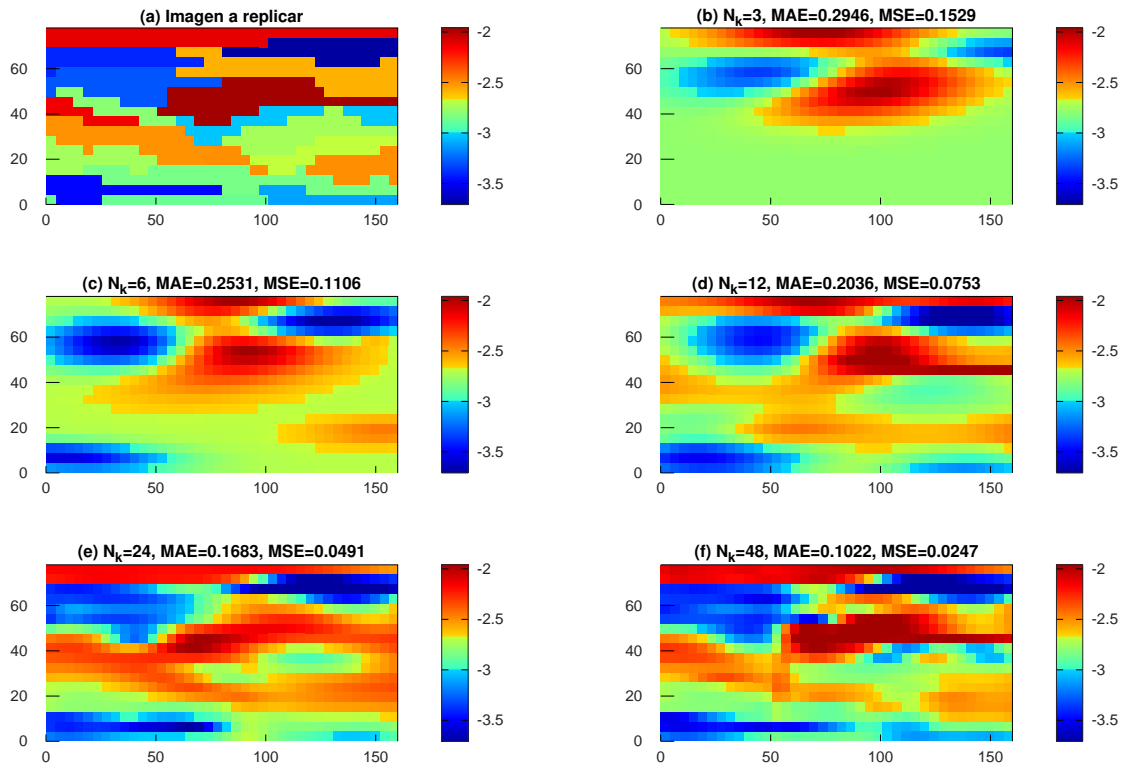


Figura 1.1: (a) Imagen a ser replicada; (b)-(f) Representación de la imagen a través de un modelo de mezcla Gaussiana con diferentes componentes. ($N_k = \{3, 6, 12, 24, 48\}$). MAE y MSE se presentan como medidas de precisión.

será el nivel de complejidad que el MMG pueda modelar. Bajo ciertos supuestos, cada componente de la mezcla representaría una capa de material. Aunque este valor puede ser estimado por el algoritmo de optimización, también puede ser fijado como una constante con la intención de detectar únicamente los principales N_k materiales que están presentes en el acuífero. Por lo tanto, también funciona como un parámetro de suavizado, que puede usarse para controlar el efecto del ruido en los datos (es decir, fluctuaciones aleatorias alrededor de los valores de presión verdaderos), para evitar el sobre-ajuste de los datos.

El resto de parámetros en el MMG se estiman mediante la optimización de una función objetivo. Estos parámetros controlan la correlación espacial de la heterogeneidad en el campo K estimado. En particular, Σ_i , la matriz de covarianza, contiene la varianza en las direcciones espaciales (horizontal y vertical) (σ_x^2 , σ_z^2), y la correlación entre ellas (ρ_{xz}). En el caso de los materiales estratificados, se puede esperar tener una mayor varianza horizontal para cada componente (es decir, correlación entre longitudes más largas), por ejemplo, una capa de material que está típicamente más extendida en la dirección horizontal que en la vertical; μ_i es la ubicación espacial del *centroide* de cada componente en las direcciones horizontal y vertical, mientras que ω_i es el peso de cada componente en la mezcla, lo que está directamente relacionado con el valor de K de cada componente/material.

Esta parametrización también tiene la ventaja de tener parámetros altamente sensibles, lo que significa que un pequeño cambio en el valor del parámetro, tiene un gran efecto en el valor de la función objetivo. Como el método de optimización utiliza el gradiente para actualizar los parámetros, el valor óptimo del parámetro sensible se encontrará más rápido. Por ejemplo, si el valor de μ_i cambia, todos los valores de K a su alrededor se verán afectados, lo que tendrá un gran efecto en la función objetivo. Por lo tanto, las áreas que reducen la función objetivo mediante la colocación de un componente Gaussiano pueden ser detectadas rápidamente. Además, si los componentes están suficientemente alejados entre sí, la correlación de los parámetros entre los distintos componentes será baja, lo que dará lugar a un proceso de optimización por zonas, que es un problema más simple de resolver.

1.2.3. Función objetivo

Aunque el enfoque propuesto está estrechamente relacionado con los métodos de agrupamiento y de clasificación estadística, la principal diferencia radica en que el campo K a aproximar con un el MMG no se conoce *a priori*, por lo que es necesario estimarlo, utilizando la presión medida cómo una variable proxy de K .

Típicamente, el error a ser minimizado en el problema inverso de TH es:

$$F_{LS}(\boldsymbol{\theta}) = \sum_{i=1}^n (h_i(\boldsymbol{\theta}) - h_i^*)^2 \quad (1.4)$$

donde h_i^* son los valores de presión observados para todas las pruebas de bombeo/inyección y $h_i(\boldsymbol{\theta}) = H_i(G(\boldsymbol{\theta}))$ son los valores de presión simulados usando el modelo del problema directo, donde $G(\boldsymbol{\theta})$ es el campo $\ln-K$ estimado.

Adicionalmente se propone la inclusión de las derivadas espaciales de los valores de presión:

$$F_d(\boldsymbol{\theta}) = \phi_f F_{LS}(\boldsymbol{\theta}) + \phi_x \sum_{i=1}^n \left(\frac{d}{dx} h_i(\boldsymbol{\theta}) - \frac{d}{dx} h_i^* \right)^2 + \phi_z \sum_{i=1}^n \left(\frac{d}{dz} h_i(\boldsymbol{\theta}) - \frac{d}{dz} h_i^* \right)^2 \quad (1.5)$$

donde ϕ_i son pesos tales que $\sum_i \phi_i = 1$, y x, z son las direcciones espaciales horizontal y vertical.

Las derivadas espaciales son especialmente útiles, ya que son más sensibles a las localizaciones de los componentes Gaussianos, siendo más fácil para el algoritmo de optimización colocar cada componente en función del cambio de presión entre dos valores de presión observados. De lo contrario, cuando sólo se utilizan valores de presión, el efecto se suaviza a lo largo de todos los valores. El peso de las derivadas espaciales podría utilizarse para incluir algún conocimiento sobre el acuífero. Por ejemplo, si se sabe que el material se distribuye en capas, un peso mayor para ϕ_z implica una función objetivo que es más sensible a los cambios en la dirección vertical.

Cómo ejemplo del efecto que surte la inclusión de las derivadas, se considera un acuífero sintético compuesto por tres capas de diferentes materiales. Suponiendo que se desconozca el valor de K para dos de las capas, las funciones objetivo para

estos dos parámetros pueden construirse variando los K estimados de cada capa y calculando el MSE de los valores de presión y el MSE de las derivadas. Las funciones objetivo construidas y sus contornos se presentan en la Figura 1.2.

Desde el punto de vista de optimización, es preferible una función objetivo creada con el MSE de las derivadas, ya que (1) un algoritmo de optimización encontrará el valor mínimo con un menor número de iteraciones debido a la alta convexidad alrededor del mínimo; (2) la función es más sensible a los cambios en los parámetros y, dependiendo de los criterios de parada, un algoritmo de optimización podría detenerse en una solución no óptima; (3) la función objetivo tiene parámetros poco correlacionados, lo que significa que un error en la estimación de uno de los parámetros tendrá un bajo impacto en la estimación del otro, resultando en una mayor precisión de los parámetros estimados.

Además, las derivadas espaciales podrían ser útiles para disminuir el impacto del efecto del *skin* (es decir, el daño de la formación alrededor del pozo), debido a que el *skin* aumenta/disminuye el valor de la presión en una constante y la derivada no cambia si se añade una constante a la función. Sin embargo, en la práctica, las derivadas espaciales de los valores de la presión se aproximan utilizando el pozo de observación más cercano. Por lo tanto, el efecto sobre la *skin* podría ser cancelado o disminuido si los pozos de observación cercanos tienen un valor similar de *skin*. En un modelo transitorio, la derivada temporal resuelve este problema, porque la derivada se estima utilizando valores de presión del mismo pozo (con igual *skin*).

Un inconveniente para el uso de las derivadas es que el nivel de ruido en los valores observados de la presión podría aumentar, pero este problema se soluciona eligiendo un valor pequeño de N_k , o usando un método de reducción de ruido como el que se presenta en el capítulo 3, el cual está diseñado para abordar datos con ruido en problemas inversos cuando se requiere la derivada, aproximando los datos de forma similar a los splines de regresión, pero utilizando diferentes soluciones del modelo del problema directo como base (o espacio de funciones) del modelo de regresión.

1.2.4. Regularización

Se añade una medida de suavidad del campo K como condición de regularización en la función objetivo. Este término puede llevar a un proceso de optimización más rápido evitando “visitar” óptimos locales que sólo ajustan a partes de los datos de

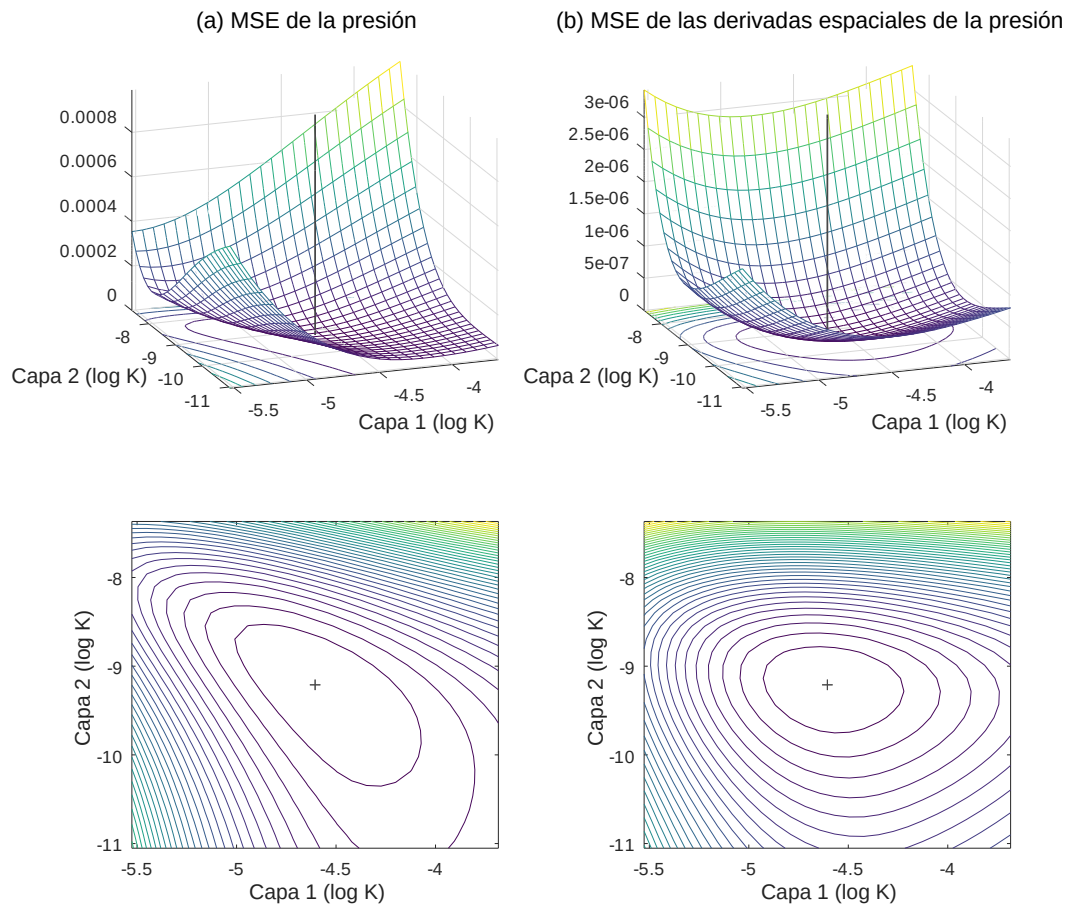


Figura 1.2: Funciones objetivo y gráficos de contorno para los valores de K de las dos capas de material cuando (a) se utiliza el error cuadrático de los valores de la presión y (b) cuando se utiliza el error cuadrático de las derivadas espaciales. La ubicación del valor óptimo está marcada con una línea sólida y una cruz.

calibración. La suavidad $S(\boldsymbol{\theta})$ de la solución se mide por:

$$S(\boldsymbol{\theta}) = \sum_{i=1}^m \left(\frac{d}{dx} G(\boldsymbol{\theta}) \right)_i^2 + \sum_{i=1}^m \left(\frac{d}{dz} G(\boldsymbol{\theta}) \right)_i^2 \quad (1.6)$$

donde m es el número de valores de K usados para la solución del modelo del problema directo.

Finalmente, la función objetivo del problema de optimización puede ser escrita como:

$$F(\boldsymbol{\theta}) = (1 - \lambda)F_d(\boldsymbol{\theta}) + \lambda S(\boldsymbol{\theta}) \quad (1.7)$$

donde λ ($0 \leq \lambda < 1$), el parámetro de regularización es especialmente útil si hay pocas mediciones de datos y/o hay datos con ruido.

El parámetro λ también puede ser usado para representar algún conocimiento sobre la heterogeneidad de los materiales que están presentes en el acuífero, donde un acuífero con materiales con valores de K similares (es decir, un caso de baja varianza) produce una distribución de K más suave (un valor de *lambda* alto). Para el análisis TH presentado en este trabajo, este parámetro se estableció en $\lambda = 0.5$.

Aunque λ y N_k pueden utilizarse como parámetros de suavizamiento, estos dos valores representan diferentes tipos de suavizamiento. Los valores altos de λ hacen que los campos K tengan una variación baja, es decir, acuíferos más homogéneos, mientras que los valores pequeños de N_k tienden a agrupar diferentes capas de materiales con K similares, pero permiten una variación alta entre grupos. En otras palabras, λ puede utilizarse para describir la heterogeneidad intracapa y N_k para la heterogeneidad entre capas.

1.2.5. Información geológica

Si valores observados del campo K están disponibles, esta información puede ser incluida en el modelo para mejorar la solución del algoritmo de optimización añadiendo un término de penalización en la función objetivo, aumentando el valor de la función cuando aumentan las discrepancias. El término de penalización $P(\boldsymbol{\theta}|\mathbf{K}^*)$

puede escribirse como:

$$P(\boldsymbol{\theta}|\mathbf{K}^*) = \frac{1}{l} \sum_{i=1}^l (\ln(K_i^*) - G(\boldsymbol{\theta})_i)^2 \quad (1.8)$$

donde K_i^* es el valor i -th observado K , $G(\boldsymbol{\theta})_i$ es el valor ln - K generado por la función de parametrización G (utilizando $\boldsymbol{\theta}$ como parámetros de entrada), en las ubicaciones i , con un total de valores observados l .

Este término de penalización se incluye en la función objetivo a través de un peso, ψ , que representa la fiabilidad de los valores observados, cuando se dispone de datos geológicos:

$$F(\boldsymbol{\theta}|\mathbf{K}^*) = (1 - \lambda)(1 - \psi)F_d(\boldsymbol{\theta}) + \lambda S(\boldsymbol{\theta}) + \psi P(\boldsymbol{\theta}|\mathbf{K}^*) \quad (1.9)$$

donde $0 \leq \psi < 1$, con valores altos de ψ para estimaciones confiables de K observados, mientras que los valores pequeños para aproximaciones. Este término de penalización no hace que los valores estimados de K sean exactamente los mismos que los observados, sino que sólo pide proximidad a ellos, durante el proceso de optimización, siendo esta proximidad más estricta dependiendo del valor de ψ .

1.2.6. Relación con modelos geológicos y geoestadísticos

Aunque el método propuesto está estrechamente relacionado con los métodos de agrupamiento y la clasificación estadística de *machine-learning*, también mantiene una relación con modelos geológicos y geoestadísticos.

Por ejemplo, los algoritmos de inversión y optimización geoestadística (tipo Newton) utilizan el jacobiano (es decir, la matriz de sensibilidad) para actualizar el valor de los parámetros. En el caso de un algoritmo de inversión geoestadística como SSLE, las covarianzas cruzadas entre los valores de presión y las conductividades se construyen utilizando la matriz jacobiana, relacionando los cambios en la presión hidráulica con los cambios en la log-conductividad (Renshaw, 1996).

Además, algunos de los parámetros del algoritmo de inversión de GM describen la heterogeneidad entre capas e intracapas. Por ejemplo, las escalas de correlación del campo K , λ_x y λ_z utilizadas en los enfoques de inversión geoestadística, son similares a las varianzas espaciales de cada componente Gaussiano ($\sigma_{i_x}^2$ y σ_z^2), ya que

estas varianzas describen la forma en que el valor de K del componente decae con la distancia desde su centro (μ_i).

El número de componentes en el MMG (N_k) está relacionado con la heterogeneidad entre capas, debido a que este parámetro describe el número de capas diferentes, necesarias para representar el acuífero, donde cuanto menor es la heterogeneidad, menor es el número de componentes, siendo $N_k = 0$ para un acuífero homogéneo.

Además, el parámetro N_k tiene una relación con el modelo geológico. En el enfoque de zonificación geológica la parametrización se realiza particionando el dominio en un conjunto de sub-dominios llamados zonas (Carrera et al., 2005), y estas zonas están relacionadas con los componentes Gaussianos del MMG, donde los K de cada componente (zona) están representados principalmente por el peso del componente, ω_i .

En un escenario ideal, cada capa del acuífero puede ser aproximada por un componente Gaussiano. Entonces, esta capa (o un conjunto de capas cercanas, con K similar) representa una zona, mientras que la ubicación de la zona y sus valores de K principales serán estimados por el algoritmo de optimización a través de los parámetros ω_i y μ_i . El número de zonas está determinado por el parámetro N_k , que puede ser fijado u optimizado, y la heterogeneidad intracapa se describirá con Σ_i , la matriz que consiste en la varianza espacial de K y la correlación entre las direcciones espaciales. Cualquier información existente sobre los datos geológicos se incorpora utilizando el término de penalización en la función objetivo (Ec. 1.9), donde cuanto mayor sea el valor de ψ , mayor será el peso de esa zona para aproximarse a ese valor, o si no se dispone de tal información, la Ecuación 1.7 se utiliza en su lugar.

Si el material en el acuífero tiene una distribución espacial de K altamente compleja, que no se asemeja a un componente Gaussiano (por ejemplo, un acuífero de roca fracturada), el aumento del número de componentes puede utilizarse para aproximarse a la distribución espacial subyacente de K , cómo se muestra en la Figura 1.1.

En general, el enfoque propuesto combina las características de un modelo geológico y un modelo geoestadístico, donde, si N_k es grande, el algoritmo GM se asemeja más a un modelo inverso geoestadístico altamente parametrizado. Por otro lado, si N_k es pequeño, el algoritmo de optimización tiende a agrupar en zonas o áreas de K similares, cómo las encontradas en un modelo geológico, y el valor de K de las zonas puede ser descrito a través del término de la penalización.

1.2.7. Implementación

Es necesario tener en cuenta algunas consideraciones para implementar el algoritmo de inversión propuesto.

Para los experimentos presentados en este trabajo, el algoritmo GM se implementó utilizando el algoritmo de optimización Levenberg-Marquardt, que utiliza una técnica conocida como regularización, con este parámetro de regularización funcionando como una combinación ponderada del paso Gauss-Newton y la de dirección de máximo descenso. Este algoritmo se utiliza ampliamente para resolver problemas no lineales de mínimos cuadrados (Heath, 1996).

En funciones objetivo altamente no lineales, la existencia de múltiples soluciones locales óptimas es un problema en el que la solución del algoritmo de inversión depende del valor inicial de los parámetros. Por ejemplo, Zhao and Illman (2018) ha mostrado que se podrían producir diferentes distribuciones de K para un acuífero dependiendo de la distribución inicial de K que se utilizó para el modelado inverso geoestadístico.

Con el fin de encontrar una mejor aproximación al óptimo global, se adopta un enfoque multistart (ver Ugray et al. (2007)). Este enfoque consiste en realizar múltiples optimizaciones utilizando diferentes valores iniciales de los parámetros, dando como resultado una solución para cada valor inicial, donde la que posee el valor mínimo de la función objetivo es utilizada como resultado final.

Aunque un enfoque multistart aumenta el tiempo de cómputo, en lugar de añadir más pruebas de bombeo/inyección para producir una función objetivo con un mejor comportamiento, el enfoque de multistart permite el uso de los datos disponibles de forma más eficiente, utilizando las diferentes soluciones para mejorar la estimación del campo K , si no se encuentra el mínimo global. También permite estimar la varianza de las estimaciones de K , utilizando las diferentes soluciones como observaciones muestrales.

La información sobre la distribución de K en el acuífero también podría utilizarse en la función objetivo (Ec. (1.9)) a fin de evitar el enfoque multistart. Sin embargo, a fin de probar las capacidades del algoritmo GM para producir representaciones precisas del campo K , en este trabajo asumimos que dicha información no está disponible, y que se utiliza la Ecuación 1.7 como función objetivo, con un valor aleatorio de los parámetros que son utilizados como estimación inicial.

Es posible utilizar un algoritmo de optimización global que tenga en cuenta

la existencia de múltiples óptimos locales en lugar de un enfoque multistart (por ejemplo, el método del Tunel (Gomez and Levy, 1982), que se utiliza en el problema de *history matching* (Gomez et al., 2001)). Sin embargo, en el algoritmo de inversión GM, se utilizaron las múltiples soluciones del enfoque multistart para estimar el campo de conductividad de Bayes (sección 1.2.8), que tiende a mejorar la solución del algoritmo de optimización.

La conductividad efectiva \bar{K} en la Ecuación 1.3 es también un parámetro que debe ser estimado por el algoritmo de optimización, donde este valor y el campo K estimado están limitados por un valor mínimo y máximo de K , que debe ser establecido como parte del proceso de inversión. Se necesitan estimaciones razonables de los límites de K (especialmente cuando las derivadas tienen un alto peso en la función objetivo), ya que es el cambio relativo entre los valores de K lo que afecta a los cambios en los valores de presión. Por lo tanto, estimaciones erróneas de los valores mínimos y máximos de K podrían resultar en un campo K espacialmente consistente con el campo K real, pero con valores de K re-escalados.

En la práctica, es necesario estimar las derivadas espaciales utilizando el método de diferencias finitas:

$$\frac{d}{dx} h_l^*(x_i, z_j) \approx \frac{h_l^*(\theta|x_i, z_j) - h_l^*(\theta|x_{i-1}, z_j)}{x_i - x_{i-1}} \quad (1.10)$$

$$\frac{d}{dz} h_l^*(x_i, z_j) \approx \frac{h_l^*(\theta|x_i, z_j) - h_l^*(\theta|x_i, z_{j-1})}{z_j - z_{j-1}} \quad (1.11)$$

donde $h_l^*(x_i, z_j)$ es el valor de presión observado para la l -ésima prueba de bombeo/inyección, en la ubicación (x_i, z_j) .

Por simplificación, se asume que los valores de presión observados en las direcciones x y z están en una configuración de cuadrícula. Sin embargo, cuando este no es el caso, el algoritmo puede ser adaptado para usar los k vecinos más cercanos, con $k = 1, \dots, n - 1$ donde n es el número de observaciones de la presión.

Finalmente, un valor pequeño de N_k implica un menor número de parámetros a ser estimados en el MMG, sin embargo se podría requerir más iteraciones del algoritmo de optimización para ajustar los datos (es decir, más tiempo de cómputo). Si N_k es grande, se puede representar mejor el campo K subyacente, logrando convergencia con menos iteraciones, pero el número de parámetros a estimar será alto, utilizando más tiempo de cómputo. Por lo tanto, existe un valor eficiente de N_k , pero el valor

real a utilizar está más relacionado con la cantidad y calidad de los datos observados. En el enfoque multistart, se pueden probar diferentes valores N_k , resultando en diferentes estimaciones del campo K .

1.2.8. Campo de conductividad de Bayes

El estimador de Bayes utiliza una probabilidad o una distribución de riesgo de los parámetros de interés (en este caso la función de riesgo es el MSE) para mejorar las estimaciones. Para encontrar el óptimo global de la función objetivo, el enfoque multistart explora diferentes valores iniciales de los parámetros, resultando en diferentes estimaciones del campo K , cada uno con un valor del MSE.

Estas diferentes soluciones pueden combinarse utilizando MSE^{-1} como ponderación, lo que da como resultado el valor esperado del campo K que minimiza la función de riesgo. Usualmente, esta nueva estimación del campo K mejora la mejor estimación del algoritmo de optimización. Llamamos a esta nueva estimación el campo de conductividad de Bayes.

Si el MSE se considera una función de riesgo y $\hat{\mathbf{K}}_i$ es un campo K , obtenido con el algoritmo de optimización (utilizando diferentes valores iniciales para los parámetros), entonces el estimador Bayes ($\hat{\mathbf{K}}_B$) se calcula con el valor esperado *a posteriori*:

$$\hat{\mathbf{K}}_B = E[\mathbf{K}|\mathbf{H}] = \int \mathbf{K} p(\mathbf{K}|\mathbf{H}) d\mathbf{K} \approx T \sum_i \frac{\hat{\mathbf{K}}_i}{\text{MSE}(\hat{\mathbf{K}}_i)} \approx T_H \sum_i \frac{\hat{\mathbf{K}}_i}{\text{MSE}(\mathbf{H}(\hat{\mathbf{K}}_i))} \quad (1.12)$$

donde \mathbf{H} son los valores de presión observados, $p(\mathbf{K}|\mathbf{H})$ es la probabilidad de que \mathbf{K} sea el verdadero campo K (para un determinado \mathbf{H}) y $T = \sum_i \text{MSE}(\hat{\mathbf{K}}_i)^{-1}$. $\text{MSE}(\mathbf{H}(\hat{\mathbf{K}}_i))$ es el MSE de los valores de presión cuando se usa $\hat{\mathbf{K}}_i$ como estimado del campo K , con $T_H = \sum_i \text{MSE}(\mathbf{H}(\hat{\mathbf{K}}_i))^{-1}$.

El campo K de Bayes, cuando se calcula utilizando las soluciones locales y/o sub-óptimas del algoritmo de optimización, puede ser interpretado como un método que resalta las características del acuífero que son consistentes entre las diferentes soluciones y suaviza aquellas con mayor incertidumbre. Además, estas diferentes soluciones del algoritmo de optimización pueden ser utilizadas para crear un mapa de incertidumbre de los valores estimados de K .

El MSE del estimador de Bayes puede calcularse y compararse con el MSE mínimo

de \hat{K}_i para confirmar que el campo K estimado se ha mejorado. Sin embargo, cuando la cantidad de datos observados es una limitante, el campo K de Bayes podría ser una mejor estimación del campo K (independientemente del valor del MSE), debido a que en este escenario habrá muchas soluciones óptimas locales o incluso globales, y el campo K de Bayes representa el consenso entre estas soluciones, siendo posiblemente una opción más “segura” para las predicciones de flujo y transporte.

1.3. Metodología y Experimentos

Para explorar en profundidad las capacidades del algoritmo de inversión GM, fue creado un generador aleatorio de campos K , el cual comienza con un acuífero homogéneo y utiliza polígonos para añadir diferentes capas de material, donde el número de capas, sus coordenadas y su valor de K se eligen al azar. Además, la transición entre las diferentes capas de materiales se simula utilizando el valor medio de K de los diferentes materiales presentes en cada celda del campo discretizado K .

Mediante el uso de este generador de campos K , se crearon 250 campos aleatorios, donde el número de materiales por acuífero varía de 8 a 20, con valores de K de $1e - 3$ a $1e + 2$ cm/s. El proceso de inversión se llevó a cabo utilizando 1-4 pruebas de bombeo, resultando en un conjunto de 1,000 experimentos.

Para todos los experimentos, se modela a partir de un acuífero bidimensional basado en la estructura del experimento de caja de arena presentado por [Illman et al. \(2010\)](#), que mide 193.0 cm de largo, 82.6 cm de alto y 10.2 cm de espesor. El acuífero fue discretizado en 741 elementos (una cuadrícula de 19 por 39) con dimensiones de elementos de 4.1 cm por 4.1 cm. En términos de condiciones de contorno, los límites superior, izquierdo y derecho del acuífero se establecen como límites de presión constantes, mientras que el inferior se trata como un límite sin flujo.

Los datos sintéticos son generados por simulación, con nueve pruebas de bombeo y 48 valores de presión observados para cada prueba (8 filas y 6 columnas), usando una tasa de bombeo constante de $Q = 0.95$ cm³/s. Los datos de las primeras 4 pruebas de bombeo (pruebas[1]-[4]) se utilizan como entrada para el modelado inverso (estos datos se denominan conjunto de datos de calibración) y los datos restantes de las pruebas 5 a 9 (conjunto de datos de validación) se utilizaron para la validación del campo K estimado con el fin de estimar el error de predicción de las caídas de presión

(o *drawdowns*). Un ejemplo de campos sintéticos de K que muestran las ubicaciones de los puertos de observación y bombeo se presenta en la Figura 1.3.

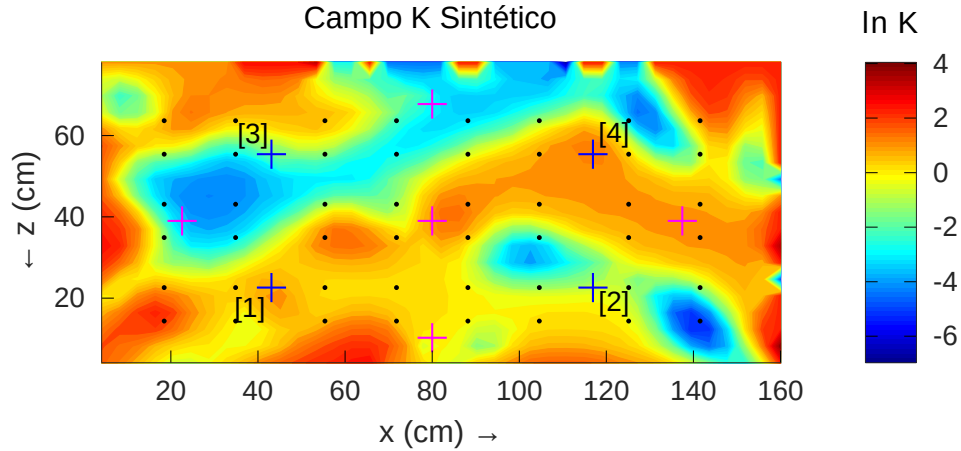


Figura 1.3: Campo esquemático de $\ln K$ simulado con el generador de campo aleatorio de K . Las cruces azules marcan la ubicación de los pozos de bombeo para la calibración, las rojas la ubicación de los pozos de bombeo utilizados para la validación y los puntos negros los pozos de observación. [1]-[4] son los puertos de bombeo para experimentos que incluyen de 1 a 4 pruebas de bombeo en el proceso de inversión.

Los experimentos pretenden medir (1) el efecto de la cantidad de datos utilizados; (2) la velocidad y exactitud en la estimación del campo K , (3) el rendimiento de cada método de inversión en la predicción de *drawdowns* y (4) la diferencia de los resultados si se asume un acuífero homogéneo. Utilizamos VSAFT2 (Xiang et al., 2009), un modelo de inversión geoestadística ampliamente utilizado en TH, para resolver los mismos problemas con el fin de comparar los resultados obtenidos a través del algoritmo GM.

Para ambos métodos de inversión se utilizó como distribución inicial de K un acuífero homogéneo con un valor de K efectivo (\bar{K}) dado por la media geométrica de los valores locales de K :

$$\bar{K} = \exp\left(\frac{1}{n} \sum_{i=1}^n \ln K_i\right) \quad (1.13)$$

donde K_i son los valores de conductividad local del campo K verdadero.

Debido a este valor inicial, la mejora real de cada algoritmo corresponde a la reducción del MSE con respecto a la solución inicial. Por lo tanto, para medir la mejora de cada método, se utiliza la siguiente medición del error del campo K estimado:

$$\begin{aligned} E_K(\text{alg}) &= 1 - \frac{\text{Error usando el valor inicial} - \text{Error usando alg}}{\text{Error usando el valor inicial}} \\ &= \frac{\sum_i \left(\ln \hat{K}_i^{(\text{alg})} - \ln K_i \right)^2}{\sum_i \left(\ln \bar{K} - \ln K_i \right)^2} = \frac{MSE^{(\text{alg})}}{Var(\ln K)} \end{aligned} \quad (1.14)$$

donde $\hat{K}_i^{(\text{alg})}$ son los estimados de K del algoritmo $\text{alg} = \{\text{GM}, \text{VSAFT2}\}$, y $MSE^{(\text{alg})}$ es el MSE del $\ln K$ para alg .

Por lo tanto, si $E_K = 1$, no hay mejora en el campo K estimado, y el error es el mismo que si se asume un acuífero homogéneo con conductividad efectiva \bar{K} . La Ecuación 1.14 está relacionada con el coeficiente de determinación (R^2) de modo que $E_K = 1 - R^2$. Es importante notar que $E_K > 1$ también podría resultar debido a que el campo K es estimado indirectamente usando pruebas de bombeo/inyección de datos de presión y es totalmente posible que un campo estimado de K pueda tener un MSE peor que el K efectivo, especialmente si hay escasez de datos para el modelado inverso, situación en la cual se espera que el problema inverso tenga múltiples soluciones.

Para el error de estimación de la calibración (E_{cal}) y de los *drawdowns* de validación (E_{val}), se utilizó la medida equivalente de la Ecuación 1.14, donde un error de 1 corresponde al mismo nivel de error que los *drawdowns* obtenidos, suponiendo un acuífero homogéneo con conductividad \bar{K} .

La simulación del problema directo y el modelado inverso se realizaron con un modelo de flujo de estado estacionario saturado e isotrópico, el cual fue programado en C++ usando el método de diferencias finitas, basado en las pautas de Wang and Anderson (1995) y el modelo matemático presentado en Harbaugh and Geological Survey (U.S.).

Todos los cálculos se realizaron en un PC con CPU i5-8250U a 1.60GHz con 8GB de RAM, usando un solo núcleo, con la tecnología *Turbo Boost* activa, resultando en una frecuencia de CPU de 3.40GHz.

1.4. Resultados

Una muestra de los 250 campos K sintéticos generados aleatoriamente se presenta en las Figuras 1.4, 1.5 y 1.6 (gráficos a-f), así cómo los campos K estimados con VSAFT2 (a*-f*) y con GM (a**-f**) cómo métodos de inversión, utilizando 4 pruebas de bombeo cómo datos de entrada. La Figura 1.4 presenta los 6 casos con la varianza de K más altas y la Figura 1.5 los 6 con las más bajas. La Figura 1.6 presenta una muestra aleatoria de 6 campos K sintéticos.

Se puede ver que ambos algoritmos detectan las ubicaciones de áreas grandes con valores de K altos y bajos, pero con respecto a las áreas pequeñas y sus formas, el algoritmo de inversión GM es más eficaz en capturar estas características, con límites de las capas más definidos, incluso para los casos en los que los campos K verdaderos no se parecen a una forma Gaussiana.

La Figura 1.6f es un ejemplo apropiado de estas capacidades, donde el algoritmo de inversión GM detectó y modeló todas las áreas pequeñas de bajo K , así cómo las áreas de alto K . Además, el algoritmo GM parece mapear mejor las capas de materiales que no son horizontales o verticales (es decir, con curvaturas), cómo lo son los casos en las Figuras 1.4a, 1.6a y 1.6e. De hecho, la dirección principal de cada capa de material se describe en el algoritmo GM a través de la correlación de las direcciones espaciales (ρ_{xz}), uno de los parámetros que se estima.

La Figura 1.7 presenta las correlaciones de Spearman entre las diferentes variables de los 1,000 experimentos para el algoritmo de inversión GM y VSAFT2. La variable más importante para reducir el error de estimación de K y el error de los *drawdowns* predichos (*drawdowns* en el conjunto de datos de validación), es el número de pruebas de bombeo utilizadas en el proceso de inversión. El descenso promedio de presión (**Mean_Drawdown**) de cada experimento también tiene un papel importante en la reducción de los errores, incluso más que la varianza de K (heterogeneidad) o su rango de valores ($\max(\ln K) - \min(\ln K)$), ya que el descenso promedio de presión está relacionada con el valor medio del campo K , lo que implica que un acuífero con alta conductividad tendrá una respuesta débil en términos de *drawdowns*, lo que dificulta la estimación del campo K debido a la relación señal/ruido.

En cuanto al desempeño de cada algoritmo, VSAFT2 muestra una mayor dependencia que GM en el número de pruebas de bombeo para reducir el error de estimación de K , además VSAFT2 tiene una correlación menor que GM entre el

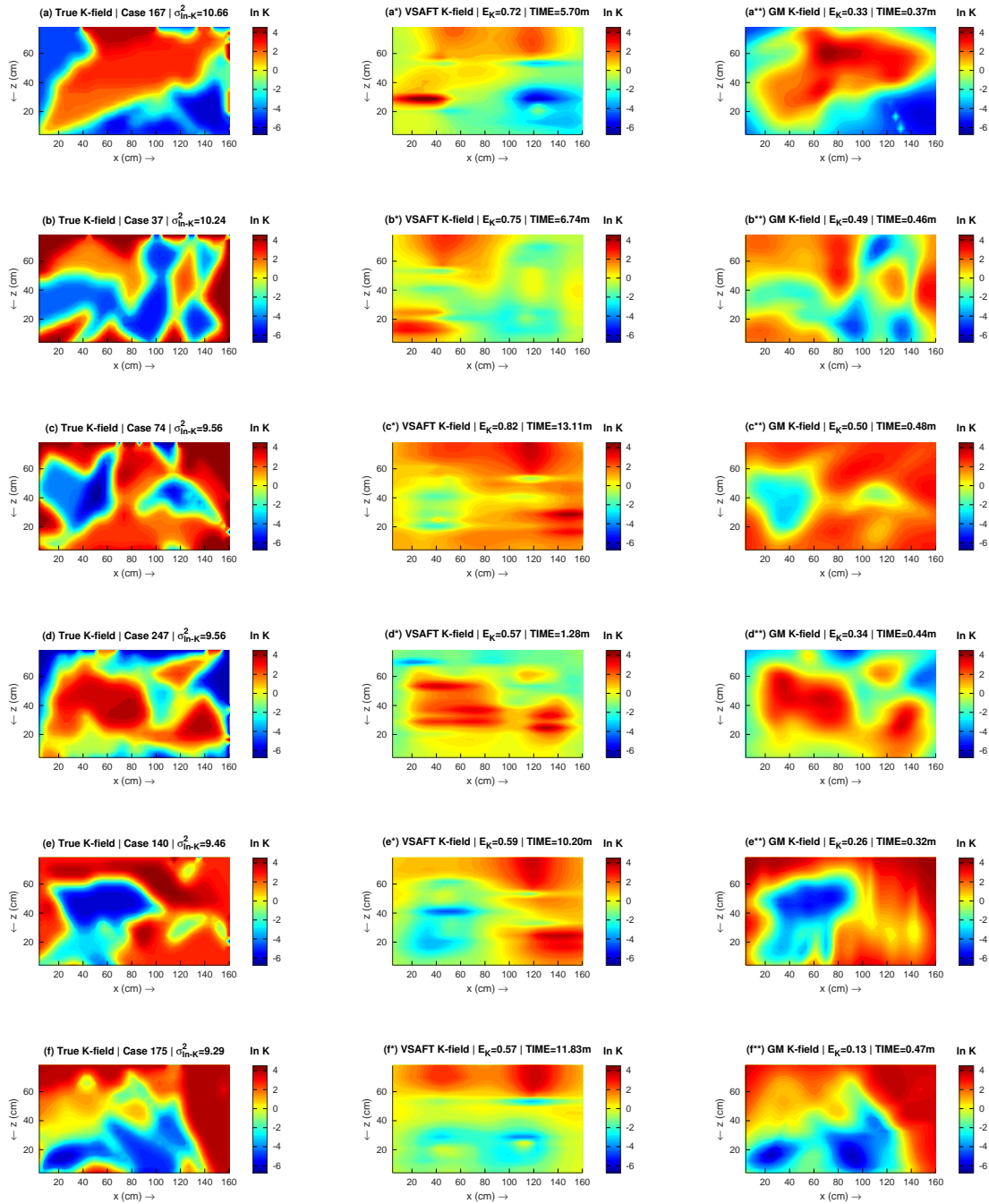


Figura 1.4: Los campos K verdaderos y los tomogramas $\ln K$ obtenidos con VSAFT² y GM cómo métodos de inversión para los 6 casos con mayor varianza de $\ln K$, utilizando 4 pruebas de bombeo con 48 valores de altura por prueba.

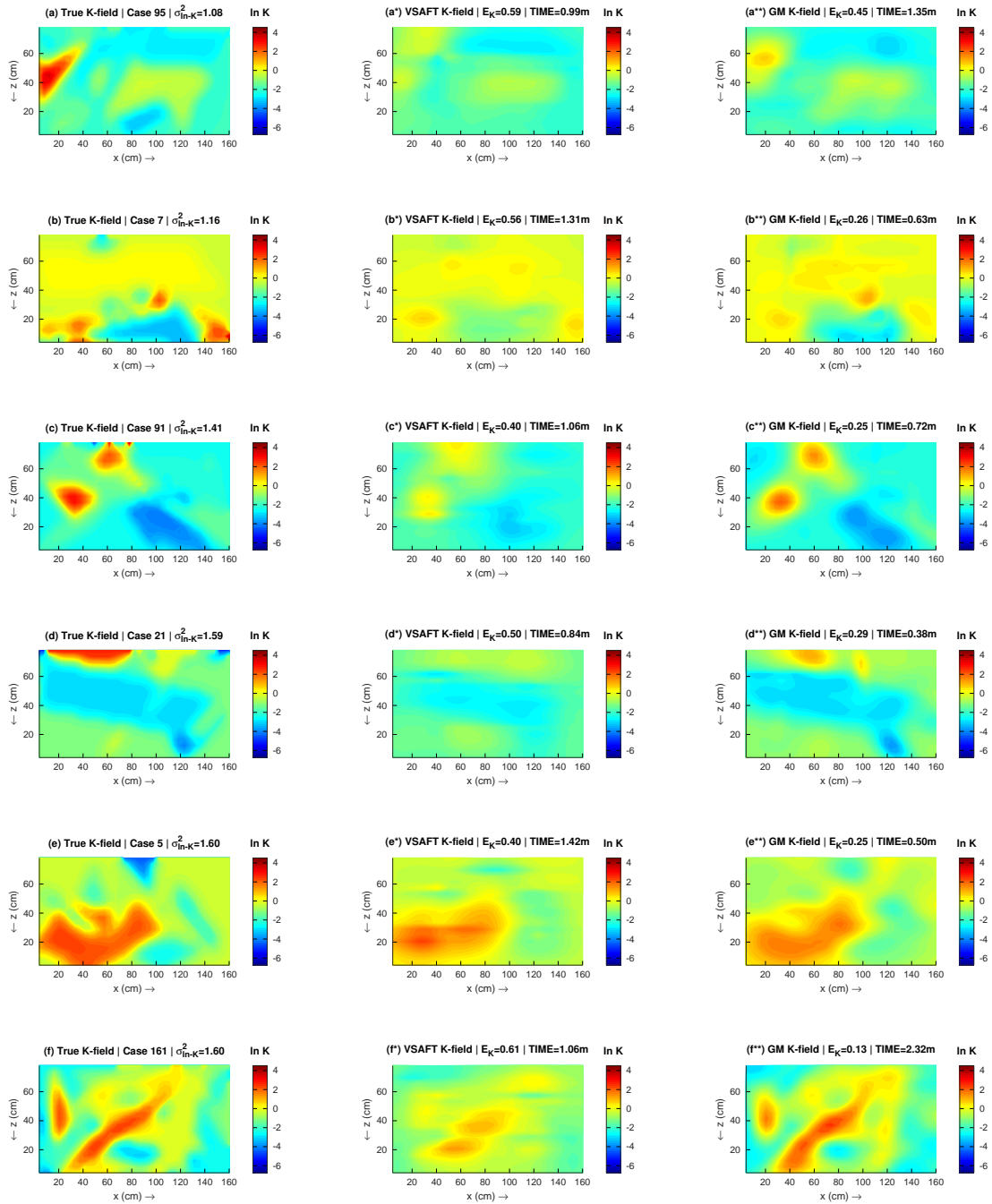


Figura 1.5: Los campos K verdaderos y los tomogramas $\ln K$ obtenidos con VSAFT² y GM cómo métodos de inversión para los 6 casos con menor varianza de $\ln K$, utilizando 4 pruebas de bombeo con 48 valores de altura por prueba.

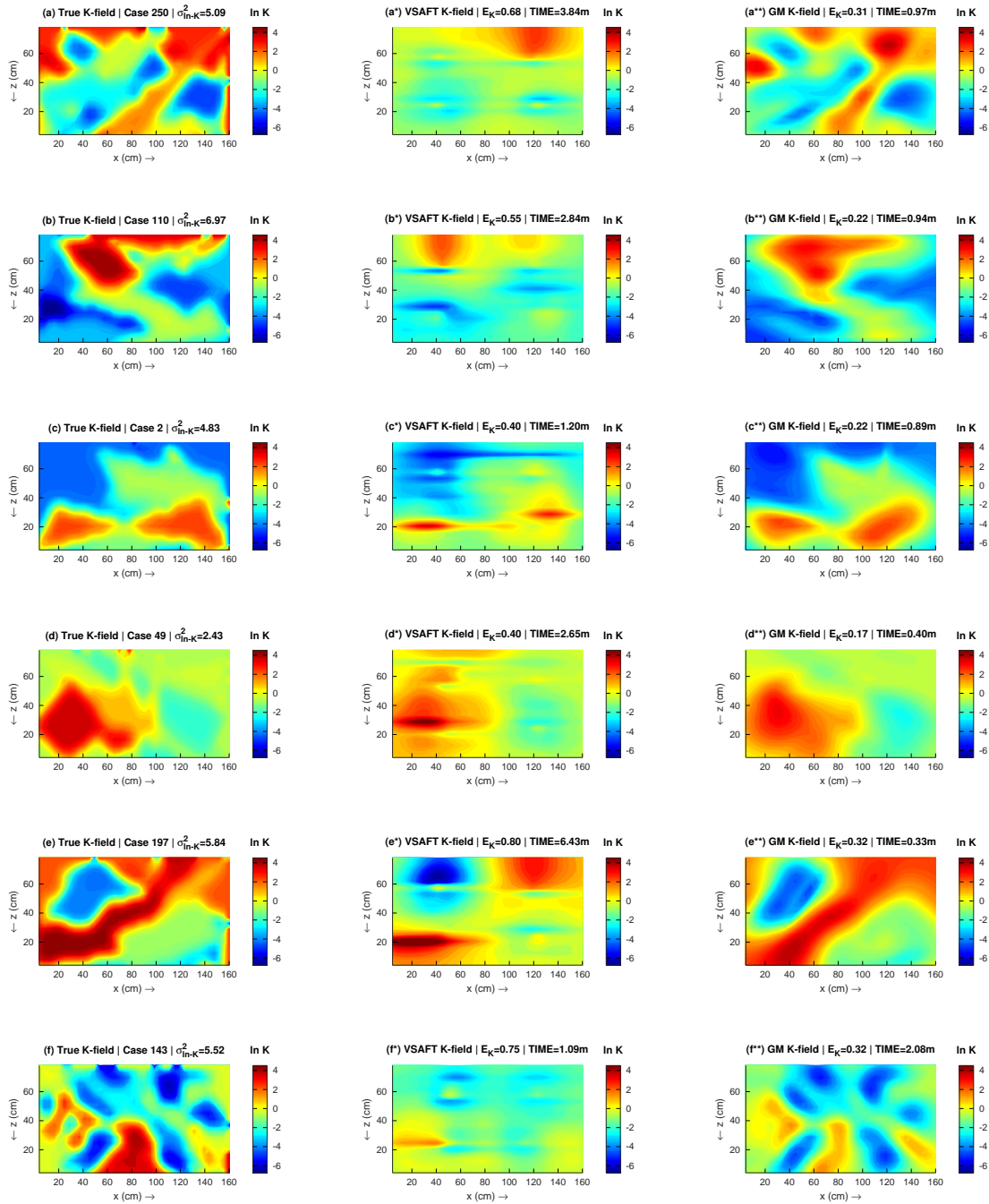


Figura 1.6: Los campos K verdaderos y los tomogramas $\ln K$ obtenidos con VSAFT2 y GM como métodos de inversión de una muestra aleatoria de 6 casos, utilizando 4 pruebas de bombeo con 48 valores de altura por prueba.

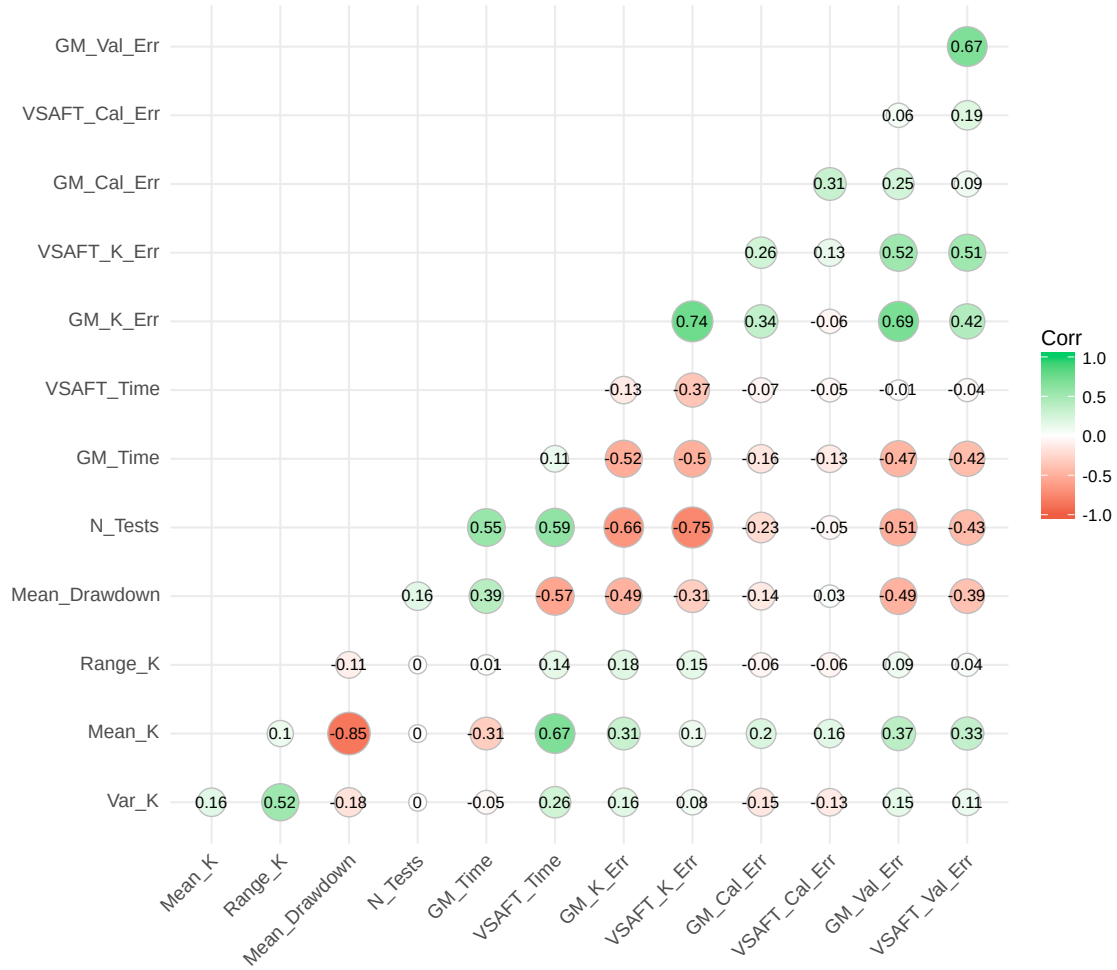


Figura 1.7: Correlograma de los errores de estimación de K (GM_K_Err, VSAFT_K_Err), Errores para los *drawdowns* en el conjunto de calibración (GM_Cal_Err, VSAFT_Cal_Err), Errores para los *drawdowns* en el conjunto de validación (GM_Val_Err, VSAFT_Val_Err), Varianza, Media y Rango de K (Var_K, Mean_K, Range_K), *Drawdown* promedio en el conjunto de calibración (Mean_Drawdown), Número de pruebas de bombeo (N_Tests) y Tiempo de cómputo (GM_Time, VSAFT_Time) de los 1,000 experimentos, para el algoritmo GM y VSAFT2.

error de estimación de K y el error de los *drawdowns* predichos, lo que sugiere que, para un error de estimación similar de K , GM tiende a tener un error menor para los *drawdowns* de validación que VSAFT2, lo que resultará en una mejor predicción de flujo y transporte.

La Figura 1.8 presenta el error de estimación promedio de K (E_K) para VSAFT2 y GM cuando se consideran diferentes números de casos y un número diferente de pruebas de bombeo ($n_tests = 1, \dots, 4$). $E_K = 1$ es equivalente al error de estimación para un campo K estimado mediante el valor \bar{K} (acuífero homogéneo).

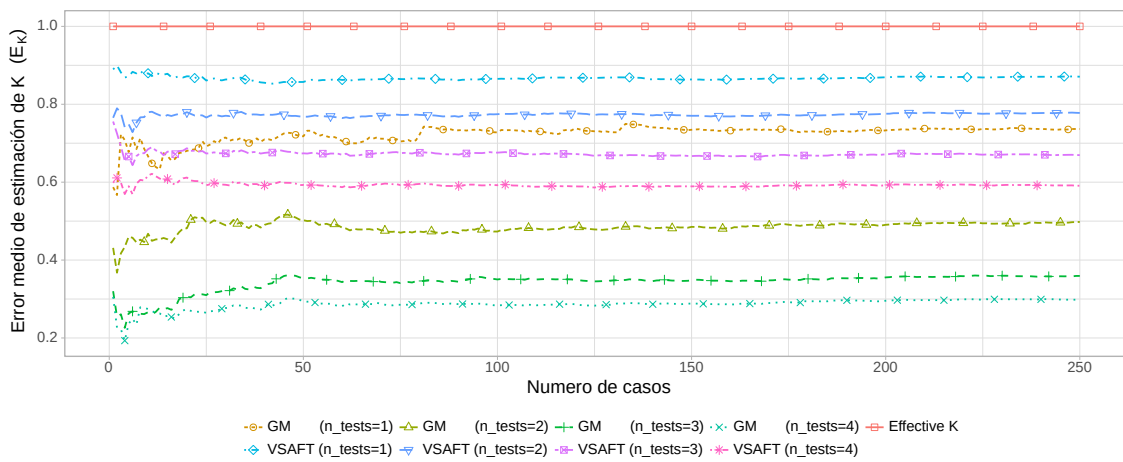


Figura 1.8: Error de estimación promedio de K (E_K) cuando GM y VSAFT2 son usados como métodos de inversión, para 1-4 pruebas de bombeo (n_test).

El error medio parece estabilizarse con aproximadamente 150 casos, lo que sugiere que este valor puede utilizarse como una estimación robusta del error esperado para cada algoritmo. La Figura 1.8 muestra que el error medio más bajo es para el algoritmo de inversión GM con 4 pruebas de bombeo seguido de los casos de 3 pruebas y 2 pruebas, siguiéndole el caso de 4 pruebas con VSAFT2. La mejora menos significativa en la estimación del campo K corresponde a VSAFT2 con 1 prueba de bombeo. También se observa que incluso para una sola prueba de bombeo, hay una mejora en la estimación del campo K , comparado con el valor efectivo de K .

La Figura 1.9 presenta box-plots de E_K para los 250 casos de cada algoritmo y

número de pruebas de bombeo, mostrando los percentiles 25, 50 y 75, así como los valores atípicos. Las medianas muestran que GM tiende a tener errores de estimación más pequeños que VSAFT2, pero también tiene varianzas más altas, especialmente cuando sólo hay una prueba de bombeo disponible para el modelado inverso.

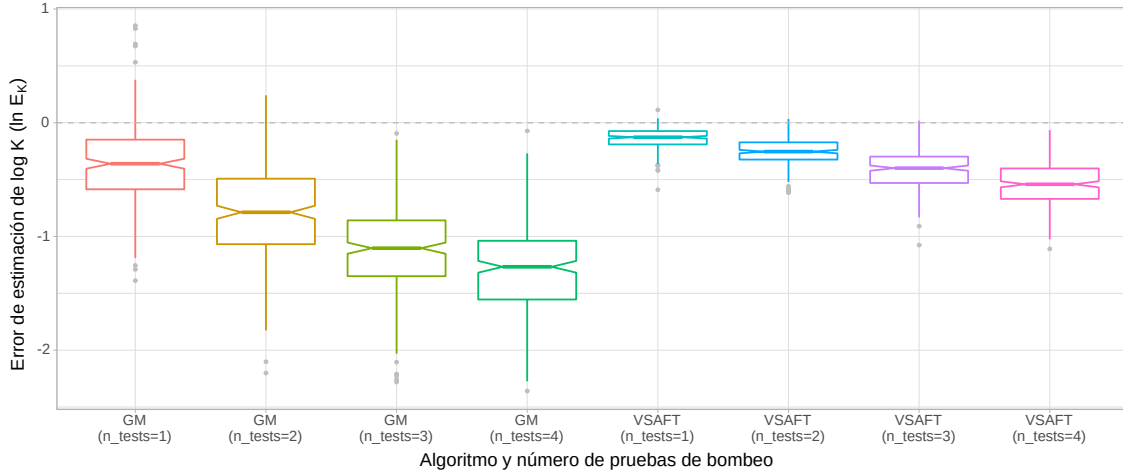


Figura 1.9: Box-plots de los errores estimados de K (E_K) para los 250 campos sintéticos de K , cuando GM y VSAFT2 se utilizan como métodos de inversión, para 1-4 pruebas de bombeo (n_tests). La línea horizontal discontinua representa el error de un K efectivo como campo K estimado.

La eficiencia de cada algoritmo es estimada por:

$$\text{eficiencia}(alg) = \frac{1 - E_K(alg)}{\text{Tiempo de cómputo}} \quad (1.15)$$

que describe cuánto se redujo el error a partir de la solución inicial por cada unidad de tiempo de cómputo (segundos).

La Figura 1.10 presenta la eficiencia media de cada algoritmo, cuyos valores se estabilizan después de promediar más de 200 casos. Los experimentos resueltos con VSAFT2 y GM presentan una diferencia significativa en eficiencia. En particular, GM con 2 y 3 pruebas de bombeo tiene la mayor eficiencia, seguido de cerca por GM con 4 pruebas de bombeo. En contraste, VSAFT2 con 1-4 pruebas de bombeo tiene la eficiencia más baja.

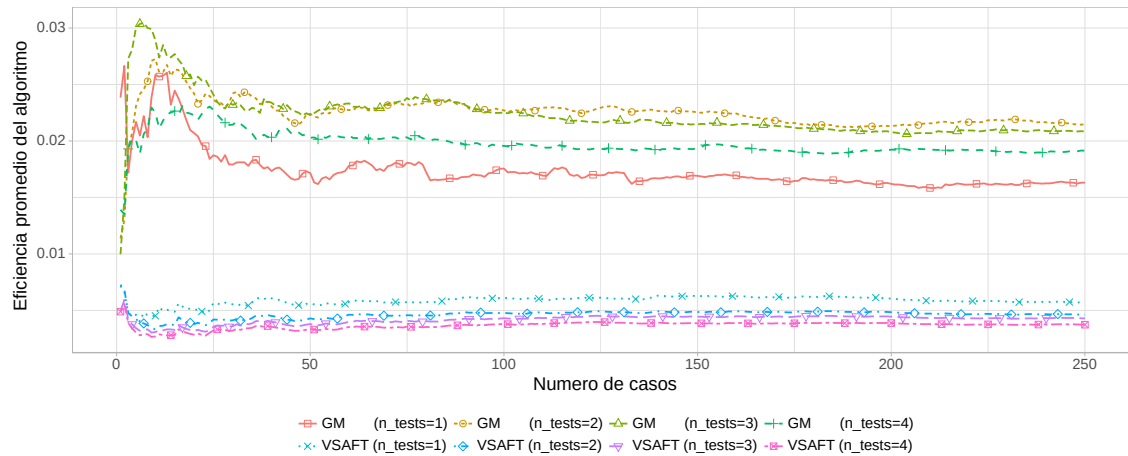


Figura 1.10: Eficiencia media del algoritmo (reducción de error por segundo de tiempo de cómputo) cuando GM y VSAFT2 se utilizan como métodos de inversión, para 1-4 pruebas de bombeo (n_tests).

En cuanto a la predicción de *drawdowns* utilizando los campos K estimados, la Figura 1.11 presenta los box-plot de los errores de *drawdowns* para el conjunto de datos de validación. Al igual que en el caso de los campos K estimados, los errores de GM tienen mayor varianza que los de VSAFT2, pero GM tiende a tener errores significativamente menores. Para VSAFT2, el aumento en el número de pruebas de bombeo tiene un impacto menor en la reducción de errores para *drawdowns* predichos.

Aunque ambos algoritmos producen valores atípicos en los que el campo K homogéneo (K efectivo) produce mejores predicciones de *drawdowns*, en general, los errores disminuyen para ambos métodos y estos valores atípicos se vuelven menos frecuentes cuando se incrementa el número de pruebas de bombeo. Para VSAFT2 94.4% de los casos tienen mejores predicciones de *drawdowns* que el caso homogéneo, y el 97.2% para GM, cuando se utilizaron 4 pruebas de bombeo.

El correlograma en la Figura 1.7 revela que los casos con altos errores de predicción de *drawdowns* están relacionados con altos valores de K , lo que genera *drawdowns* pequeños a lo largo del acuífero debido a que los cambios en la presión se dividen y propagan rápidamente, lo que da lugar a respuestas débiles por parte del acuífero.

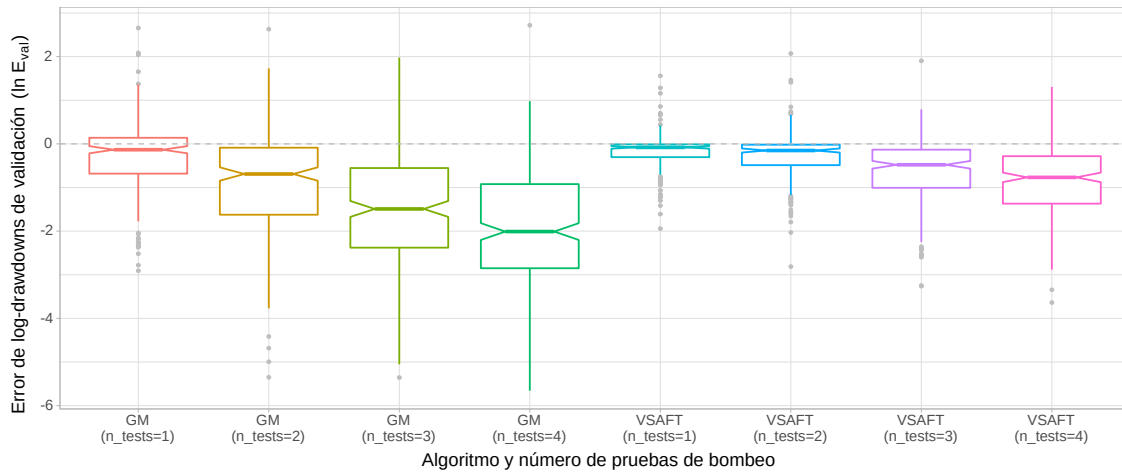


Figura 1.11: Box-plots de los errores de *drawdowns* de validación para los 250 campos K sintéticos, cuando GM y VSAFT2 se utilizan como métodos de inversión, para 1-4 pruebas de bombeo (n_tests). La línea horizontal discontinua representa el error de un K efectivo como campo K estimado.

La Figura 1.12 presenta los box-plots de los tiempos de cómputo para cada conjunto de experimentos, teniendo GM tiempos significativamente menores y distribuciones más simétricas, lo que sugiere un comportamiento más predecible. En el caso de VSAFT2, los tiempos de cómputo altos están correlacionados con valores altos de K , pero esta correlación es en la otra dirección para GM.

Las Figuras 1.13 y 1.14 presentan los gráficos de dispersión de la *drawdowns* observados vs simulados para los conjuntos de datos de calibración y validación, utilizando los 250 campos K sintéticos, con 4 pruebas de bombeo para la calibración y 5 pruebas para la validación. Los resultados se presentan para los campos K estimados a través de (a) el valor efectivo de K , (b) VSAFT2 y (c) GM.

En cuanto al conjunto de datos de calibración, tanto VSAFT2 como GM reproducen con alta precisión los *drawdowns*. En contraste, para el conjunto de datos de validación, GM fue significativamente mejor en la reproducción de los *drawdowns* con un $R^2 = 0.876$, en comparación con VSAFT2 con $R^2 = 0.747$. Puede verse que los K homogéneos no fueron capaces de reproducir adecuadamente los *drawdowns*, teniendo un efecto similar tanto para calibración como para validación.

La información presentada en las Figuras 1.8-1.12 se resume en el Cuadro 1.1, a

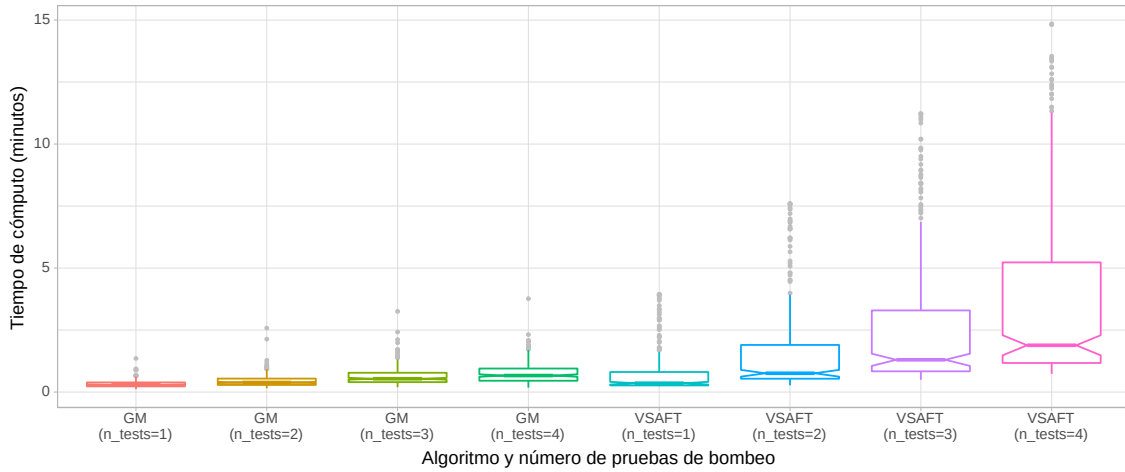


Figura 1.12: Box-plots de los tiempos de cómputo para los 250 campos sintéticos de K , cuando GM y VSAFT2 se utilizan como métodos de inversión, para 1-4 pruebas de bombeo (n_{tests}).

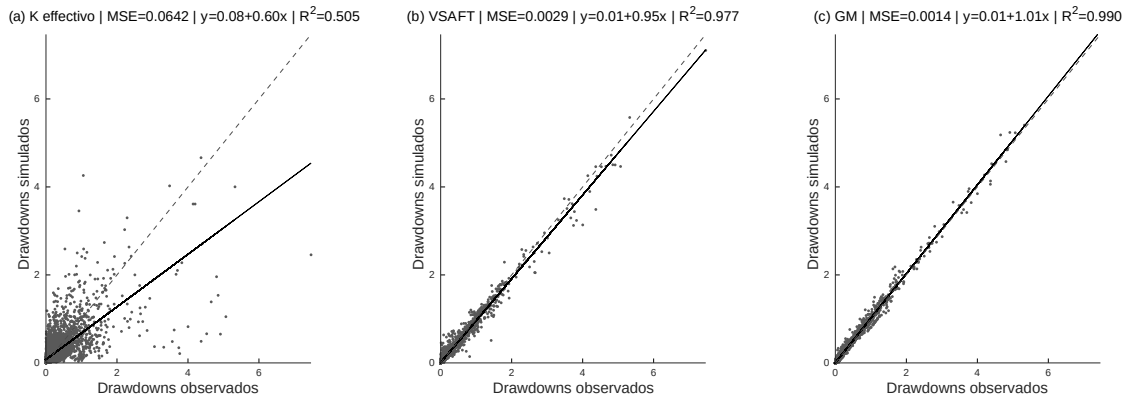


Figura 1.13: Gráficas de dispersión de los *drawdowns* simulados vs observados para el conjunto de datos de calibración de los 250 experimentos, cuando el campo K se estima usando (a) el K efectivo, (b) VSAFT2 y (c) GM, con 4 pruebas de bombeo y 48 valores de presión por prueba. La línea de 45 grados representa una coincidencia perfecta. La línea sólida corresponde a una regresión lineal, mostrando el intercepto y el coeficiente de determinación (R^2).

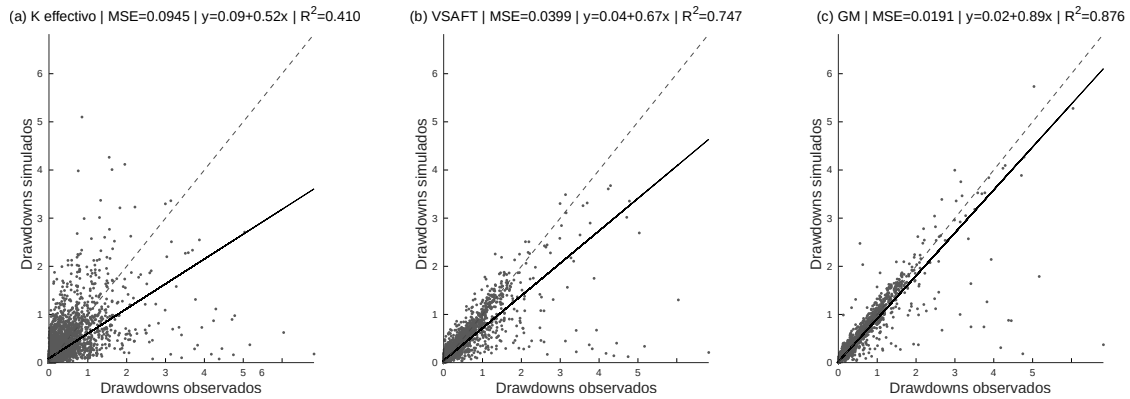


Figura 1.14: Gráficas de dispersión de los *drawdowns* simulados vs observados para el conjunto de datos de validación de los 250 experimentos, cuando el campo K se estima usando (a) el K efectivo, (b) VSAFT2 y (c) GM, con 4 pruebas de bombeo y 48 valores de presión por prueba. La línea de 45 grados representa una coincidencia perfecta. La línea sólida corresponde a una regresión lineal, mostrando el intercepto y el coeficiente de determinación (R^2).

través de las medianas de las diferentes medidas de error, así como los tiempos de cómputo y las eficiencias para el problema de TH de los 250 campos K sintéticos.

En cuanto al error de los campos K estimados, al utilizar 2 pruebas de bombeo, el algoritmo de inversión GM fue capaz de lograr un error menor que VSAFT2 cuando utiliza 4 pruebas de bombeo, además GM requirió sólo el 20% de tiempo de cómputo. Para 4 pruebas de bombeo, GM es 5 veces más eficiente que VSAFT2, y obtuvo el 29% del error que tiene VSAFT2 en términos de *drawdowns* predichos.

Es de notar que, cuando el número de pruebas de bombeo aumenta, las diferencias entre ambos métodos son más significativas para los *drawdowns* predichos, a favor del algoritmo GM, además, el correlograma de la Figura 1.7 también presenta una relación más fuerte entre la precisión de los campos K estimados por el algoritmo GM y los *drawdowns* predichos. Esta característica parece deberse a la naturaleza de la parametrización del MMG, que produce valores continuos y monótonos de disminución/incremento de K a partir del *centroide* de cada capa de material, funcionando como una interpolación de valores, que parece estar más relacionada con la naturaleza de las capas de materiales, resultando en tomografías de K más

Cuadro 1.1: Mediana de los errores de K y de *drawdowns* para el problema de TH, usando 250 casos sintéticos.

N_{tests}	Algoritmo	E_K	E_{Cal}	E_{Val}	Tiempo	Eficiencia
1	VSAFT2	0.8802	0.0636	0.9224	0.3573	0.0041
	GM	0.6973	0.0397	0.8745	0.2907	0.0163
2	VSAFT2	0.7758	0.0689	0.8589	0.7575	0.0039
	GM	0.4549	0.0275	0.5013	0.3852	0.0215
3	VSAFT2	0.6708	0.0595	0.6193	1.3010	0.0038
	GM	0.3318	0.0179	0.2250	0.5365	0.0193
4	VSAFT2	0.5819	0.0582	0.4640	1.8861	0.0034
	GM	0.2817	0.0142	0.1340	0.6654	0.0174

N_{tests} =Número de pruebas de bombeo utilizadas, Algoritmo=Método de inversión, E_K =Error para los campos K estimados, E_{Cal} =Error para los *drawdowns* de calibración, E_{Val} =Error para los *drawdowns* de validación, Tiempo=Minutos utilizados para estimar el campo K , Eficiencia=Reducción del error de K por segundo de tiempo de cómputo. Un error igual a 1 representa el caso de un campo K homogéneo con valor \bar{K} .

consistentes y nítidas. Además, el algoritmo de inversión GM utiliza el campo de conductividad de Bayes como su estimación final de K , que es el consenso de las diferentes soluciones encontradas por el algoritmo.

1.4.1. Caso no Gaussiano / Acuífero fracturado

Para probar el método propuesto, con un acuífero no Gaussiano, se creó un acuífero sintético fracturado ($\sigma_{\ln K}^2 = 0.3271$), el cual se presenta en la Figura 1.15a.

En el Cuadro 1.2 se muestran las diferentes medidas de error y el tiempo de cómputo para los campos de conductividad estimados utilizando ambos métodos de inversión. Se puede observar que se necesitó una mayor cantidad de datos para obtener buenas estimaciones del campo de conductividad.

Además, para este acuífero las diferencias entre ambos algoritmos parecen ser más significativas. Utilizando 10 pruebas de bombeo, GM produce una quinta parte del error que VSAFT tiene, y requiere sólo el 15% del tiempo de cómputo. Además, el error obtenido con VSAFT con 10 pruebas de bombeo es similar al error que GM tiene con sólo 4 pruebas de bombeo.

En las Figuras 1.15b y 1.15c se presentan los tomogramas (campos K estimados) para cada algoritmo de inversión y se puede observar que aunque VSAFT es capaz

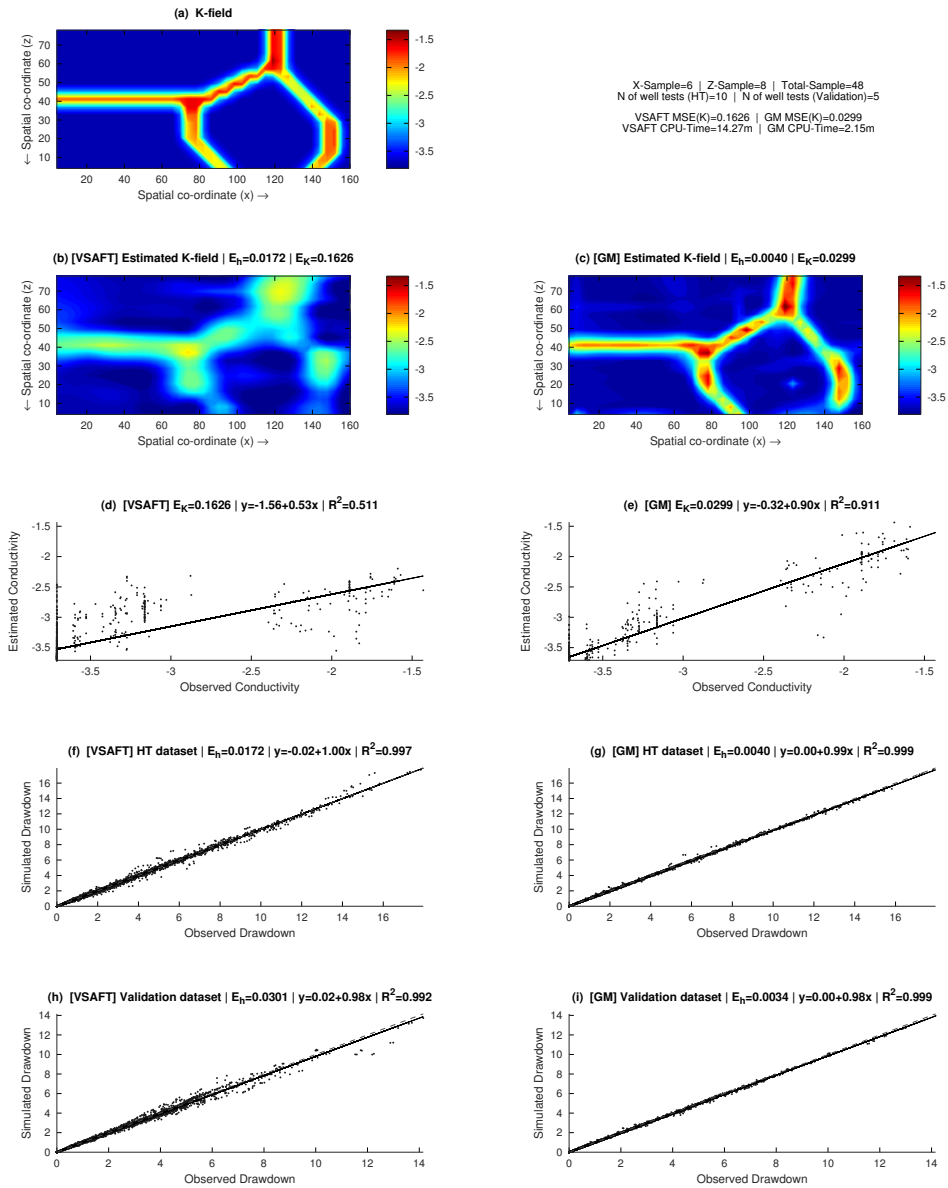


Figura 1.15: (a) Campo K sintético de un acuífero fracturado y los campos K estimados con (b) VSAFT2 y (c) GM como métodos de inversión, utilizando 10 pruebas de bombeo con 48 valores de presión por prueba.

Cuadro 1.2: Tiempo de cómputo y diferentes medidas de error para el problema de TH, en un acuífero sintético fracturado ($\sigma_{\ln K}^2 = 0.3271$).

N_{tests}	Algoritmo	$\hat{\sigma}_{\ln K}^2$	E_K	E_{Cal}	E_{Val}	Tiempo
4	VSAFT2	0.1154	0.9560	0.0454	0.1015	4.85
	GM	0.2654	0.5167	0.0121	0.0261	0.98
6	VSAFT2	0.1706	0.8040	0.0261	0.0670	7.46
	GM	0.2166	0.3922	0.0166	0.1283	1.93
8	VSAFT2	0.1536	0.5940	0.0197	0.0498	10.22
	GM	0.2461	0.1755	0.0122	0.0146	1.43
10	VSAFT2	0.1799	0.4971	0.0172	0.0301	14.27
	GM	0.2913	0.0914	0.0040	0.0034	2.15

N_{tests} =Número de pruebas de bombeo utilizadas, Algoritmo=Método de inversión, $\hat{\sigma}_{\ln K}^2$ =Varianza del campo K estimado, E_K =Error para los campos K estimados, E_{Cal} =Error para los *drawdowns* de calibración, E_{Val} =Error para los *drawdowns* de validación, Tiempo=Minutos utilizados para estimar el campo K .

de detectar las principales áreas de conductividad, GM reproduce mejor las fracturas, siendo más clara la forma de éstas.

La capacidad de GM para detectar rápidamente las principales áreas de conductividad utilizando los cambios en los valores de la presión y para producir formas complejas mediante la combinación de múltiples núcleos Gaussianos, parece tener un mayor impacto en este tipo de acuíferos.

1.4.2. Caso real / Caja de arena

El Cuadro 1.3 presenta los resultados de la estimación del campo K para el experimento de caja de arena realizado por Illman et al. (2010), utilizando VSAFT y GM como métodos de inversión.

Se puede observar que con 8 pruebas de bombeo la cantidad de tiempo requerida con GM es casi una quinta parte del tiempo utilizado con VSAFT y la varianza del campo K estimado es similar en ambos casos, así como el error de los valores de presión.

Aunque no es posible calcular el error de estimación del campo K , hay algunas indicadores de que tan buena es la estimación. Por ejemplo, el valor medio de $\ln K$ se reporta en el Cuadro 1.3 como $\ln K_G$, y en ambos casos, las estimaciones son

similares al valor de -2.56 , reportado en [Illman et al. \(2010\)](#), usando 48 muestras de núcleo. Aunque las varianzas de $\ln K$ son ligeramente diferentes al valor estimado usando muestras de núcleo ($\sigma_{\ln K}^2 = 0.868$), la varianza obtenida con el algoritmo GM es más cercana a ese valor.

Cuadro 1.3: Tiempo de cómputo y diferentes medidas de error para el problema de TH en un experimento de caja de arena.

N_{tests}	Algoritmo	$\hat{\sigma}_{\ln K}^2$	E_{Cal}	E_{Val}	$\ln K_G$	Tiempo
2	VSAFT2	0.7053	0.0102	0.0552	-2.28	2.58
	GM	0.7498	0.0088	0.0521	-2.34	0.92
4	VSAFT2	0.8424	0.0066	0.0720	-2.32	5.70
	GM	0.7611	0.0105	0.0509	-2.28	1.18
8	VSAFT2	1.1272	0.0081	0.0464	-2.31	10.83
	GM	1.0370	0.0054	0.0417	-2.35	2.40

N_{tests} =Número de pruebas de bombeo utilizadas, Algoritmo=Método de inversión, $\hat{\sigma}_{\ln K}^2$ =Varianza del campo K estimado, E_{Cal} =Error para los *drawdowns* de calibración, E_{Val} =Error para los *drawdowns* de validación, $\ln K_G$ es el valor medio del campo $\ln K$, Tiempo=Minutos utilizados para estimar el campo K .

La Figura 1.16 muestra los tomogramas de K generados con cada método de inversión y una fotografía del acuífero sintético que es superpuesta. Las áreas de altas K se indican en rojo, donde es posible ver que cada una de estas áreas corresponde a una capa diferente de arena. Un valor estimado de K de cada tipo de arena se presenta en [Illman et al. \(2010\)](#) (Tabla 1), donde los cuatro tipos de arena con K más altas incluyen (de K más altas a más bajas), las arenas 16/30, 20/30, 20/40 y #12. Examinando la Figura 1.16 se observa que la arena 16/30 es el único tipo que no se muestra en rojo en los tomogramas de K . Esto podría deberse al hecho de que este tipo de arena tiene el segundo volumen más pequeño en el experimento.

VSAFT detectó cómo áreas de alta K todas las capas con tipos de arena 20/30, 20/40 y #12 con la única excepción de la capa 13, que es una arena 20/30. Sin embargo, el algoritmo GM fue capaz de detectar la capa 13 cómo un área de alta K . En el caso de VSAFT, no es hasta [Berg and Illman \(2011\)](#) que se detecta la capa 13, usando tomografía hidráulica transitoria, la cual requiere más datos y significativamente más tiempo de cómputo.

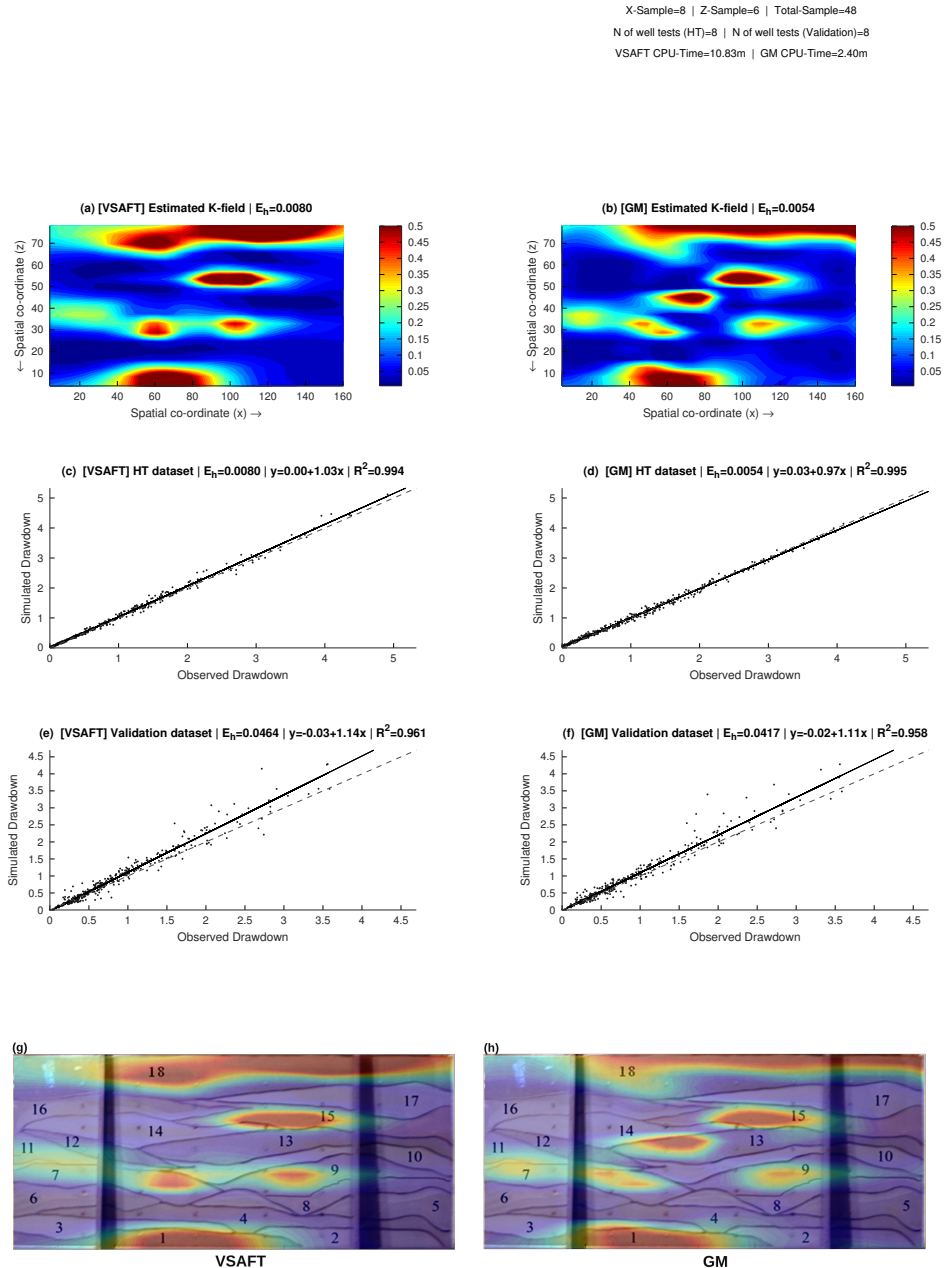


Figura 1.16: Campos K estimados para el experimento de caja de arena, obtenidos con (a) VSAFT2 y (b) GM como métodos de inversión, utilizando 8 pruebas de inyección con 47 valores de presión por prueba. Imagen superpuesta de los tomogramas y la fotografía del arenero (g)-(h). La línea de puntos de 45 grados representa una coincidencia perfecta.

1.5. Resumen y conclusiones

La tomografía hidráulica se ha convertido en uno de los métodos más robustos para caracterizar la heterogeneidad de los parámetros hidráulicos como la conductividad hidráulica, que es indispensable para la predicción de flujo y transporte. No obstante, se necesita un tiempo de cómputo elevado y una cantidad significativa de datos para obtener buenas estimaciones del campo K . Un método altamente parametrizado como un algoritmo de inversión geoestadística actualiza todos los valores de K durante cada iteración, lo que podría resultar en una actualización más lenta del campo K , conduciendo a tomogramas menos definidos.

El desarrollo de nuevos algoritmos de inversión que hagan mejoras en este aspecto son esenciales para continuar explorando y desarrollando los experimentos de TH. El objetivo principal de este trabajo fue proponer, y probar extensivamente un nuevo método de inversión para reducir la carga computacional y aumentar la precisión de los tomogramas de K .

Para lograr esto, se hizo uso de una nueva forma de parametrizar el campo K . La parametrización del MMG implementada en el algoritmo de inversión GM es capaz de cambiar rápidamente la complejidad del campo K generado a través de pequeños cambios de los parámetros (por ejemplo, el peso de un componente y su ubicación). Esta parametrización del campo K juega un papel importante en el modelado, permitiendo que el algoritmo de inversión reduzca el número de parámetros a ser estimados, resultando en una optimización más rápida y al mismo tiempo, posee parámetros más sensibles, los cuales pueden ser estimados más temprano en el proceso de optimización, produciendo un rápido “esquema” del campo K , usando pocas iteraciones del algoritmo. Además, la parametrización MMG puede controlar la complejidad de la distribución espacial del campo K estimado, cambiando el parámetro N_k , pudiendo describir distribuciones K muy complejas cuando N_k es grande o sólo detectar y agrupar las zonas principales de alto y bajo K cuando N_k es pequeño. La precisión se incrementó al incluir una nueva función objetivo que utiliza las derivadas espaciales, así como un nuevo estimador Bayesiano del campo K , que toma el consenso de las diferentes soluciones.

Con un gran conjunto de campos K sintéticos en los que se variaron las formas, varianzas, rangos y conductividades, se realizaron múltiples experimentos con diferentes números de pruebas de bombeo, lo que permitió realizar un estudio exhaustivo

de los efectos que podrían tener un impacto en el rendimiento de los métodos de TH. Las mediciones de los errores mostraron una convergencia a lo largo del número de experimentos, lo que sugiere resultados robustos.

Este estudio lleva a las siguientes conclusiones principales:

1. Aunque todos los resultados de los métodos de TH presentaron un aumento en la precisión de los campos de K estimados y la predicción de *drawdowns* (incluso con sólo una prueba de bombeo), el algoritmo de inversión GM produjo mejores estimaciones de los campos K . En particular, cuando se utilizaron 4 pruebas de bombeo para el modelado inverso, el algoritmo GM dio como resultado el 48 % del error que produjo VSAFT2, y el 29 % del error para los *drawdowns* de validación, requiriendo sólo el 35 % del tiempo de cómputo.
2. Estos resultados también sugieren que el algoritmo de inversión GM permite reducir la cantidad de datos necesarios para realizar TH, donde, para los campos K estimados, el algoritmo GM logró un error menor que VSAFT2, al utilizar 2 pruebas de bombeo en lugar de 4, con un 20 % del tiempo de cómputo.
3. La distribución de los tiempos de cómputo reveló una mejora significativa en este aspecto para el algoritmo de inversión GM, que presentaba distribuciones más simétricas con menos valores atípicos, lo que sugiere tiempos de cómputo más predecibles, independientemente de la naturaleza del problema.
4. La parametrización del MMG y el estimador Bayesiano de los campos K parecen producir campos K más consistentes con la naturaleza de los acuíferos, con diferentes capas de materiales, a través de la generación de una solución por consenso, resultando en mejores predicciones de *drawdowns*.
5. El correlograma mostró que una de las variables que más afectaba el rendimiento de los métodos de TH era el valor efectivo de K , donde los valores altos de K en el acuífero producen *drawdowns* pequeños, resultando en respuestas débiles del acuífero, lo que requerirá aumentar la tasa de bombeo/inyección para mejorar la relación señal/ruido.
6. En general, los resultados sugieren que el algoritmo de inversión GM es robusto y aplicable a casos reales de experimentos de TH, mejorando los tiempos de

cómputo y precisión.

Parte II

Caracterización de yacimientos de petróleo

Capítulo 2

Optimización ad-hoc para el ajuste de curvas tipo mediante un modelo de flujo de triple porosidad y geometría fractal

Las pruebas de pozos son hoy en día un procedimiento fundamental para verificar las propiedades de los yacimientos, pozos y del fluido, tales como la permeabilidad, el efecto *skin*, la presión media del yacimiento, las heterogeneidades, los límites y la conectividad hidráulica. En este trabajo se presenta una metodología basada en la optimización numérica y en métodos estadísticos para obtener una caracterización precisa y fiable de yacimientos carbonatados vulgares naturalmente fracturados, utilizando un Modelo Fractal de Triple Porosidad. Se realiza un análisis de sensibilidad para identificar el efecto y la magnitud sobre la presión de cada parámetro involucrado en el modelo, para así reducir la estimación de errores y eliminar múltiples interpretaciones debido a la precisión numérica. Los resultados se ilustran con pruebas de pozos transitorios de yacimientos reales de México.

El análisis automatizado de pruebas de pozos (AWTA, por sus siglas en inglés), también llamado ajuste automático de curvas tipo, es hoy en día uno de los principales componentes de la interpretación de pruebas de pozos.

A pesar de los avances en teorías matemáticas como la regresión no lineal, la optimización global y el análisis numérico y la existencia de potentes ordenadores,

el AWTA se sigue llevando a cabo con un bajo nivel de automatización real. Los ingenieros generalmente siguen un enfoque de prueba y error cuando usan software de prueba de pozos comerciales.

En el presente trabajo se presenta una metodología para desarrollar un algoritmo de optimización *ad-hoc* al problema, basándose en los análisis de sensibilidad, lo que permite alcanzar un mayor grado de automatización en la caracterización de yacimientos, incluso para modelos altamente complejos (con múltiples parámetros desconocidos), usando cómo ejemplo un modelo de flujo de triple porosidad y geometría fractal, para reservorios vulgares naturalmente fracturados.

La metodología presentada y resultados comparativos cuando la metodología propuesta no es utilizada, fueron publicados (utilizando un modelo con un menor número de parámetros que el aquí presentado) en [Minutti et al. \(2016\)](#).

2.1. Descripción del problema

El petróleo se encuentra almacenado dentro de un medio poroso (roca), que presenta tres posibles medios de flujo: fracturas, vóculos (cavidades) y matriz (roca porosa propia de la roca almacén). El flujo a través de estos medios (y por lo tanto la presión en el pozo) depende de varios parámetros, por ejemplo las proporciones presentes de aceite en cada uno de los medios, o la facilidad que tiene el petróleo de moverse en cada uno de ellos (permeabilidad).

Cuando se perfora un pozo en un yacimiento petrolero, es posible realizar una prueba de presión del pozo, la cual consiste en cerrar la producción del mismo y registrar el aumento de presión.

Al considerar aspectos cómo la geometría del pozo, la estructura del medio poroso y las características de flujo del petróleo, se puede generar un modelo matemático que a partir de los parámetros que describen el yacimiento explica cómo se ve afectada la presión en el pozo a lo largo del tiempo.

Es por medio de estos parámetros que describen al yacimiento que es posible estimar la presión a tiempos mayores, y por medio de esta presión, determinar la productividad del pozo. Aunque estos parámetros no se conocen a priori, es posible estimarlos al comparar los distintos valores de presión que genera el modelo matemático contra los datos medidos durante la prueba de presión en el pozo; por lo tanto es necesario un modelo que sea capaz de describir adecuadamente los cambios

de presión en el pozo y un método para estimar los parámetros del modelo que corresponden a los datos observados.

El enfoque estándar para el análisis automatizado de pruebas de pozos (AWTA) en la literatura de ingeniería petrolera se basa en los métodos de estimación de parámetros de mínimos cuadrados no lineales (NLS) o de mínimo valor absoluto (LAV, ver [Bonalde and Ramones \(1994\)](#), [Menekse et al. \(1995\)](#) y [Vieira and Rosa \(1996\)](#)). Los problemas de NLS y LAV son problemas de optimización y requieren para su solución de métodos robustos de optimización global (ver [Gomez et al. \(2001\)](#) y [Gomez et al. \(2014b\)](#)). Los métodos locales, como el conocido Levenberg-Marquardt, no pueden garantizar la convergencia con los valores óptimos de los parámetros cuando se optimizan funciones objetivo complejas como las que se obtienen a partir de modelos de flujo complejos. Los métodos de regresión no lineal se han utilizado durante mucho tiempo en AWTA para calcular las regiones de confianza para las estimaciones de parámetros (véase [Joseph and Koederitz \(1983\)](#)). En los últimos años, estos métodos han recibido mayor atención (véase [Dastan and Horne \(2010\)](#)). El análisis de sensibilidad e incertidumbre (otra teoría matemática bien establecida para el estudio de cómo la incertidumbre en el resultado de un modelo puede relacionarse con diferentes fuentes de incertidumbre en sus entradas), se ha utilizado principalmente en history matching (ver por ejemplo [Genrich and Sommer \(1989\)](#), [Watson \(1989\)](#) y [Schiozer et al. \(2005\)](#)).

Los datos usados en AWTA son generalmente la presión transitoria registrada en un pozo *versus* el tiempo cuando el pozo es bombeado a una tasa constante (*Drawdown*) o cerrado (*Buildup*). La interpretación de estas pruebas se ve facilitada en gran medida por el uso de la log-derivada ([D. Bourdet and Pirard, 1989](#)). Sin embargo, el cálculo de la derivada de los datos medidos puede ser un problema difícil debido al ruido inherente a los datos procedentes del proceso de medición y del propio dispositivo. En general, el problema de la diferenciación numérica está mal planteado debido a que pequeños errores en los datos pueden causar grandes cambios en la derivada correspondiente. Así, considerando esta información extra contenida en la log-derivada, se puede mejorar la convergencia de los métodos de optimización que logran un buen ajuste a la presión y a su log-derivada.

La metodología propuesta para abordar este problema puede resumirse en los siguientes pasos:

1. Emplear métodos de análisis numérico para resolver eficazmente los proble-

mas de cálculo numérico con las soluciones analíticas o semi-analíticas de los modelos de flujo.

2. Realizar un análisis de sensibilidad en las soluciones del modelo de flujo para identificar las regiones de influencia de los parámetros del modelo para la presión calculada. Un análisis de sensibilidad también permite determinar las correlaciones entre los parámetros del modelo. Por ejemplo, en la práctica habitual de la ingeniería, el valor del *skin* se determina a partir de cálculo de una línea recta en gráficos semi-log o logarítmicos (véase [Everdingen \(1953\)](#)) conocidos como curvas tipo. Mediante este procedimiento, la línea recta se ajusta sólo a un intervalo determinado de la curva de presión. Esto muestra que la influencia del parámetro del *skin* es más determinante sólo en una determinada región temporal de la curva tipo.
3. Basándose en los resultados del análisis de sensibilidad, se desarrolla un modelo de regresión no lineal más adecuado para la tarea de identificación de parámetros, el cual consiste en una secuencia de optimizaciones ponderadas de NLS determinadas por la región de tiempo de influencia de los diferentes parámetros del modelo.
4. Incluir la log-derivada de los datos sin ruido como parte de las optimizaciones ponderadas de NLS.
5. Utilizar el método de optimización desarrollado para resolver el modelo de regresión no lineal. Las funciones objetivo de cualquier modelo de regresión para datos de presión son muy complejas, pueden mostrar valles planos o pronunciados que dificultan la convergencia de los métodos de optimización para alcanzar el óptimo global. Los métodos locales, como Levenberg-Marquardt, pueden fallar en el cálculo de estimaciones de parámetros significativos.

La metodología descrita en los pasos anteriores permite, dado un modelo de flujo y una curva de presión transitoria, calcular estimaciones de parámetros con un alto grado de confianza de forma automática. Se observó que los pasos uno a tres son necesarios sólo una vez para un modelo de flujo dado.

2.2. El modelo de flujo de triple porosidad y geometría fractal

El estudio de los procesos de flujo en yacimientos carbonatados vugulares naturalmente fracturados ha recibido una mayor atención en los últimos años debido a que se han encontrado varios de estos yacimientos en todo el mundo con importantes reservas y producción de petróleo y gas. Diferentes autores han propuesto modelos de triple porosidad para estos reservorios (Abdassah, 1986; Liu et al., 2002; Wu et al., 2004; Camacho-Velázquez et al., 2005; Wu et al., 2006) cómo extensiones del modelo de doble porosidad de Warren and Root (1963).

Para el presente trabajo se ha utilizado un modelo de geometría fractal de triple porosidad y que asume que hay flujo solo en fracturas, basado en el modelo presentado por Balankin and Elizarraraz (2012). Utilizando variables sin dimensiones, la ecuación diferencial para la presión $p_{Df}(r_D, t_D)$ sobre el sistema de fractura, para un pozo totalmente penetrado, viene dada por:

$$r_D^{1-\gamma-\beta} \frac{\partial}{\partial r_D} \left[(r_D)^{\beta+1} \frac{\partial p_{Df}}{\partial r_D} \right] - \left(\omega_m \frac{\partial p_{Dm}}{\partial t_D} + \omega_v \frac{\partial p_{Dv}}{\partial t_D} + \omega_f \frac{\partial p_{Df}}{\partial t_D} \right) = 0, \quad (2.1)$$

La presión $p_{Dm}(r_D, t_D)$ en los bloques de matriz está conformado por la ecuación:

$$\omega_m \frac{\partial p_{Dm}}{\partial t_D} = \lambda_{fm} (p_{Df} - p_{Dm}) + \lambda_{vm} (p_{Dv} - p_{Dm}), \quad (2.2)$$

y la presión $p_{Dv}(r_D, t_D)$ en la red de vugulos viene dada por la ecuación diferencial:

$$\omega_v \frac{\partial p_{Dv}}{\partial t_D} = \lambda_{fv} (p_{Df} - p_{Dv}) + \lambda_{vm} (p_{Dm} - p_{Dv}), \quad (2.3)$$

Con el siguiente conjunto de condiciones iniciales y de frontera:

$$p_{Di} = 0 \quad (2.4)$$

$$-C_D \frac{\partial p_{Dwf}}{\partial t_D} + \left[(r_D)^{\beta+1} \frac{\partial p_{Df}}{\partial r_D} \right]_{r_D=1} = -1 \quad (2.5)$$

$$p_{Dwf} = \left[p_{Df} - \frac{s}{r_w} (r_D)^{\beta+1} \frac{\partial p_{Df}}{\partial r_D} \right]_{r_D=1} \quad (2.6)$$

con $i = f, m, v$, y donde el conjunto de parámetros a identificar está dado por:

$$\mathbf{x} = (\omega_v, \omega_f, \lambda_{mf}, \lambda_{mv}, \lambda_{vf}, s, C_D, dL, \Delta) \in \mathfrak{R}^9 \quad (2.7)$$

con $\Delta = \gamma/2$, $dL = \beta + 1 + \Delta$, que corresponden a los parámetros de la geometría fractal.

Además, en casos reales, para transformar variables sin dimensiones en cantidades dimensionales, necesitamos dos parámetros más α_t y α_p , los cuales también requieren de estimación.

Dado que el modelo de triple porosidad y geometría fractal incluye múltiples parámetros, el ejercicio de caracterización es más complejo.

2.2.1. El problema de la identificación de parámetros

El objetivo del análisis de una prueba de pozos es determinar estimaciones razonables de los parámetros del modelo. Aunque estos parámetros no se conocen a priori, es posible estimarlos comparando los valores de presión y su derivada logarítmica generada por el modelo matemático (para algunos valores de los parámetros del modelo) con la presión real (y su derivada logarítmica) medida durante la prueba del pozo, minimizando una función objetivo:

$$\begin{aligned} \text{mín } F(\mathbf{x}) &= \text{mín } \|P(\mathbf{x}) - \text{data}\| \\ \text{s.a. } &L \leq \mathbf{x} \leq U \end{aligned} \quad (2.8)$$

donde P es la presión generada por el modelo de flujo a partir del vector de parámetros \mathbf{x} , data los datos medidos durante la prueba de pozo, L y U los límites inferior y superior para \mathbf{x} . En este caso la función objetivo utilizada está dada por:

$$F(\mathbf{x}) = \left(\left\| \frac{P(\mathbf{x}) - \text{data}}{\text{data}} \right\|_2 + 1 \right) \cdot \left(\left\| \frac{P(\mathbf{x})_{der} - \text{data}_{der}}{\text{data}_{der}} \right\|_2 + 1 \right) - 1 \quad (2.9)$$

donde el subíndice *der* hace referencia a la log-derivada de los datos.

La solución óptima de 2.8 da como resultado el valor de los parámetros desconocidos del pozo y del yacimiento.

Sin embargo, puede haber múltiples estimaciones óptimas del problema inverso, la presencia de ruido en los datos y el hecho de que en el modelo se permite que el flujo vaya tanto a las fracturas cómo en los vugulos.

Es así que en el proceso de caracterización, la precisión juega un papel muy importante. Esto puede ser verificado en la Figura 2.1, donde si cambiamos ligeramente un parámetro (ω_v), el efecto en la curvas de la log-derivada es considerable. Así, un pequeño cambio en un parámetro puede producir un gran cambio en la solución del modelo, lo que muestra la necesidad de encontrar los parámetros con una precisión muy alta para poder reproducir los datos medidos.

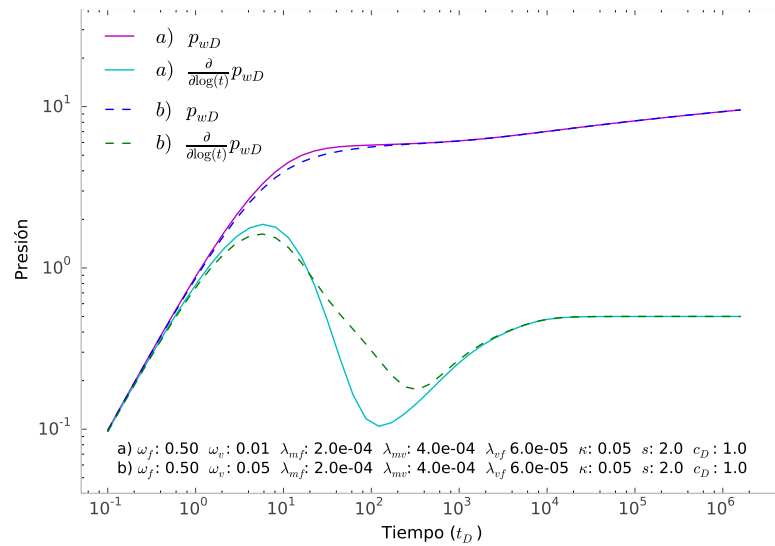


Figura 2.1: Curvas tipo: comparación con la modificación de un parámetro (ω_v).

Para conseguir la precisión necesaria para caracterizar el yacimiento con este modelo de triple porosidad es necesario estudiar el efecto estadístico de cada parámetro en la presión y en qué intervalo de tiempo el efecto es más significativo. Esto se logra mediante un análisis de sensibilidad. Con esta información se diseña una secuencia específica de pasos para mejorar la identificación de los parámetros del modelo.

Un ejemplo de este enfoque se presenta en la Figura 2.2, donde a una función no lineal de 2 parámetros (b y c) determinan que rango de valores de los datos de

entrada tienen mayor efecto en los parámetros a estimar, asignando en la función objetivo un mayor peso a esos datos. El lado izquierdo de la gráfica muestra una función objetivo (y su contorno) donde todos los datos de entrada tienen el mismo peso y del lado derecho una función objetivo donde se optimizaron los pesos con el fin de minimizar el error de estimación.

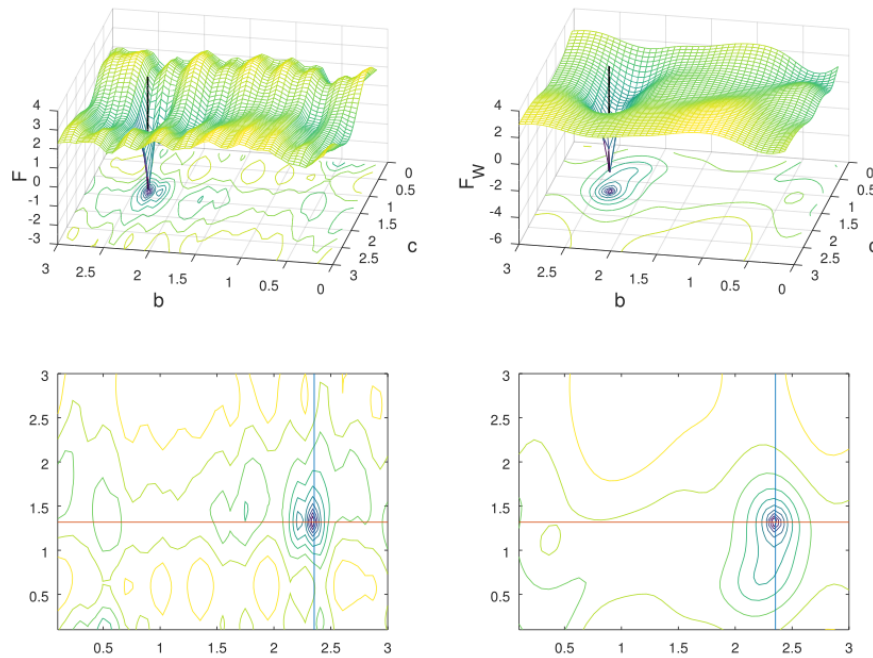


Figura 2.2: Función objetivo sin pesos y con pesos optimizados.

Se observa que la forma de la función objetivo en el segundo caso presenta características más deseables para el proceso de minimización de la función, con menor número de mínimos locales, una forma más convexa alrededor del mínimo y parámetros menos correlacionados. Esta última característica es de gran importancia, ya que si los parámetros presentan una gran dependencia entre ellos (alta correlación), ello implica que el optimizador requiere de un buen estimado de uno de los parámetros para estimar adecuadamente el otro, lo cual implica una constante actualización de

ambos valores (convergencia lenta).

Una función objetivo como la de los pesos optimizados, permite identificar los parámetros de una forma más independiente del valor de los otros parámetros y disminuye el riesgo de caer en óptimos locales.

2.3. Análisis de sensibilidad

Al estudiar la sensibilidad de los distintos parámetros del modelo con respecto a la respuesta que genera (presión en el pozo), es posible crear métodos de optimización *ad-hoc* que tomen en cuenta el efecto de cada parámetro, y así estimar los parámetros de manera más precisa y en menor tiempo de cómputo.

El modelo fractal de triple porosidad para yacimiento infinito devuelve la presión $\bar{p}_{wD}(\mathbf{x}, \tau_D)$, correspondiente a la solución sin dimensiones en el espacio de Laplace (τ_D), y por medio de inversión numérica (algoritmo de Stehfest [Stehfest \(1970\)](#)) se obtiene la presión $p_{wD}(\mathbf{x}, t_D)$, correspondiente a la solución sin dimensiones, en tiempo real sin dimensiones (t_D).

En los experimentos aquí presentados se estudia el efecto de los parámetros:

$$\mathbf{x} = (\omega_v, \omega_f, \lambda_{mf}, \lambda_{mv}, \lambda_{vf}, s, C_D, dL, \Delta) \in \mathfrak{R}^9$$

con respecto a la presión en tiempo real $p_{wD}(\mathbf{x}, t_D)$, esto mediante la generación de un gran número de casos sintéticos donde se asigna aleatoriamente los valores de los distintos parámetros, para tiempos también distribuidos aleatoriamente, considerando los siguientes rangos para cada parámetro:

$$\begin{aligned} 0 &\leq \omega_i \leq 1, & i = f, v \\ 1e-9 &\leq \lambda_i \leq 1e-3, & i = mf, mv, vf \\ -4 &\leq s \leq 50 \\ 0 &\leq C_D \leq 1500 \\ 1e-2 &\leq t_D \leq 1e+8 \\ 0 &\leq dL \leq 3 \\ 0 &\leq \Delta \leq 2 \end{aligned}$$

siempre cumpliendo la restricción $\omega_f + \omega_v \leq 1$. Los valores aleatorios de los λ 's son generados para su exponente, es decir que $\log_{10}(\lambda)$ sigue una distribución uniforme en los experimentos.

2.3.1. Sensibilidad: parámetro-tiempo

Para estudiar el efecto de cada parámetro a lo largo del tiempo, fue necesario generar 500,000 juegos de parámetros aleatorios para cada uno de los 100 diferentes valores de tiempo utilizados. Posteriormente se midió la correlación de los valores de presión obtenidos, con respecto a cada uno de los parámetro (en los distintos valores de tiempo) mediante una correlación no lineal (correlación de *Spearman*).

De esta forma el efecto que tienen los parámetros sobre la presión puede apreciarse más claramente, ya que entre más cercano es el valor de la correlación (en valor absoluto) a 1, quiere decir que ese parámetro es el que domina en ese valor de tiempo. El signo de la correlación es indicativo de que si en general la presión tiende a subir cuando el valor del parámetro aumenta (correlación positiva), o tiende a disminuir al aumentar el valor del parámetro (correlación negativa).

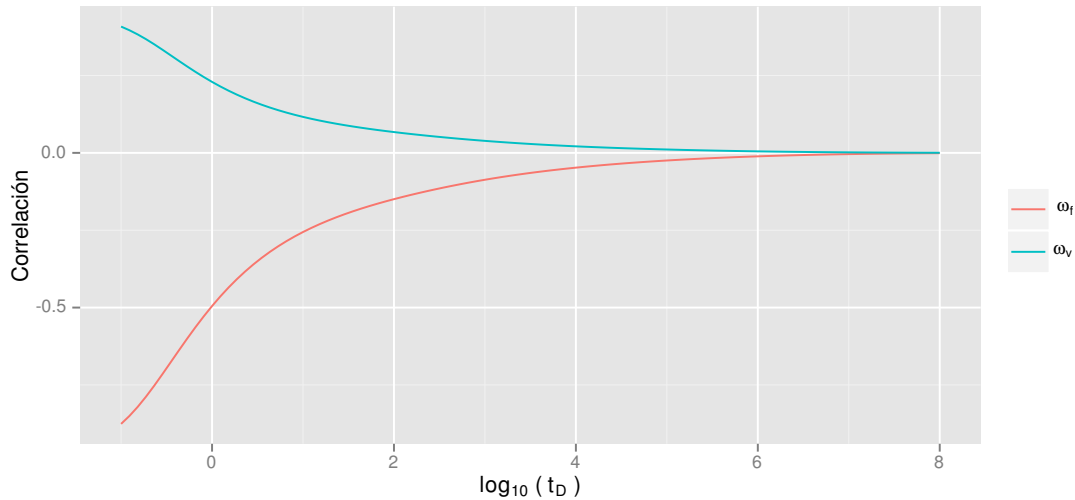


Figura 2.3: Correlación de ω_f y ω_v con la presión, en diferentes tiempos.

En la Figura 2.3 se muestra el efecto de los parámetros ω_f y ω_v , donde se puede

ver que el efecto (en magnitud) es mucho mayor para ω_f , lo cual es de esperar ya que el modelo asume un $\kappa = 1$, es decir, asume que el flujo es principalmente en fracturas.

Sin embargo, más importante aun son los tiempos en los que estos dos parámetros tienen mayor efecto sobre la curva de presión, siendo los tiempos cortos los más relevantes, por lo que si, durante el proceso de optimización, a la función objetivo se le da mayor peso a los tiempos cortos, esto nos permitirá estimar con mayor precisión los ω 's.

También puede observarse en la gráfica que ω_f tiene una correlación negativa, por lo que se espera que a mayor valor de ω_f , menor valor de la presión. El caso contrario ocurre para ω_v .

Por último, ya que el efecto de ω_f sobre la presión es mayor que el de ω_v , será más fácil estimar con precisión ω_f que ω_v , debido a que pequeños cambios sobre ω_f tendrán mayor impacto sobre la presión (y por lo tanto sobre la función objetivo), lo cual hace a estos cambios más fácilmente detectables que los que produce ω_v .

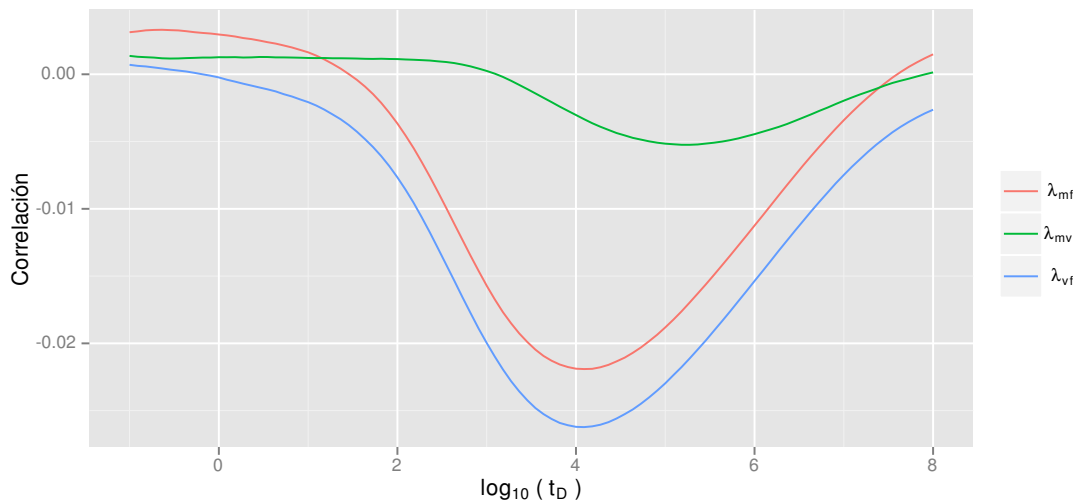


Figura 2.4: Correlación de λ_{mf} , λ_{mv} y λ_{vf} con la presión, en diferentes tiempos.

En la Figura 2.4 se presenta el efecto de los parámetros λ 's, donde se observa que λ_{mf} y λ_{vf} tienen un efecto muy similar entre sí, mientras que el efecto de λ_{mv} es

mucho menor, por lo que este último se espera tenga, considerablemente, un mayor error de estimación.

Además, el efecto de los λ 's se presenta principalmente en tiempos intermedios, por lo que son estos tiempos los que deben tener un mayor peso en la función objetivo durante la optimización, para así determinar estos parámetros con mayor precisión, ya que es en esta región donde se concentran los cambios producidos por estos parámetros.

En cuanto al efecto en la presión, la mayor parte del tiempo los parámetros λ 's tienen una correlación negativa, lo que implica que entre mayor es su valor, menor es el valor de presión. Sin embargo, debe notarse que su correlación con la presión es muy pequeña, por lo que afectarán muy poco a la misma.

En cuanto a los parámetros fractales, dL y Δ , en la Figura 2.5 se muestra su efecto en los diferentes valores de tiempo. Puede observarse que el parámetro de mayor influencia es dL , el cual, de todos los parámetros, es el que más está correlacionado con cambios en la presión a tiempos medios y largos (llegando a un valor de correlación de hasta $\rho \approx -0.94$).

Ya que ambos parámetros tienen una influencia significativa sobre la presión a lo largo de todo el intervalo de tiempo, su determinación no tendrá dificultad (en especial si se utilizan tiempos largos), por lo que se puede esperar una buena precisión en su estimación.

En los análisis anteriores no se considera el efecto de los parámetros de pozo (*i.e.* $C_D = s = 0$), esto debido al fuerte efecto que tienen sobre los parámetros de yacimiento, especialmente el almacenamiento (C_D).

En la Figura 2.6 se estudian los parámetros de pozo (s, C_D), el cual es un ejemplo de cómo un parámetro que tiene un efecto de gran magnitud, puede suprimir el efecto de otro(s) parámetro(s). En tiempos cortos, el almacenamiento (C_D) tiene una correlación de 1, lo que implica que el valor de la presión puede ser determinado enteramente por este parámetro.

El parámetro de daño (s) desplaza la presión a lo largo de todo el rango de tiempo, sin embargo, al tomar en cuenta el almacenamiento, este influye fuertemente en los valores de la presión, dando la impresión de que el efecto del daño solo aparece en tiempos intermedios y largos. Sin embargo lo que en realidad sucede es que el efecto del almacenamiento es mucho más fuerte en magnitud que el del daño, por lo cual “obscurece” el efecto del daño.

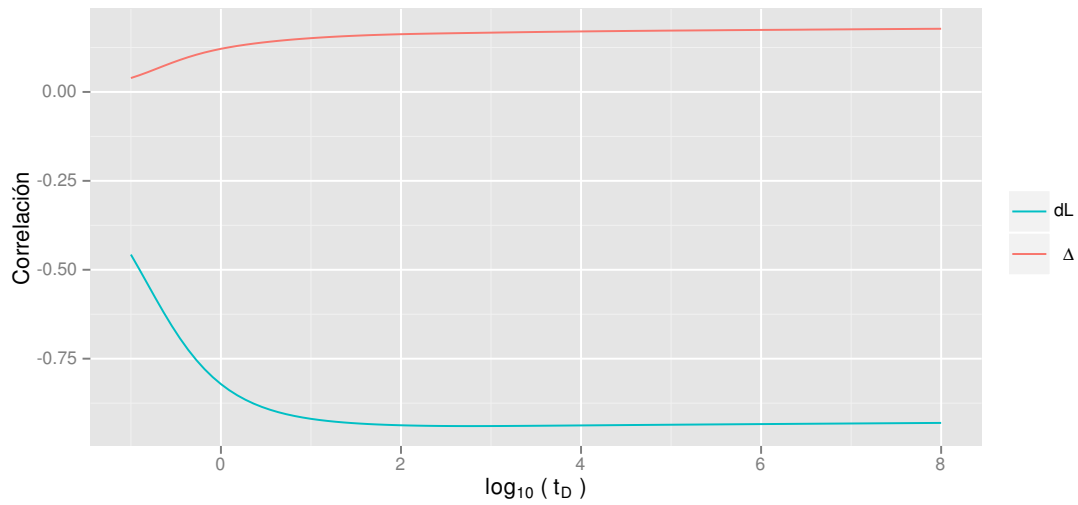


Figura 2.5: Correlación de dL y Δ con la presión, en diferentes tiempos.

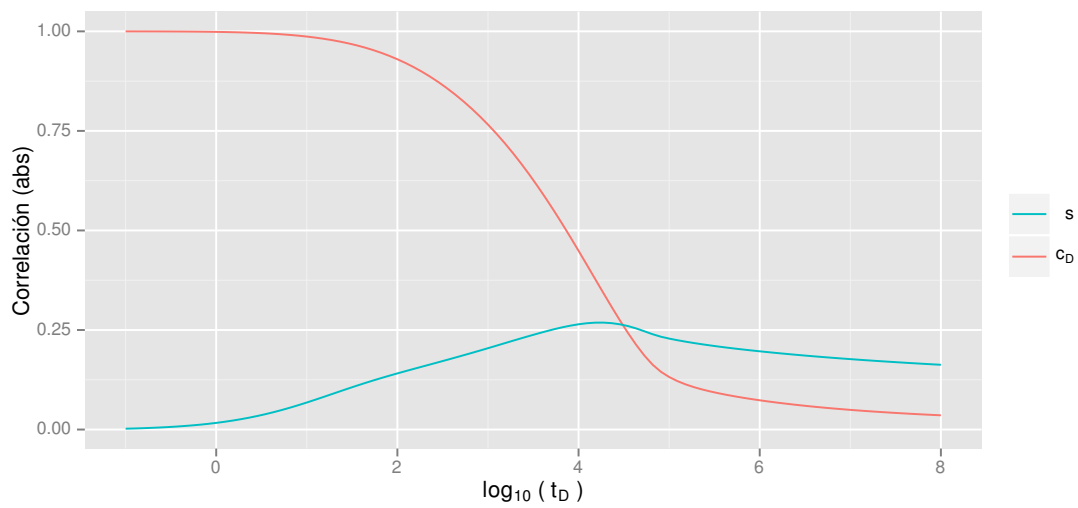


Figura 2.6: Correlación de los parámetros de pozo con la presión, en diferentes tiempos.

El estudio del efecto que posee el almacenamiento en los valores de presión es de gran importancia para la correcta determinación de los parámetros en el problema inverso, ya que este puede suprimir el efecto de los parámetros de yacimiento; por lo que es necesario estudiar los parámetros de yacimiento y de pozo por separado, ya que si se tomaran en conjunto, daría la impresión de que ningún parámetro tiene efecto a tiempos cortos (a excepción del parámetro de almacenamiento).

En la Figura 2.7 se presenta en la primera gráfica la correlación en valor absoluto de todos los parámetros de yacimiento, pudiendo observar cómo para tiempos cortos, ω_f y dL son los parámetros que más afectan la presión, y en tiempos largos son dL y Δ . Los λ 's son los parámetros que menos influyen en la presión.

Sin embargo, al considerar el efecto de almacenamiento (segunda gráfica), todos los parámetros tienden a una correlación de cero a tiempos cortos, y esto es debido a que el almacenamiento obscurece el efecto de todos los demás parámetros. Es por ello que en el proceso de optimización debe tomarse en cuenta este problema, considerando que los parámetros que tienen mayor efecto a tiempos cortos, tal vez no muestren su efecto sino hasta tiempos medios, debido al almacenamiento, por lo que los tiempos medios siempre formarán parte de la estimación de todos los parámetros de yacimiento.

2.3.2. Sensibilidad: parámetro-valor

Anteriormente se estudió la influencia de cada parámetro sobre la presión, y se determinó cómo cada uno influye de forma distinta en diferentes puntos del tiempo. Ahora se estudia la magnitud de estos cambios en la presión, ya que un parámetro puede ser el que más cambios en la presión genere en un punto determinado del tiempo, pero no necesariamente ese cambio en la presión es grande, y si los cambios que genera son pequeños, ello conlleva a un problema al momento de estimar el valor del parámetro.

El experimento nuevamente consiste en generar valores aleatorios de los parámetros y de los tiempos (los tiempos con espaciado logarítmico), y obtener el valor de presión correspondiente. Posteriormente, para valores específicos de cada parámetro se mide el valor medio de la presión, por lo que para cada valor del parámetro se tiene su efecto medio en la presión cuando el resto de parámetros (y tiempos) varían aleatoriamente en su rango. Esto nos permite conocer la influencia promedio en la

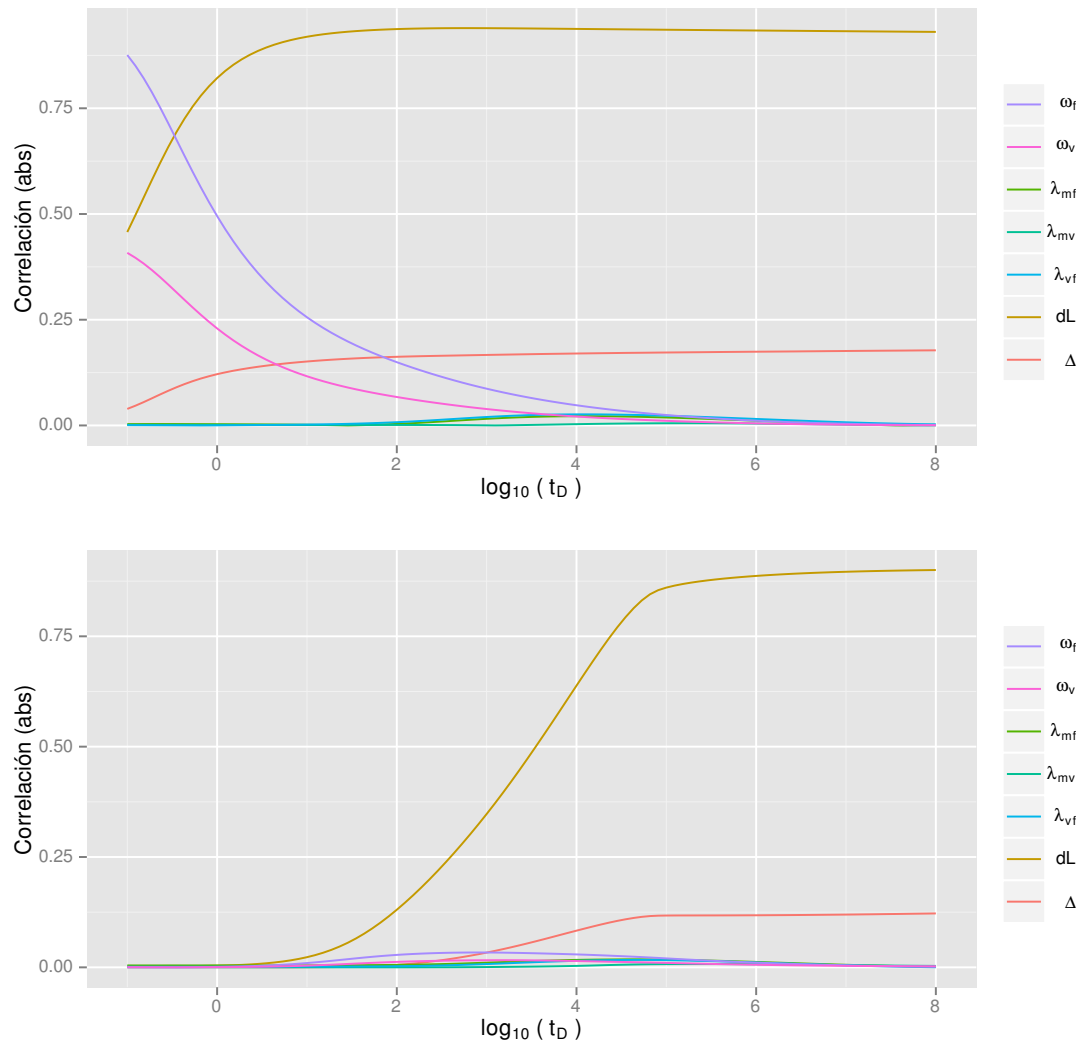


Figura 2.7: Correlación de los parámetros de yacimiento con la presión, en diferentes tiempos, cuando no existe almacenamiento y cuando sí existe.

presión que tiene cada parámetro a lo largo de su rango.

El valor a graficar es $\Delta p_{wD} = \bar{p}_{wD} - \min(\bar{p}_{wD})$, donde \bar{p}_{wD} es el valor medio de la presión para cada valor del parámetro, por lo que Δp_{wD} tiene como mínimo valor al 0 y representa el cambio medio en la presión que genera cada valor del parámetro.

Para este experimento en particular fue necesario generar para cada parámetro **36 millones de evaluaciones de la función de presión** $p_{wD}(\mathbf{x}, t_D)$, para así lograr obtener la tendencia que existe en los datos.

En la Figura 2.8 se muestra el resultado del experimento para los parámetros ω_f y ω_v . Para ω_f se observa que el efecto en la presión es mayor cuando el parámetro tiende a 0. En contraste ω_v tiene el efecto contrario, sin embargo, de los anteriores análisis (Figura 2.3) sabemos que el parámetro dominante en la presión es ω_f , por lo que este efecto “espejo” es más probable que sea debido a la restricción $\omega_f + \omega_v \leq 1$, la cual provoca que si el valor de ω_v es cercano a 1, necesariamente el de ω_f es cercano a 0.

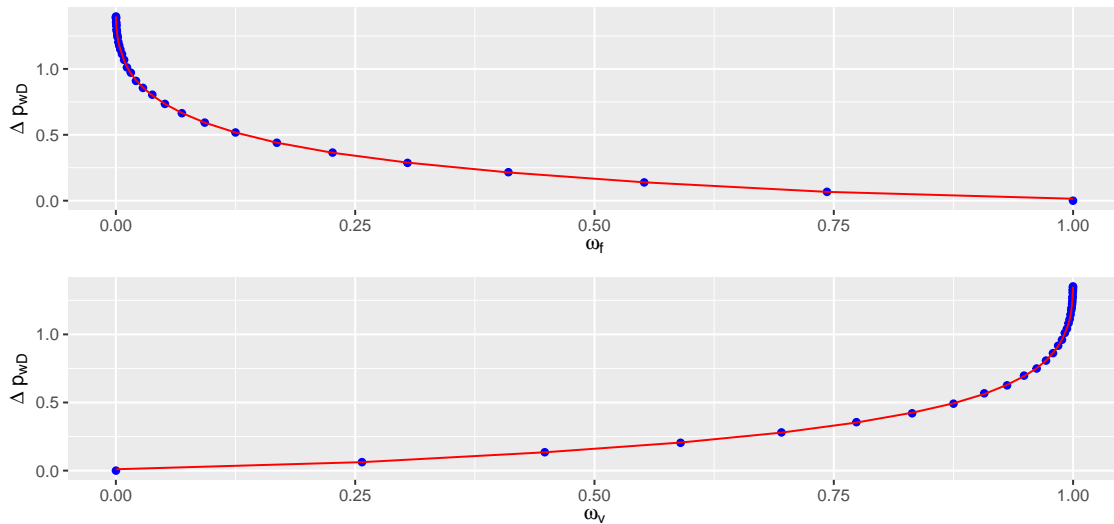


Figura 2.8: Efecto promedio de ω_f y ω_v en la presión.

Otra conclusión que se deriva de la gráfica es que puede existir un mayor error de estimación para valores grandes de ω_f que para valores pequeños, ya que para valores grandes del parámetro, el cambio en la presión es similar entre ellos, por lo

que en esta región es más difícil diferenciar que valor del parámetro causó el cambio en presión.

En la Figura 2.9 se presenta el resultado del experimento para λ_{mf} , λ_{mv} y λ_{vf} . Se observa que entre menor es el valor del parámetro, tiende a ser mayor el valor en el cambio de presión, sin embargo debemos esperar tener mayores errores de estimación para valores pequeños del parámetro, ya que, cómo en el caso anterior, es una región donde el cambio en la presión es similar entre distintos valores del parámetro, lo cual dificulta el discriminar entre los diferentes valores del mismo.

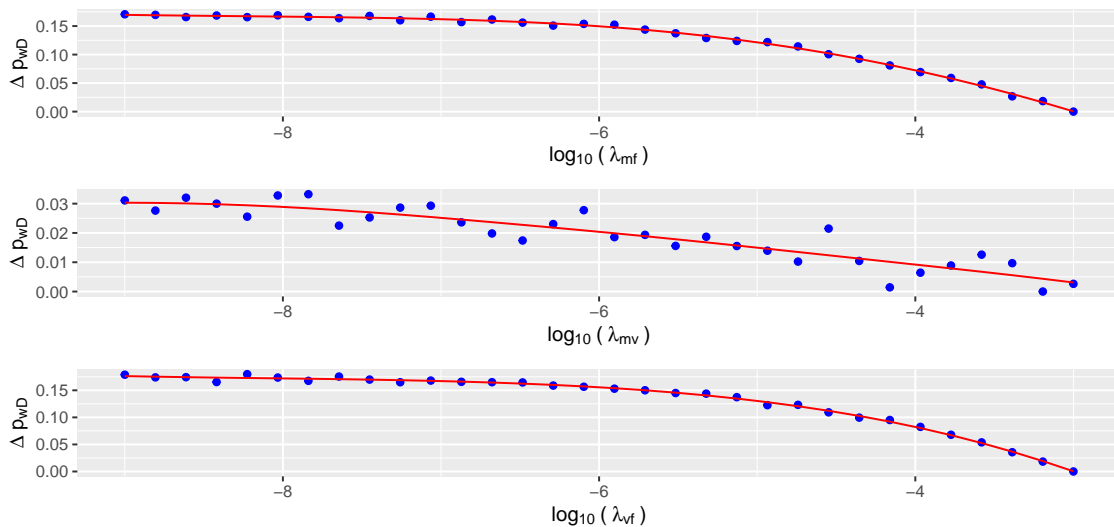


Figura 2.9: Efecto promedio de λ_{mf} , λ_{mv} y λ_{vf} en la presión.

Puede observarse que el cambio generado en la presión es mucho menor para λ_{mv} (aproximadamente la quinta parte del efecto de los otros dos λ 's). Esto se debe a que el modelo asume flujo en fracturas, por lo que la interacción entre matriz y vugulos no es significativo. Además, los análisis de correlación también mostraban que este parámetro era poco influyente en todo el rango de tiempo (Figura 2.4), y ahora se observa que el cambio que genera sobre la presión no es significativo (es por ello mismo que las mediciones de este parámetro presentan más ruido), por lo que es de esperar que la precisión en la estimación de este parámetro no sea buena.

En cuanto a los parámetros fractales dL y Δ , en la Figura 2.10 se observa que dL es por mucho el parámetro que mayor efecto tiene en los valores de presión, la

cual tiende a aumentar cuando el parámetro es más pequeño. Adicionalmente, Δ también tiene un efecto importante sobre la presión, aumentando la misma cuando el valor es mayor.

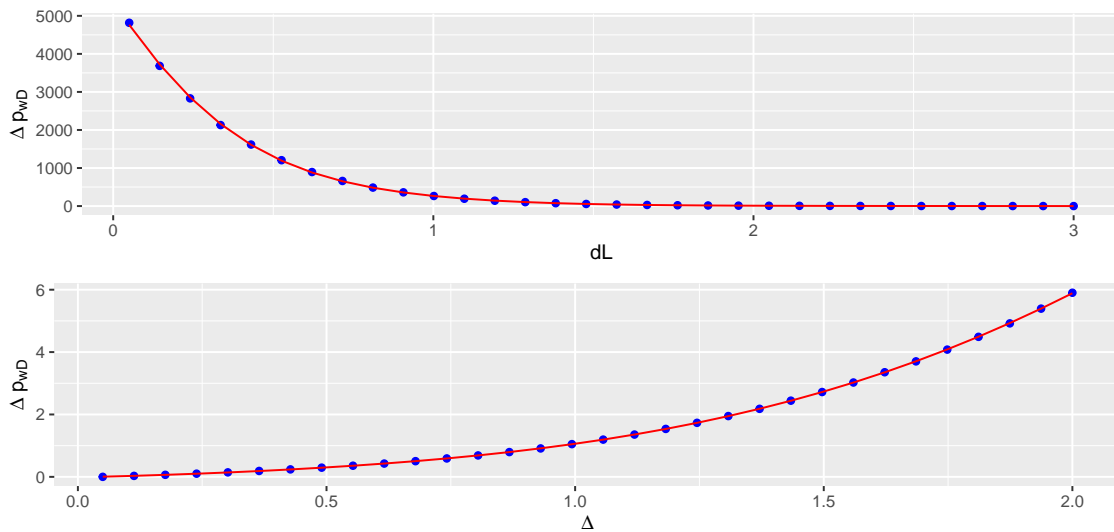


Figura 2.10: Efecto promedio de dL y Δ en la presión.

En cuanto a los parámetros de pozo (Figura 2.11), se observa en s el efecto lineal que se espera del mismo, donde por cada unidad que aumenta el parámetro, también aumenta una unidad la presión. Para C_D se presenta un cambio de tipo logarítmico, donde a mayor valor del parámetro, menor es el valor de la presión.

2.3.3. Conclusiones del análisis de sensibilidad

De los análisis de sensibilidad se generan varios resultados que son de gran importancia a considerar en el proceso de caracterización:

- Los parámetros dL , Δ , C_D y s son altamente sensibles, por lo que es posible obtenerlos con una gran precisión, en especial C_D para tiempos cortos y el resto en tiempos medios y largos.
- El parámetro ω_f es mucho más sensible que ω_v , por lo que se puede esperar mayor precisión en la estimación de ω_f , y la estimación de ambos parámetros

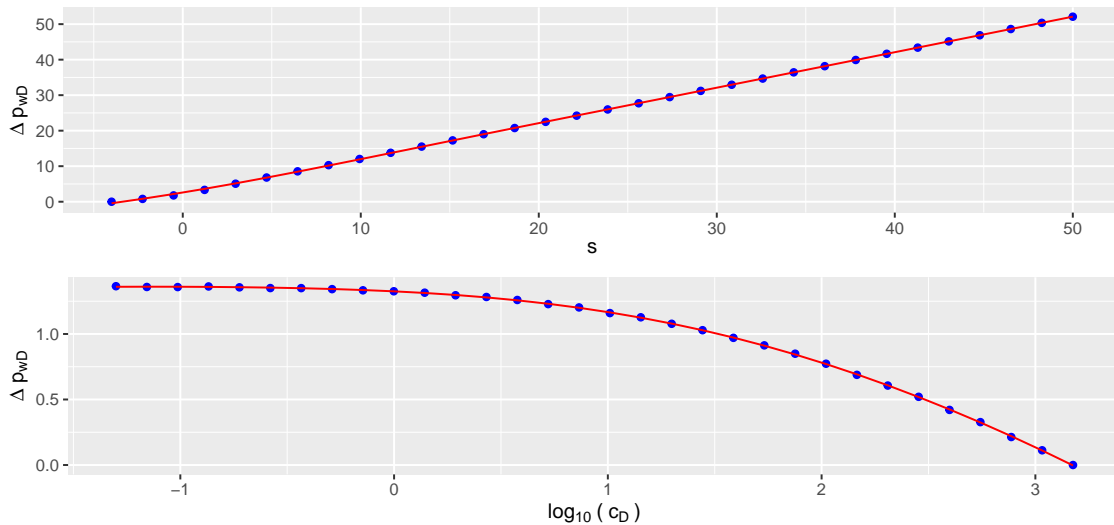


Figura 2.11: Efecto promedio de s y C_D en la presión.

debe ser más precisa si se da mayor peso a tiempos cortos y medios.

- Los parámetros λ 's son los menos sensibles de todos, por lo que se puede esperar un mayor error de estimación en estos parámetros, en especial para λ_{mv} , el cual casi no afecta a la presión. Para poder estimar estos parámetros poco sensibles es necesario dar mayor peso a los tiempos medios, ya que es en esta región donde los λ 's muestran su efecto.
- Para los λ 's podemos esperar un mayor error de estimación cuando el valor del parámetro es pequeño, y en el caso de ω 's, cuando ω_f es grande.

Adicionalmente, ya que varios de los parámetros tienen un efecto muy diferente en los cambios de presión que generan, es aconsejable e incluso necesario para un proceso de optimización eficiente determinar primero los parámetros que más influyen en la presión, o determinar en conjunto aquellos que tienen un efecto similar, ya que por ejemplo, si el algoritmo de optimización intenta identificar a la par parámetros con efectos muy dispares cómo lo es dL y λ_{mv} , cualquier pequeño cambio que se genere en dL puede influir mucho más en la presión (y por tanto en la mejora de la función objetivo) que un cambio grande en λ_{mv} , por lo que en esta situación el valor de este

parámetro poco sensible puede resultar casi aleatorio.

Es así que parámetros tan sensibles como dL , C_D o s , pueden ser determinados con gran precisión sin necesidad de tener buenos estimados para ω 's y λ 's, por lo que una estrategia adecuada sería fijar los valores de los parámetros más sensibles a lo largo del proceso de optimización (dependiendo de la sensibilidad del parámetro), para que estos no afecten la estimación de los parámetros menos sensibles.

Otro enfoque posible es una optimización secuencial, donde en cada iteración se trata de optimizar en orden y por separado los parámetros más sensibles, o en conjunto solo aquellos que tienen un efecto (en magnitud) similar entre sí, a la vez que se da mayor peso a los rangos de datos que son más sensibles al parámetro en turno, siendo el peso determinado por la propia sensibilidad estimada para los diferentes valores de la variable tiempo. Este enfoque de optimización secuencial es el que se adopta para el algoritmo *ad-hoc*.

2.4. Estimación de los parámetros

Al considerar los resultados de los análisis de sensibilidad se puede generar un algoritmo de optimización *ad-hoc* al problema. Es así que en esta sección se utiliza ese algoritmo con el fin de estudiar que tan buenos resultados se pueden obtener durante el proceso de caracterización, y a su vez obtener información de las relaciones de parámetros.

Para contrastar los distintos cambios propuestos se realiza un experimento en el cual se generan datos sintéticos de presión para **10,000 combinaciones** aleatorias para los valores de los parámetros, que representan el valor exacto de los parámetros para cada experimento, por lo que les llamaremos conjuntos de parámetros óptimos. Es para estos parámetros óptimos que tratamos de estimar su valor mediante el uso del algoritmo de optimización *ad-hoc*, y dado que se conocen los valores de los parámetros que generaron cada uno de los casos sintéticos, podemos medir los errores de estimación que se cometen en cada caso.

El rango para la generación de los valores de los parámetros y su optimización es el mismo que se utilizó en el análisis de sensibilidad, donde consideramos distribuciones uniformes tanto para la generación de los valores de los parámetros como para su optimización, a excepción de los parámetros λ 's, donde la distribución uniforme es en relación a la variable $\log_{10}(\lambda_i)$.

2.4.1. Error de estimación

El error de estimación para cada uno de los parámetros se calcula mediante:

$$err(\hat{\theta}) = \frac{|\theta - \hat{\theta}|}{R(\theta)}$$

donde $\hat{\theta}$ corresponde a la estimación del valor exacto del parámetro θ y $R(\theta)$ el valor de su rango, por ejemplo, $R(C_D) = 1500 - 0 = 1500$. Por lo que hace referencia a un error absoluto normalizado.

Ya que el error de estimación es tal que $err(\hat{\theta}) \in [0, 1]$, y por construcción $\theta/R(\theta) \sim U(0, 1)$; si la estimación del parámetro θ fuera aleatoria, es decir, $\hat{\theta} \sim U(0, 1)$, entonces el valor esperado del error sería:

$$E(err(\hat{\theta})) = E\left(\frac{|\theta - \hat{\theta}|}{R(\theta)}\right) = \frac{1}{3}$$

Por lo tanto, entre más cercano es el error medio de estimación a 0.33, mayormente se aproxima a una estimación aleatoria.

2.4.2. Resultados

Tras realizar las optimizaciones para cada uno de los 10,000 casos sintéticos, se obtuvieron los siguientes errores de estimación:

Cuadro 2.1: Media y mediana del error de estimación.

ω_f	ω_v	ω_m	λ_{mf}	λ_{mv}	λ_{vf}	s	C_D	dL	Δ
Media del error									
0.0100	0.0755	0.0758	0.1341	0.2119	0.1310	$7.13e-03$	$4.74e-06$	0.0036	0.0437
Mediana del error									
0.0011	0.0361	0.0369	0.0512	0.1321	0.0514	$5.83e-04$	$1.67e-07$	0.0013	0.0179

En el Cuadro 2.1 se presenta tanto el error medio de estimación, cómo la mediana del error. Aunque en todos los casos se estimaron los 10 parámetros del modelo, por

simplicidad de interpretación se presenta el error de estimación para cada parámetro.

El error medio de estimación nos sirve (en general) cómo una medida del error que se espera cometer para cada parámetro, y la mediana del error nos habla sobre la dispersión del error, por ejemplo, para el parámetro ω_f sabemos que el 50 % de los casos tienen un error menor o igual a 0.11 %, por lo que para ese parámetro tenemos una precisión de casi 3 decimales, en el 50 % de los casos.

Cómo se anticipó a partir de los análisis de sensibilidad, se obtuvo un mayor error de estimación para ω_v que para ω_f . A su vez, los parámetros fractales tuvieron un error de estimación pequeño, y los λ 's fueron los parámetros con mayor error, siendo λ_{mv} el que tuvo mayor error de estimación, con 13 % para la mediana del error.

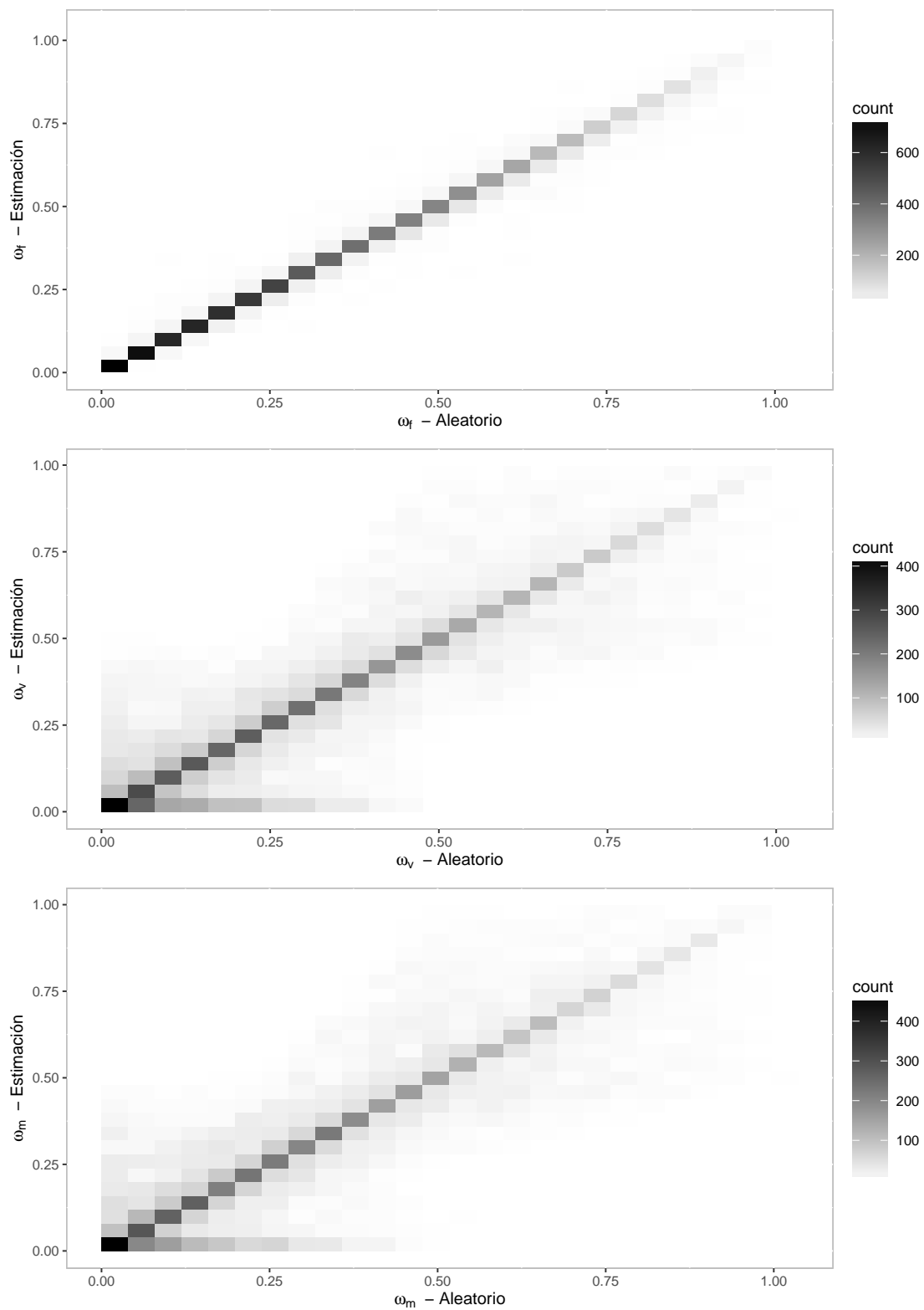
También se observa que los parámetros ω_v y ω_m , así como λ_{mf} y λ_{vf} , tienen errores de estimación muy similares. Esto es un resultado esperado debido a que existe una equivalencia entre estos parámetros, los cuales se intercambian para generar la segunda solución analítica del modelo, donde se genera exactamente los mismos valores de presión al intercambiar estos parámetros.

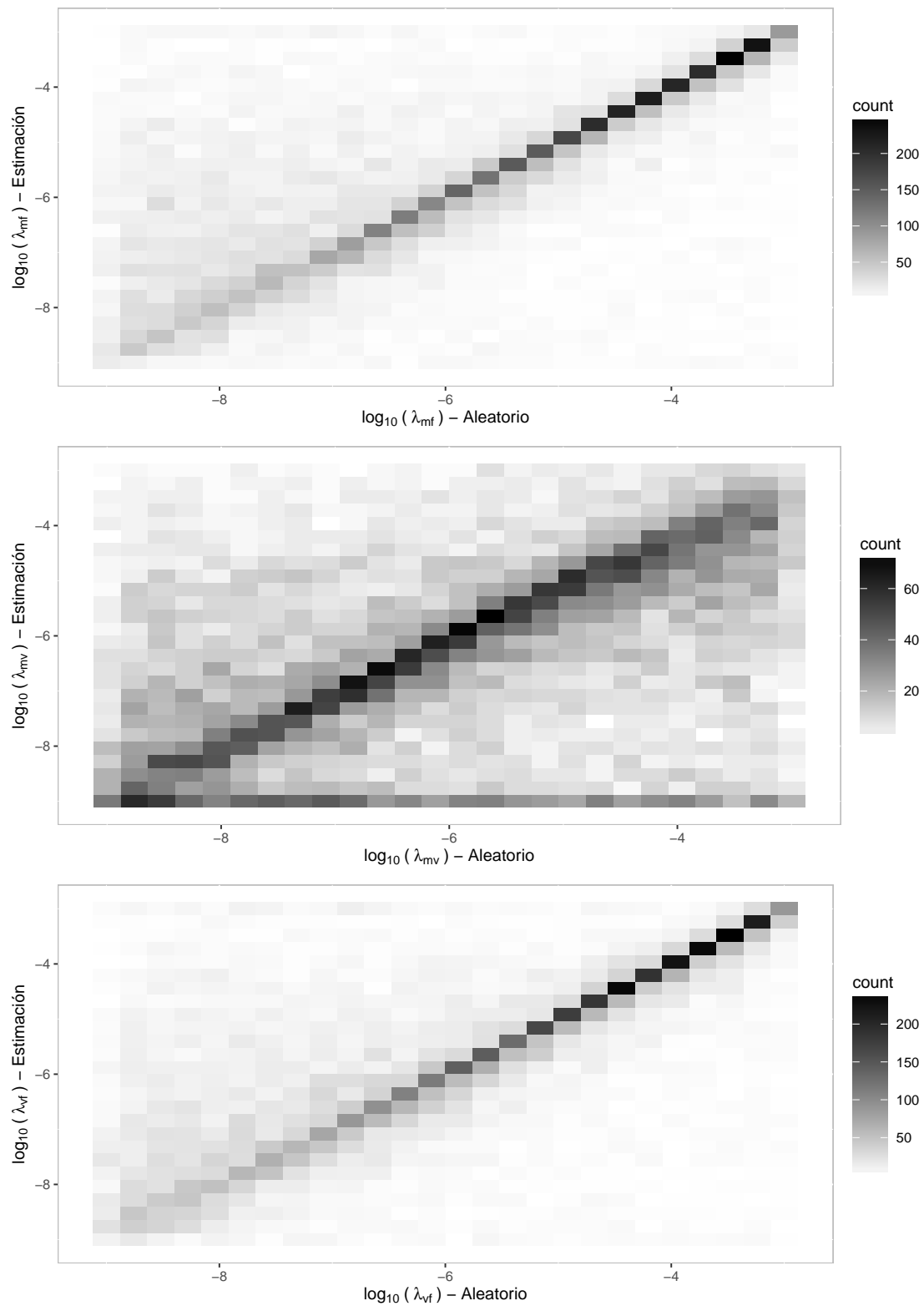
En las Figuras 2.12-2.15 se presenta la dispersión del valor a estimar *vs* el valor estimado, esto para cada uno de los parámetros del modelo. Es así que el caso ideal es una línea recta y sólida, de pendiente 1; puede verse que esta es la situación para los parámetros s , C_D y dL , que son los que tienen menor error de estimación.

Para ω_f se observa también que existe una muy buena estimación del parámetro, aunque existe un menor número de casos para valores grandes de ω_f (o de cualquier ω), lo cual es producto de la restricción $\omega_f + \omega_v \leq 1$, ya que para valores mayores a 0.5 de un ω , impide que otro ω tenga también un valor grande.

Puede verse que λ_{mv} es el parámetro que muestra mayor dispersión en su estimación (mayor error), lo cual es resultado de su poco efecto sobre la presión. Para λ_{mf} y λ_{vf} el error se presenta cuando el valor del parámetro es muy pequeño, lo cual también se dedujo sucedería a partir de los análisis de sensibilidad, donde valores pequeños de λ 's tenían un efecto muy similar entre sí, por lo cual se espera sea difícil diferenciarlos durante la optimización.

Por último, aunque Δ es un parámetro con poco error de estimación, se observa en la gráfica (Figura 2.14) que cuando hay error de estimación, este se presenta principalmente en los valores extremos que el parámetro puede tomar.

Figura 2.12: Dispersión de los valores a estimar de ω 's vs estimación.

Figura 2.13: Dispersión de los valores a estimar de λ 's vs estimación.

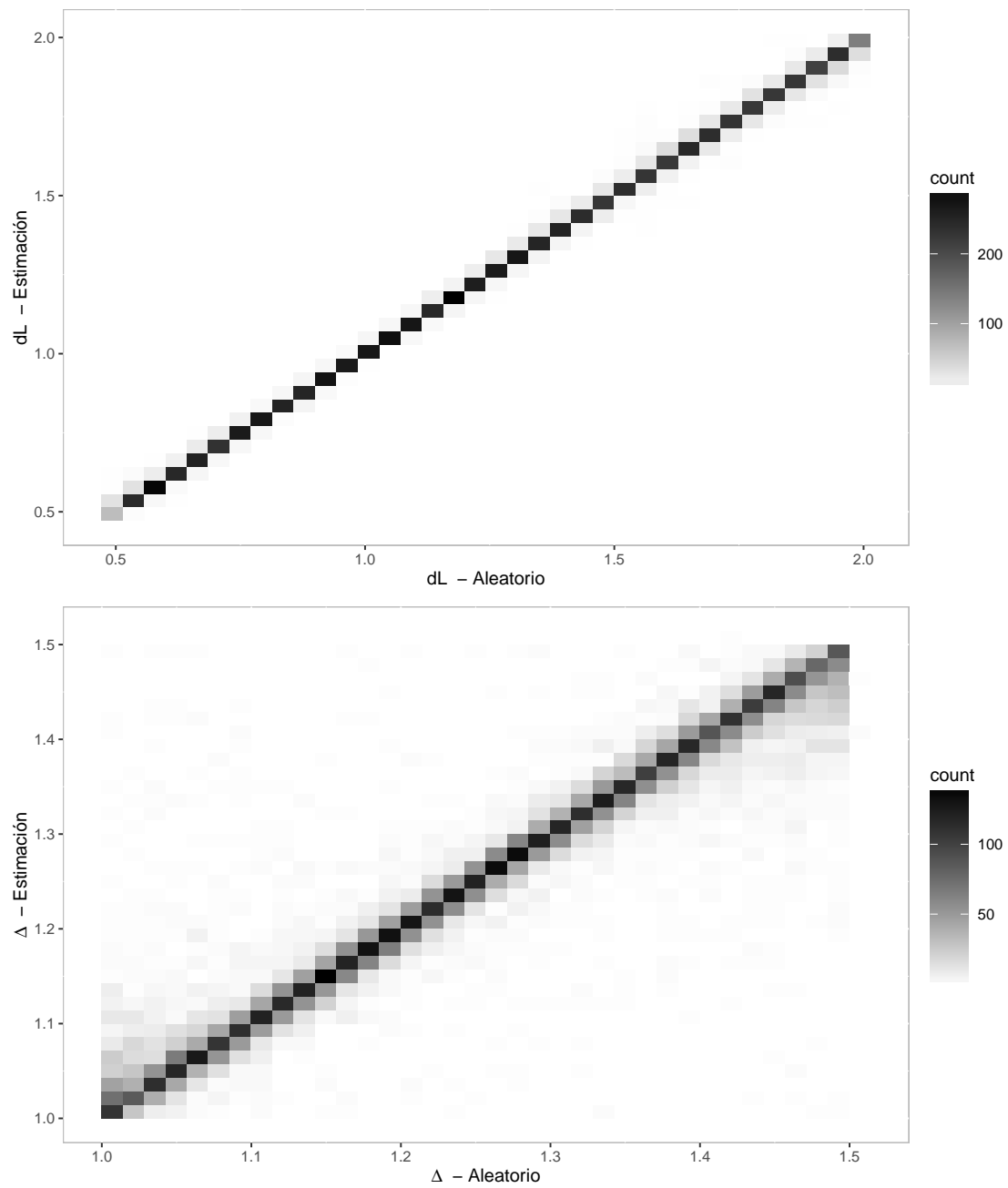


Figura 2.14: Dispersión de los valores a estimar de dL y Δ vs estimación.

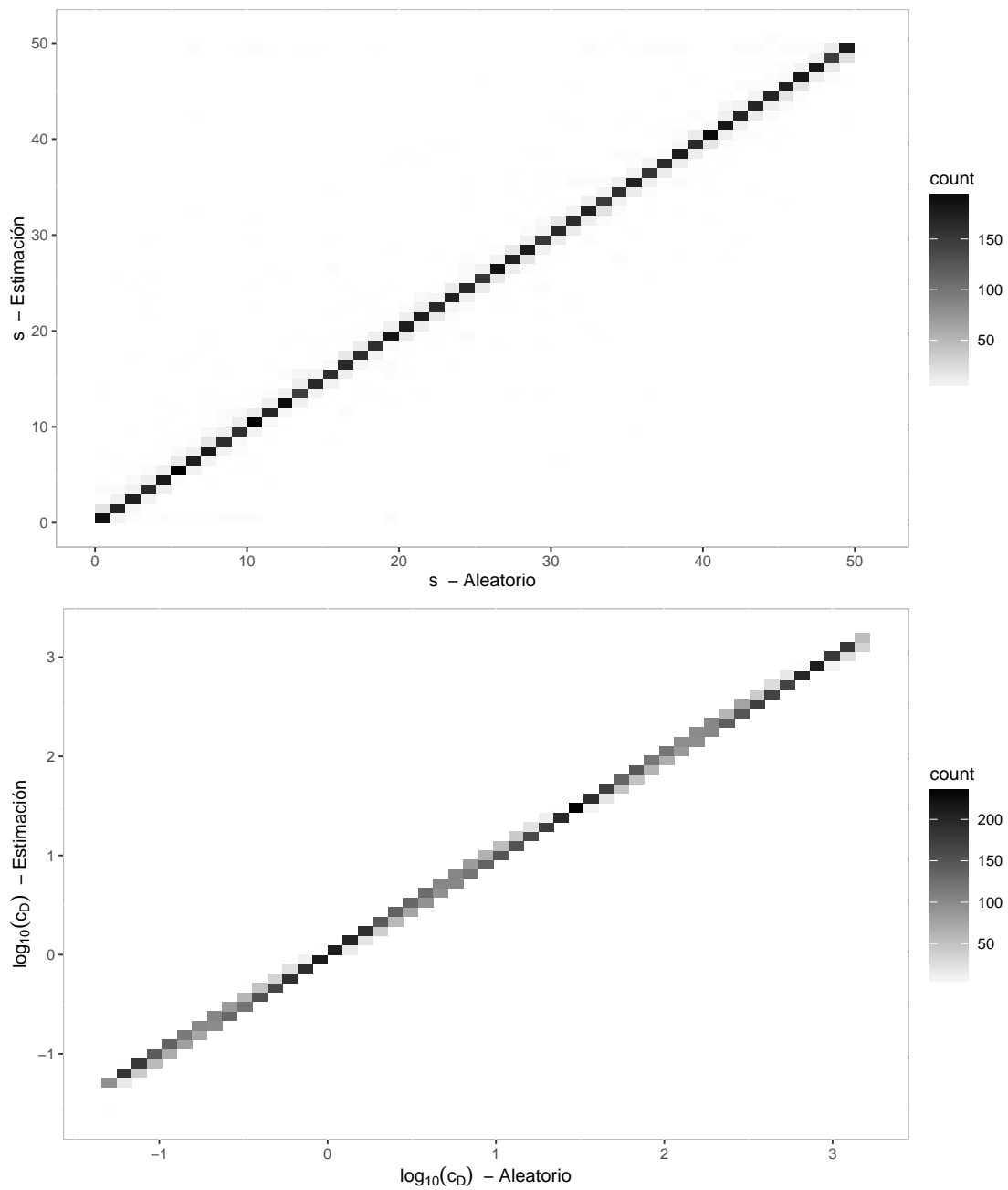


Figura 2.15: Dispersión de los valores a estimar de s y C_D vs estimación.

2.4.3. Dependencia entre parámetros y el error de estimación

En los análisis previos se ha estudiado principalmente el efecto de cada uno de los parámetros y en este apartado se estudia su efecto conjunto en el error de estimación. Esto se hace al analizar la correlación (correlación parcial) que tiene cada uno de los parámetros con los distintos errores de estimación obtenidos.

En la Figura 2.16 se muestra gráficamente las correlaciones obtenidas entre los distintos errores de estimación. Un valor cercano a -1 o 1 indica una fuerte correlación (negativa o positiva, respectivamente) mientras que un valor cercano a 0 indica falta de correlación.

En un principio se observa que la correlación entre los distintos valores de los parámetros es pequeña, lo cual es algo esperado, ya que en los casos sintéticos se generaron aleatoriamente (utilizando el rango de los parámetros). La única correlación entre los parámetros que es de relevancia, es entre los ω 's, la cual es debida a la restricción $\omega_f + \omega_v + \omega_m = 1$.

En cuanto a los errores de estimación (prefijo *E-*), existe una fuerte correlación positiva entre los errores de estimación de ω_v y ω_m , es decir que si se comete un error grande en ω_v , sucederá lo mismo con ω_m . Lo anterior es un resultado esperado, ya que ω_m se estima mediante $1 - \omega_f - \omega_v$, sin embargo ω_f influye fuertemente en la presión, por lo que este se estima con buena precisión, por lo que solo influye el valor de ω_v al determinar a ω_m .

La siguiente correlación de importancia es entre el valor de C_D y su error de estimación, siendo mayor el error de estimación entre más grande es el valor de C_D . Sin embargo, debido a que el error de estimación de C_D es muy pequeño, no es de relevancia. Lo mismo ocurre con la correlación negativa entre el valor del daño (s) y el error de C_D , o entre el error de s (*E-s*) y el error de C_D (*E-cD*).

Dentro de las correlaciones de interés se encuentran la correlación (positiva) entre *E-dL* y *E-Delta*, significando que entre mayor es el error de Δ , mayor será el error de dL , y viceversa. Considerando que del Cuadro 2.1 sabemos que el error de estimación de Δ es mayor que el de dL , significa que se requiere aun mayor precisión en la estimación de dL para poder aumentar la precisión de Δ .

Una dependencia cómo la anterior usualmente significa que ambos parámetros en algún punto producen un efecto similar sobre la función estudiada (la presión en el pozo), generando un efecto de ofuscamiento entre ambos.

También se observa una correlación (negativa) entre el valor de dL y el error en

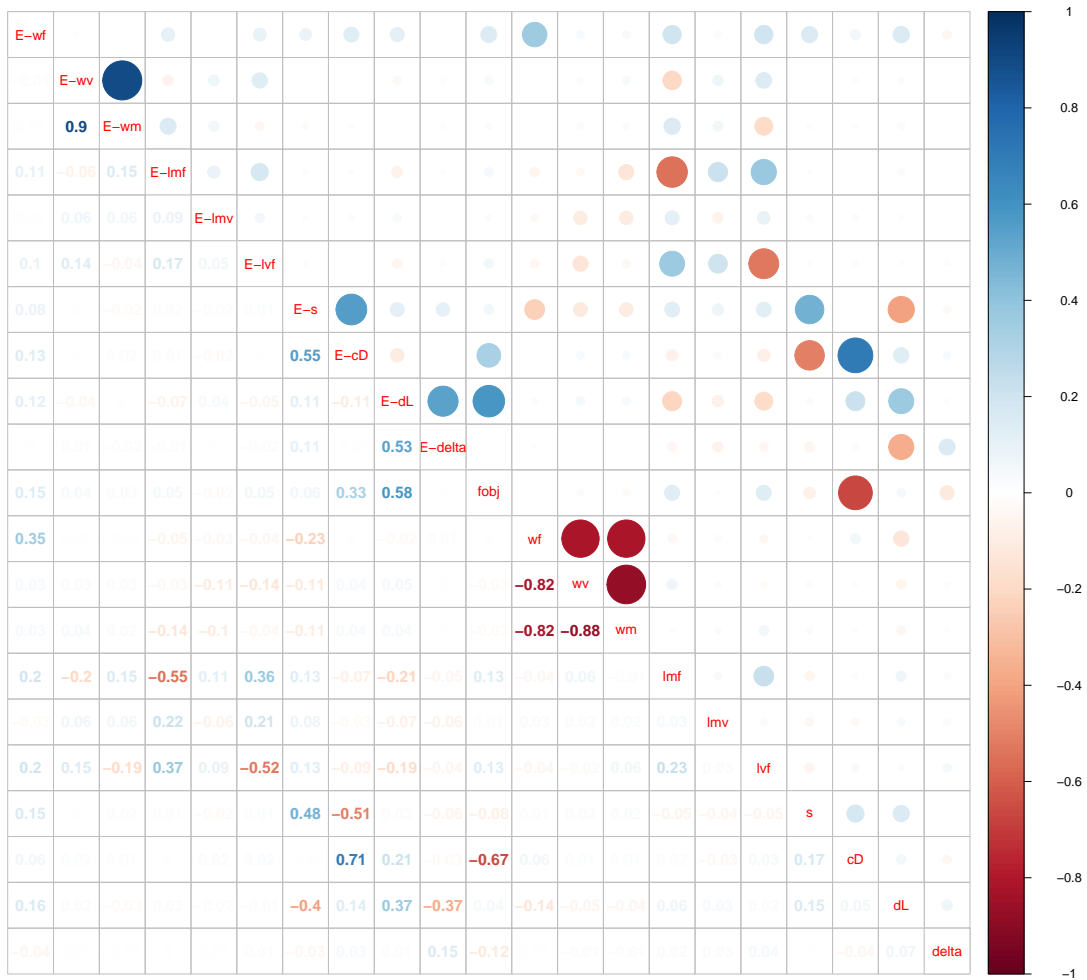


Figura 2.16: Correlación parcial (de Spearman) entre errores de estimación y valores de los parámetros.

Δ , significando que en cuanto más pequeño el valor de dL , mayor el error en Δ . A su vez se observa una correlación de magnitud similar (pero positiva) entre el valor de dL y el error de dL , significando que a valores grandes de dL les corresponde un mayor error de estimación.

En resumen, el valor de dL afecta a la estimación de Δ cuando su valor es pequeño y a la estimación de dL cuando es grande. Además, nuevamente se ve que existe una relación entre el valor de los λ 's y su error, siendo mayor su error cuando el valor del λ es menor.

2.5. Fallas y almacenamiento variable

En esta sección se estudia el efecto en la sensibilidad de los parámetros así cómo de su estimación, al agregar fallas y almacenamiento variable al modelo. Para la sensibilidad se analiza el caso de una y dos fallas, a presión constante y de flujo cero, así cómo del uso del modelo de almacenamiento variable de Fair (1981), siendo en este modelo el vector de parámetros a estimar:

$$x = (\omega_f, \omega_v, \lambda_{mf}, \lambda_{mv}, \lambda_{vf}, s, \mathbf{C}_D, dL, \Delta, L_D, C_{\phi D}, \alpha_{\phi D}) \in \mathfrak{R}^{12}$$

2.5.1. Análisis de sensibilidad

Sensibilidad: parámetro-tiempo

Al igual que los experimentos de sensibilidad anteriormente realizados, para estudiar el efecto de cada parámetro a lo largo del tiempo, fue necesario generar 500,000 juegos de valores aleatorios para los parámetros, para cada uno de los 100 diferentes valores de tiempo utilizados. Posteriormente se midió la correlación de los valores de presión obtenidos, con respecto a cada uno de los parámetros.

Cómo se mencionó anteriormente, entre más cercano es el valor de la correlación (en valor absoluto) a 1, quiere decir que ese parámetro es el que domina en ese valor de tiempo. El signo de la correlación indica si en general la presión tiende a subir cuando el valor del parámetro aumenta (correlación positiva), o tiende a disminuir al aumentar el valor del parámetro (correlación negativa).

Para estudiar la sensibilidad de las fallas y el almacenamiento variable se utiliza

el siguiente rango para los parámetros:

$$\begin{aligned} 0 &\leq L_D \leq 3,000 \\ 0 &\leq C_{\phi D} \leq 100 \\ 0 &< \alpha_{\phi D} \leq 20 \\ n_{\alpha} &= \{2, 3, 4\} \end{aligned}$$

En cuanto más cercana es la distancia de una falla, más rápidamente se ve reflejado su efecto en la curva de presión-tiempo, esto debido a que el efecto de la falla consiste en sumar o restar un valor a la presión, y a partir de que aparece el efecto de la falla en la curva de presión-tiempo, este seguirá existiendo a lo largo de la misma.

Es debido a lo anterior que, aunque los análisis de sensibilidad no toman un valor específico para la distancia de la falla (usa todo el rango de valores), se espera que su efecto sea visible principalmente a tiempos largos, lo cual se confirma en la Figura 2.17.

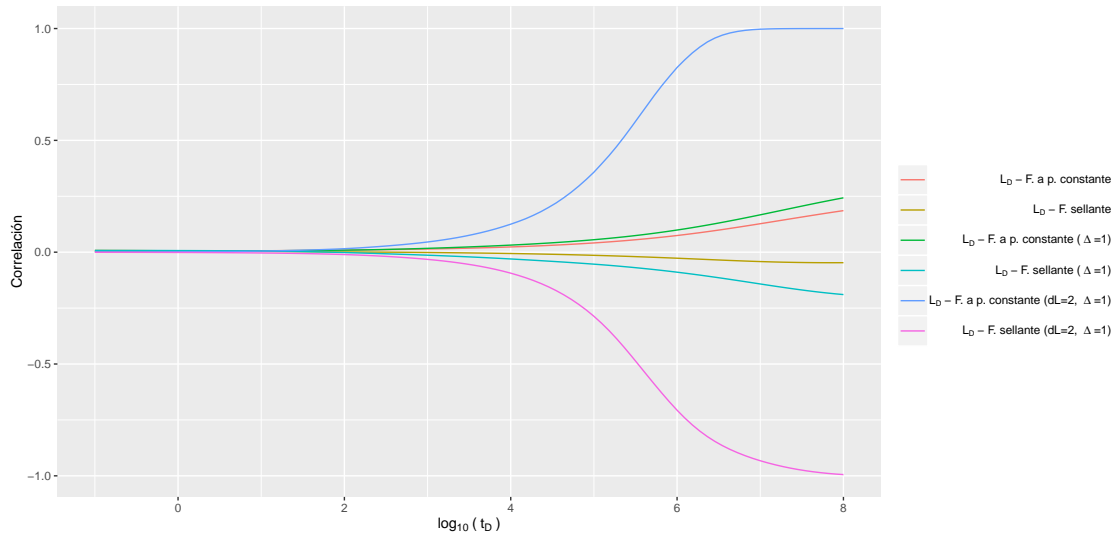


Figura 2.17: Correlación de L_D con la presión, en diferentes tiempos.

Cómo es de esperar, se observa que la falla a presión constante tiene una correlación positiva con la presión (aumenta la presión) y la falla a flujo cero (falla sellante),

una correlación negativa.

Un resultado relevante de la sensibilidad de ambas fallas es que estas no son completamente simétricas, ya que se ve afectada por los parámetros fractales (principalmente por Δ). Puede verse que, aunque el efecto de las fallas aparece mayormente a tiempos largos, la magnitud de su efecto se ve afectada dependiendo del valor de Δ :

- Cuando Δ no se encuentra fijo, la sensibilidad de la falla a presión constante es mayor que la sellante, esto sugiere que cuando la falla es sellante, Δ y L_D (distancia de la falla) tienen un efecto “similar” hasta cierto punto, haciendo que el efecto de uno se confunda parcialmente con el efecto del otro.
- Al fijar Δ , la magnitud del efecto de ambos tipos de fallas se vuelve similar, y al fijar también a dL , el efecto se vuelve completamente simétrico.

De lo anterior puede esperarse que las fallas tengan un efecto en la estimación de Δ , aumentando su error de estimación.

En la Figura 2.18 se presenta el caso de dos fallas, donde nuevamente se observa que para el caso de fallas sellantes, la magnitud del efecto es menor. El parámetro $n_\alpha = 180/\alpha$ (donde α es el ángulo entre las dos fallas) no muestra ser muy sensible.

En la Figura 2.19 se presenta la correlación de cada uno de los parámetros involucrados en el almacenamiento variable. Todos ellos tienen efecto a tiempos cortos y su correlación es suficiente como para que ninguno tenga problemas de estimación. A diferencia de las fallas, no se encontró que otros parámetros afecten en su estimación, y los mismos solo afectan a la estimación de otros parámetros en el mismo grado que ya lo hacía el almacenamiento no-variable.

Sensibilidad: parámetro-valor

Para el caso de fallas, en la Figura 2.20 se muestra el efecto medio que tiene el valor del parámetro L_D para cada una de las fallas. Se observa que cerca del valor 1,000 el efecto sobre la presión es casi cero, ya que en cuanto más lejana está la falla, más tarda su efecto en aparecer en la curva presión-tiempo, y en algún momento no aparecerá su efecto en absoluto en el periodo de tiempo manejado. Por lo anterior, es de esperar que entre más lejana se encuentre la falla, mayor será su error de estimación.

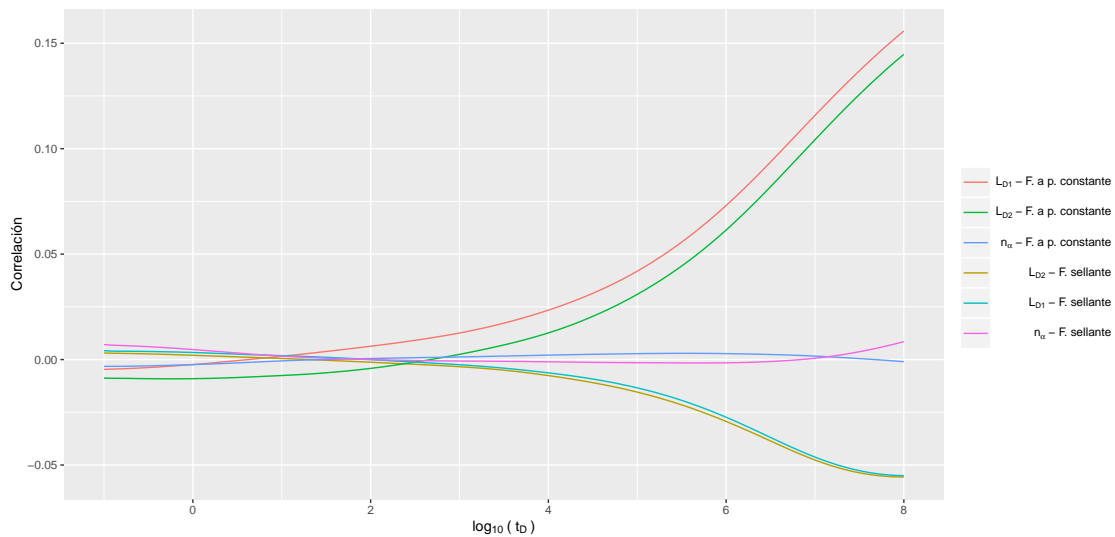


Figura 2.18: Correlación de $L1_D$, $L2_D$ y n_α con la presión, en diferentes tiempos.

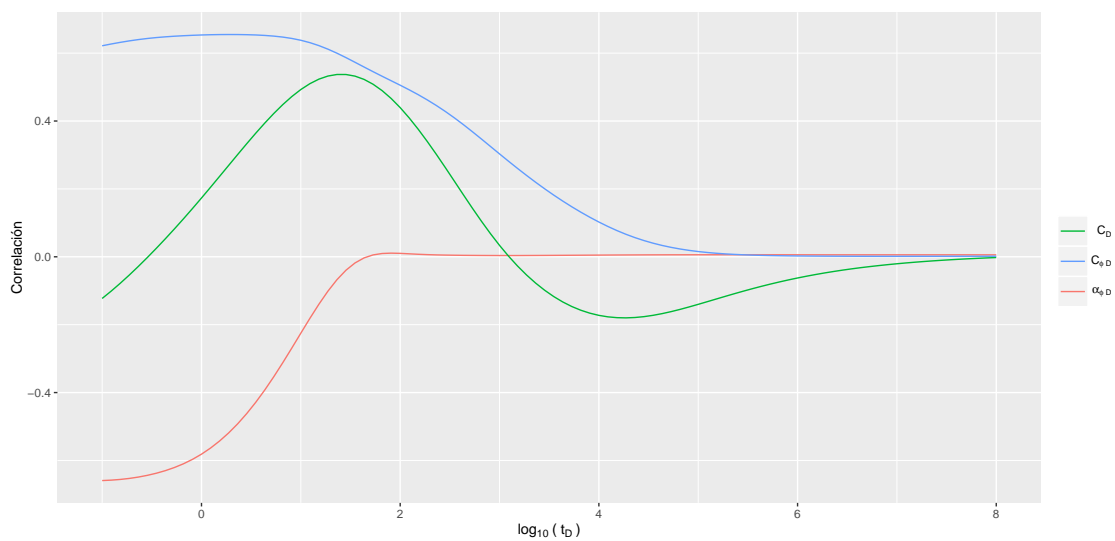


Figura 2.19: Correlación de C_D , $C_{\phi D}$ y $\alpha_{\phi D}$ con la presión, en diferentes tiempos.

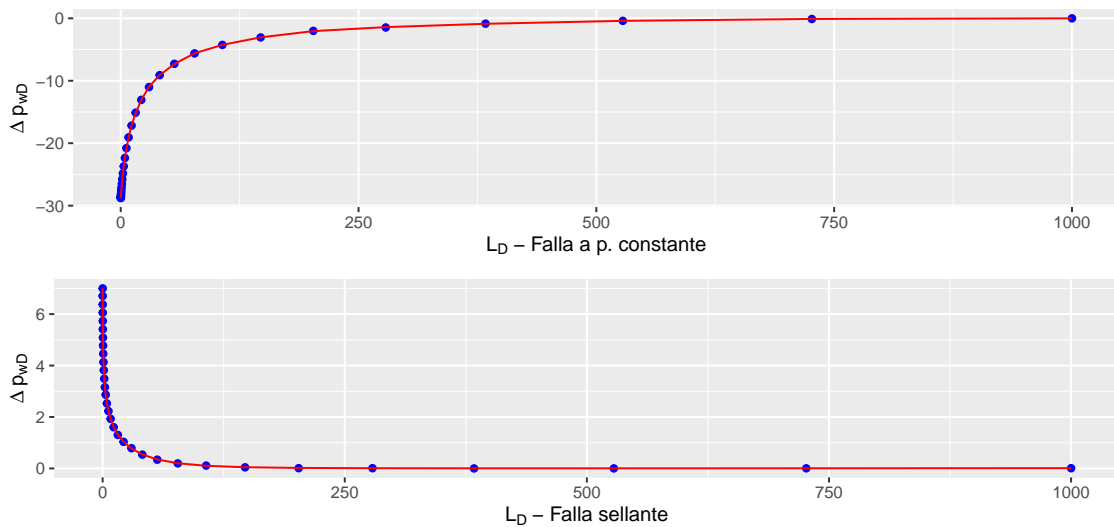


Figura 2.20: Efecto promedio de L_D en la presión, para una fallas a presión constante y fallas sellantes.

2.5.2. Estimación de los parámetros

Se realiza nuevamente el experimento de optimización para un gran número de casos sintéticos (7,000 para fallas a presión constante y 7,000 para fallas sellantes), para así medir los errores de estimación de cada parámetro, usando el método de optimización *ad-hoc*, ahora incluyendo entre los parámetros desconocidos la distancia de la falla y los parámetros de almacenamiento variable. Al aumentar el número de parámetros se espera que también aumente el error de estimación, pero también se espera observar cómo estos nuevos parámetros afectan o se ven afectados por el resto de parámetros.

Resultados

En el Cuadro 2.2 se presentan la media y mediana de los errores de estimación de los parámetros para el modelo con fallas y almacenamiento variable. Puede observarse que algunos de los errores de estimación ahora son mayores, pero, como se esperaba, Δ es el parámetro que se ve más afectado por la inclusión de las fallas, ya que como se mostró en los análisis de sensibilidad, Δ puede verse afectado por el comportamiento

de L_D en algunas partes de la curva.

Cuadro 2.2: Media y mediana del error de estimación, incluyendo fallas y almacenamiento variable.

ω_f	ω_v	ω_m	λ_{mf}	λ_{mv}	λ_{vf}	s	C_D
Media del error							
0.0552	0.0983	0.0987	0.1765	0.2419	0.1757	0.0417	0.0053
Mediana del error							
0.0125	0.0610	0.0604	0.0807	0.1678	0.0802	0.0073	0.0001
dL	Δ	L_D	$C_{\phi D}$	$\alpha_{\phi D}$			
Media del error							
0.0174	0.1580	0.1489	0.0177	0.0260			
Mediana del error							
0.0071	0.1006	0.0922	0.0008	0.0009			

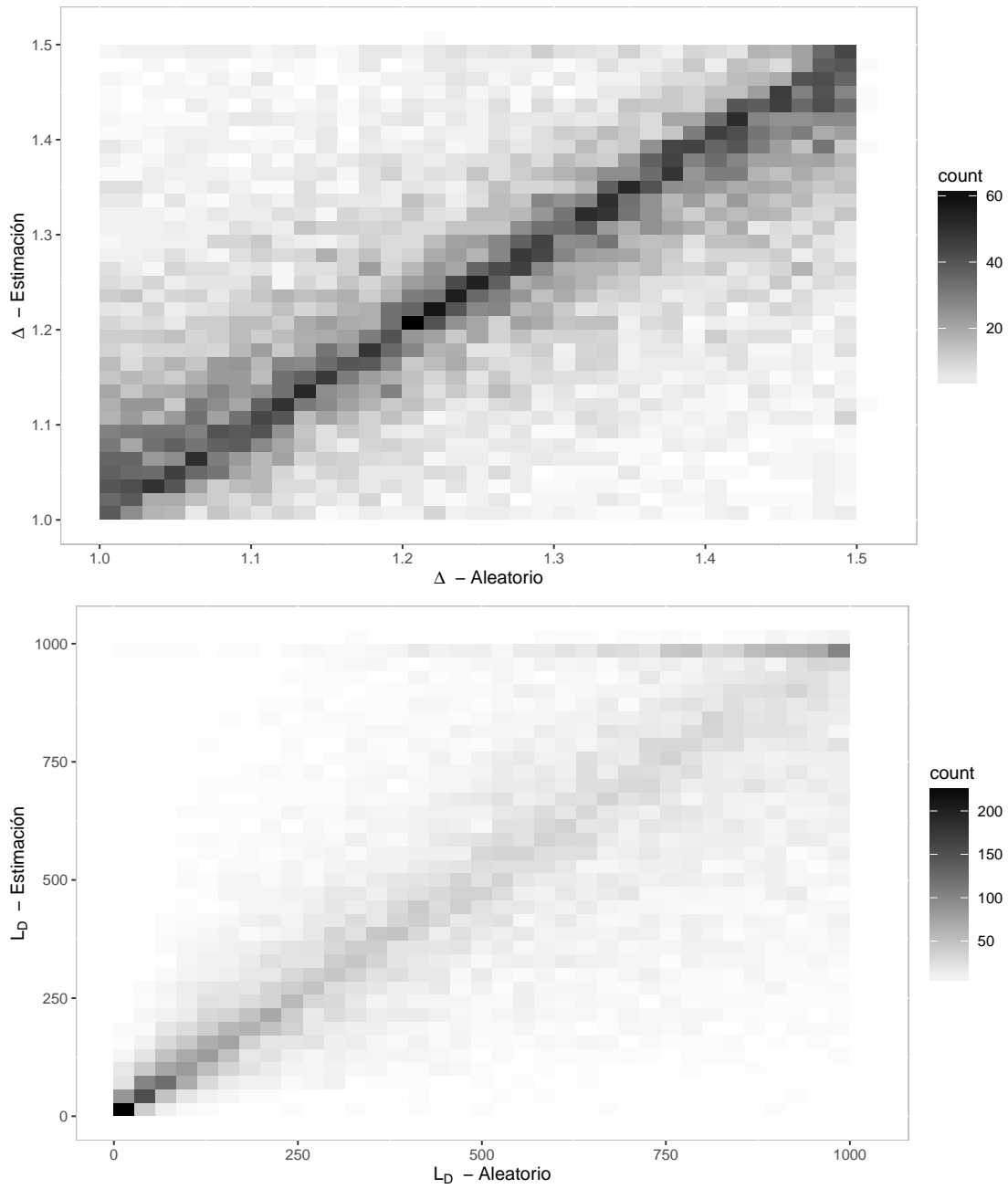
En las Figuras 2.21 y 2.22 se presenta la dispersión del error de estimación para los nuevos parámetros y para Δ . Al comparar el valor a estimar *vs* estimación, se observa que el error de Δ ahora es mayor y que el error de dL aumenta cuando el valor de L_D también aumenta.

Esto último también se dedujo a partir de los análisis de sensibilidad, ya que si las fallas se presentan a mayores distancias, tienen menor impacto sobre la presión al tardar más tiempo en mostrar su efecto. Para los parámetros de almacenamiento variable puede verse que la estimación es altamente precisa y no representan una dificultad.

Dependencia entre parámetros y el error de estimación

En la Figura 2.23 se muestran las nuevas correlaciones de los parámetros y los errores de estimación ahora que se han agregado fallas y almacenamiento variable.

Puede verse que el error de Δ (*E-Delta*) está altamente correlacionado con el

Figura 2.21: Dispersión de los valores a estimar de Δ y L_D vs estimación.

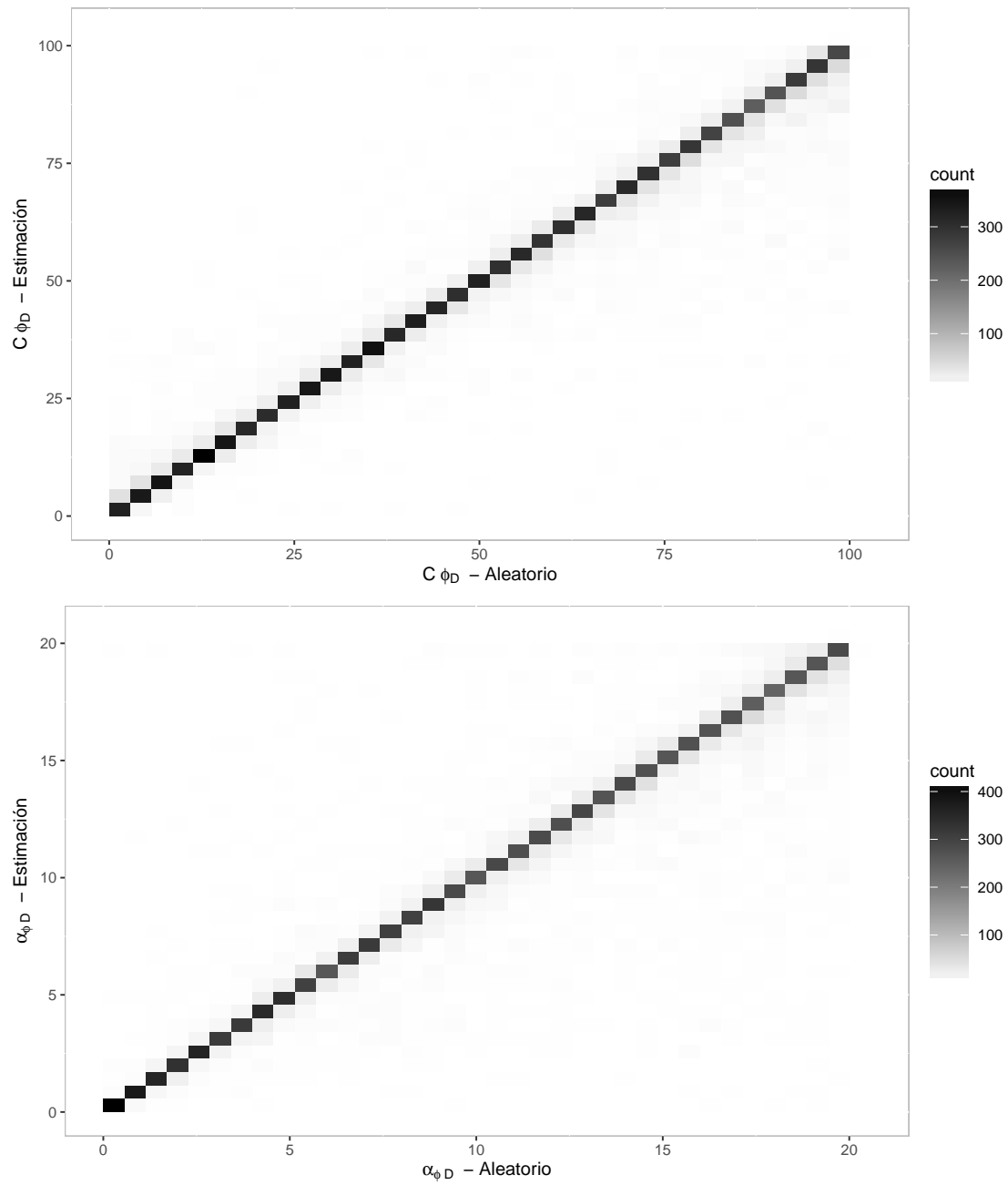


Figura 2.22: Dispersión de los valores a estimar de C_{ϕ_D} y α_{ϕ_D} vs estimación.

error en L_D ($E-LD$), por lo que es necesaria una buena precisión en Δ para lograr una buena estimación de L_D y viceversa.

Se observa que el valor de L_D se encuentra correlacionado negativamente con el error en Δ , por lo que es cuando la falla se encuentra cerca del pozo que puede aumentar el error de estimación en Δ . También se confirma la correlación entre el valor de L_D y su error de estimación, siendo mayor el error en tanto es mayor el valor de L_D .

Por último, puede verse que los errores de los parámetros $C_{\phi D}$ y $\alpha_{\phi D}$ solo se encuentran correlacionados entre ellos, y no con algún otro parámetro.

2.6. Casos reales

En esta sección se muestran los resultados del ajuste en tres casos reales, usando el método heurístico diseñado *ad-hoc*, para el problema.

En cada ajuste se presentan dos soluciones, ya que el modelo genera la misma curva de presión para el conjunto de parámetros \mathbf{x} y para \mathbf{x}' , donde en \mathbf{x}' se intercambian el valor de ω_v con ω_m , y el valor de λ_{mf} con λ_{vf} . Por lo que en un ajuste, siempre existirán dos conjuntos de parámetros que generan la misma curva de presión. Sin embargo, con información adicional de los yacimientos es posible determinar cual de las dos soluciones obtenidas es la que caracteriza mejor al yacimiento.

Adicionalmente se presentan los parámetros α_p y α_t , que corresponden al paso de dimensiones a sin dimensiones para la presión y el tiempo, respectivamente, por lo que para estos casos se determinó en total 11 parámetros.

Para el primer caso, el pozo F1, se eliminó parte de los datos al final de la prueba, debido a la posibilidad de que el efecto que se observa sea ruido causado por el tiempo de producción de la prueba, adicionalmente se presenta el caso F1B, donde se realiza un ajuste sin eliminar esa información.

2.6.1. Pozo F1

Al ajustar las curvas con el método de optimización *ad-hoc* se obtuvo el ajuste de la Figura 2.24, cuyos parámetros se describen en el Cuadro 2.3. Se eliminó parte de los datos debido a la posibilidad de que sea ruido causado por el efecto generado por el tiempo de producción.

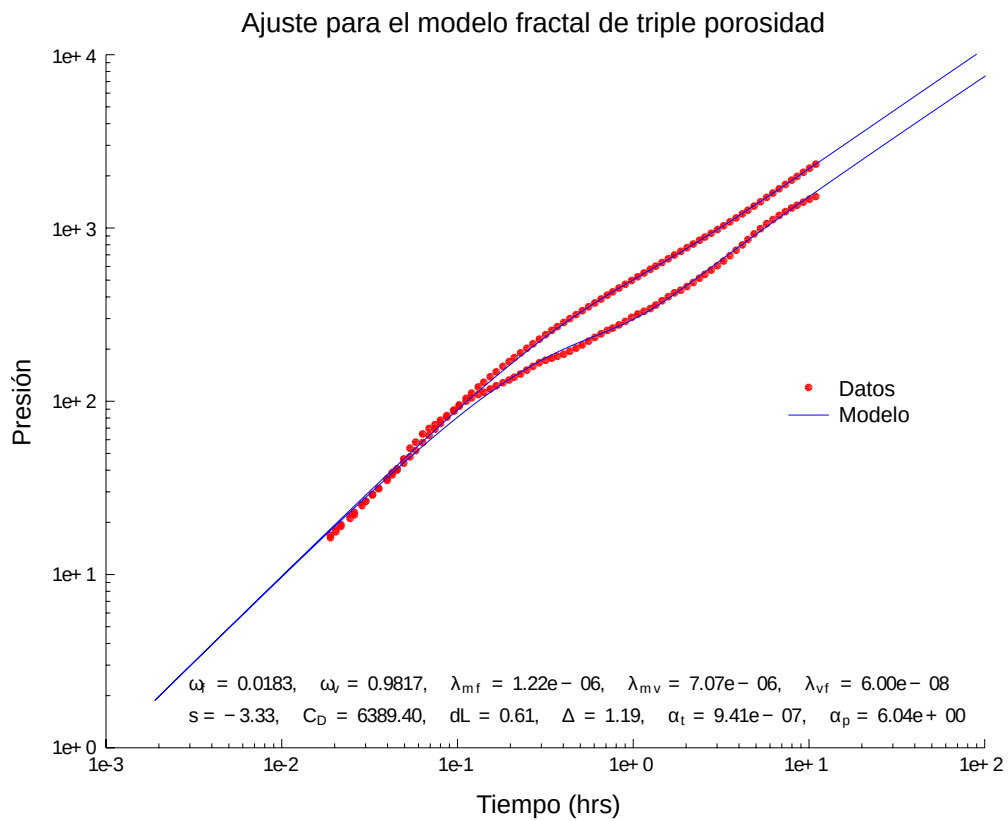


Figura 2.24: Ajuste para el pozo F1 (datos eliminados debido al tiempo de producción)
- Método *ad-hoc*.

Cuadro 2.3: Parámetros para el pozo F1 (datos eliminados debido al tiempo de producción) - Método *ad-hoc*.

	Solución 1	Solución 2
ω_f	0.018253	0.018253
ω_v	0.981750	1.4526e-10
λ_{mf}	1.2152e-06	5.9994e-08
λ_{mv}	7.0738e-06	7.0738e-06
λ_{vf}	5.9994e-08	1.2152e-06
s	-3.3331	-3.3331
C_D	6389.4	6389.4
dL	0.60571	0.60571
Δ	1.1890	1.1890
α_t	9.4099e-07	9.4099e-07
α_p	6.0392	6.0392

2.6.2. Pozo F1B

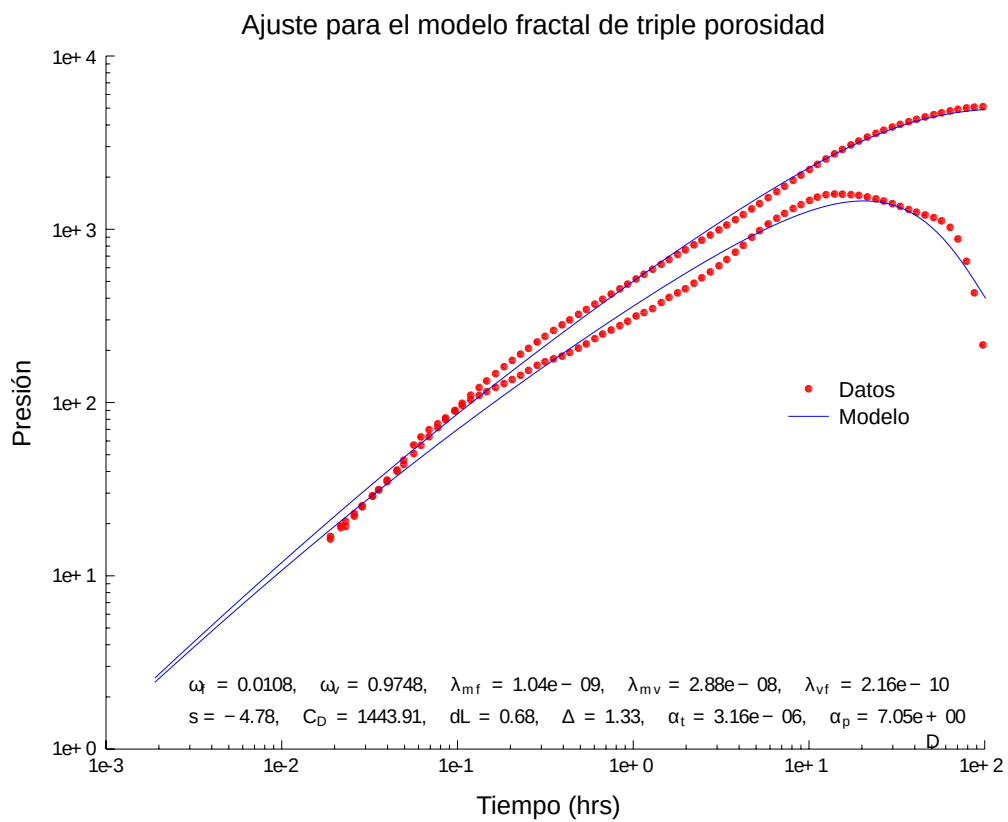
Cuando en el pozo F1 no se eliminan los datos del último periodo de tiempo, se obtiene el ajuste de la Figura 2.25, cuyos parámetros corresponden al Cuadro 2.4.

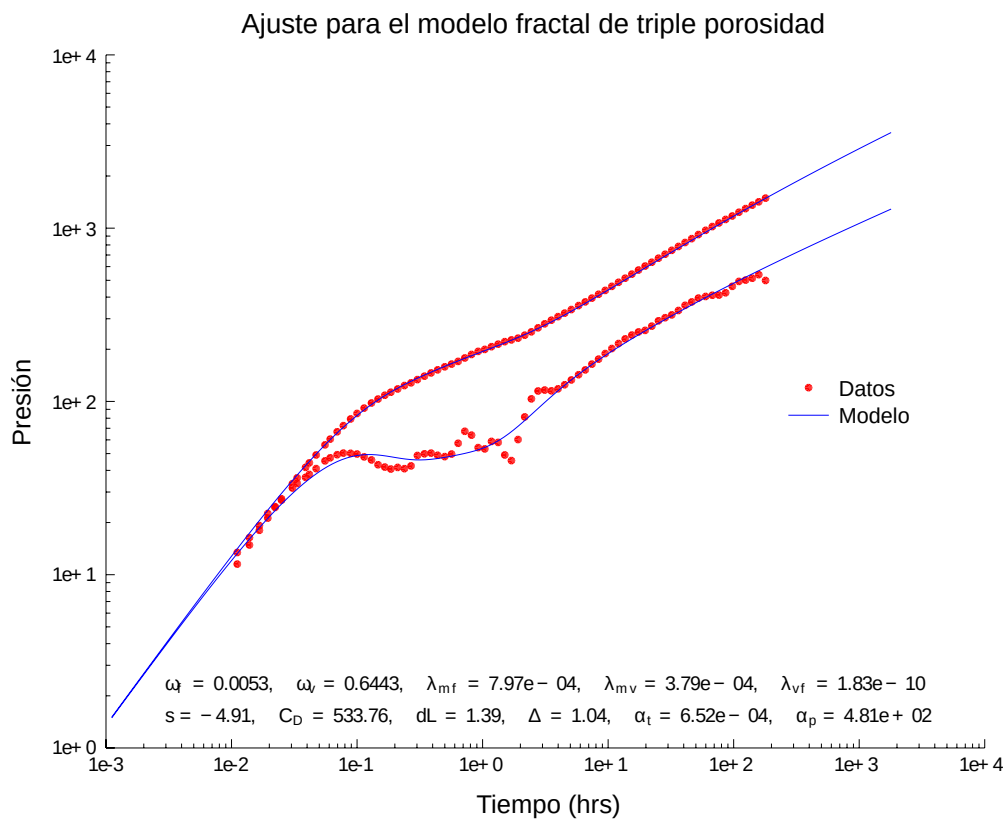
Cuadro 2.4: Parámetros para el pozo F1 (sin eliminar datos) - Método *ad-hoc*.

	Solución 1	Solución 2
ω_f	0.010797	0.010797
ω_v	0.974830	0.014372
λ_{mf}	1.0383e-09	2.1622e-10
λ_{mv}	2.8802e-08	2.8802e-08
λ_{vf}	2.1622e-10	1.0383e-09
s	-4.7766	-4.7766
C_D	1443.9	1443.9
dL	0.68439	0.68439
Δ	1.3346	1.3346
α_t	3.1580e-06	3.1580e-06
α_p	6.9885	6.9885

2.6.3. Pozo F2

Para el pozo F2 se obtuvo el ajuste de la Figura 2.26 cuyos parámetros se muestran en el Cuadro 2.5.

Figura 2.25: Ajuste para el pozo F1 (sin eliminar datos) - Método *ad-hoc*.

Figura 2.26: Ajuste para el pozo F2 - Método *ad-hoc*.

Cuadro 2.5: Parámetros para el pozo F2 - Método *ad-hoc*.

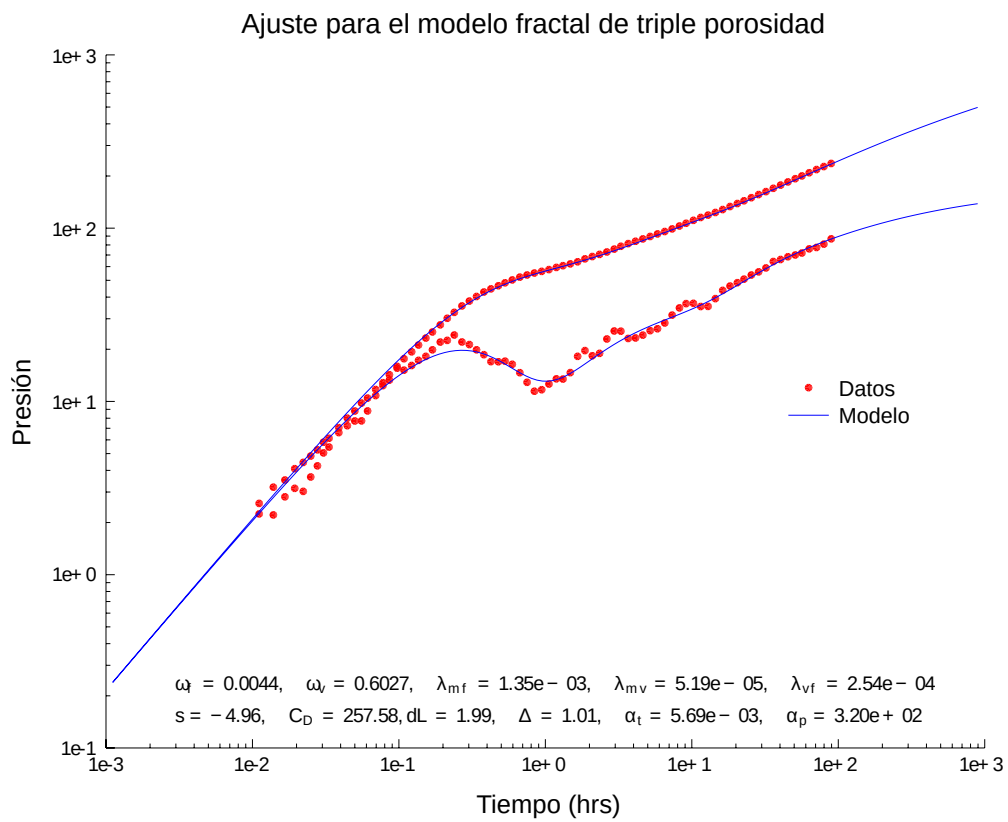
	Solución 1	Solución 2
ω_f	0.005278	0.005278
ω_v	0.644320	0.350400
λ_{mf}	7.9658e-04	1.8344e-10
λ_{mv}	3.7860e-04	3.7860e-04
λ_{vf}	1.8344e-10	7.9658e-04
s	-4.9142	-4.9142
C_D	533.76	533.769
dL	1.3909	1.3909
Δ	1.0367	1.0367
α_t	6.5203e-04	6.5203e-04
α_p	480.64	480.64

2.6.4. Pozo F3

Por último, para pozo F3 el ajuste que se obtuvo con el método *ad-hoc* se muestra en la Figura 2.27 y sus parámetros en el Cuadro 2.6.

Cuadro 2.6: Parámetros para el pozo F3 - Método *ad-hoc*.

	Solución 1	Solución 2
ω_f	0.004395	0.004395
ω_v	0.602700	0.392900
λ_{mf}	1.3531e-03	2.5421e-04
λ_{mv}	3.7860e-04	3.7860e-04
λ_{vf}	2.5421e-04	1.3531e-03
s	-4.9645	-4.9645
C_D	257.58	257.58
dL	1.9860	1.9860
Δ	1.0102	1.0102
α_t	5.6893e-03	5.6893e-03
α_p	320.11	320.11

Figura 2.27: Ajuste para el pozo F3 - Método *ad-hoc*.

Capítulo 3

Reducción del ruido en los datos mediante un enfoque de *machine-learning*

La caracterización de yacimientos es un problema inverso en el que se estiman los valores de los parámetros asociados a las propiedades del medio poroso. En este problema, los datos de presión y la log-derivada se utilizan en el proceso de ajuste. Sin embargo, la diferenciación numérica necesaria para generar la curva de la log-derivada es un problema mal planteado en el que el ruido de los datos puede propagarse en gran medida. Este ruido puede producir varios mínimos locales espurios en la función objetivo e impide obtener una aproximación precisa a los parámetros. Por lo tanto, se requiere un método eficiente de reducción de ruido para lograr la precisión deseada en los valores de los parámetros.

En este trabajo, exploramos tres métodos de reducción de ruido aplicándolos a datos de pruebas de pozo. Estos métodos se basan en diferencias finitas de varios pasos y splines, pero utilizando un enfoque de *machine-learning*.

Además, se presenta un nuevo método para reducir el ruido, el cual consta de generar una curva óptima a partir de una combinación lineal de múltiples curvas de datos, obtenidas de aproximaciones a la solución del problema inverso.

Se muestra cómo estos métodos que utilizan la información del modelo del problema directo, son más eficientes para la reducción del ruido. Los métodos propuestos pueden utilizarse en muchos problemas inversos en los que existe un modelo matemá-

tico que describe el fenómeno que se estudia, y que utiliza mediciones del fenómeno (datos) para estimar los parámetros desconocidos.

Los análisis y resultados presentados en esta sección fueron publicados en [Minutti et al. \(2018\)](#).

3.1. Descripción del problema

En muchos problemas inversos en los que los parámetros del modelo del problema directo se estiman usando datos observados, el ruido en los datos puede ser un problema importante, especialmente en modelos complejos (es decir, con un gran número de parámetros).

La caracterización de yacimientos de petróleo mediante el análisis de pruebas de pozos es uno de estos problemas, donde la log-derivada (la derivada de la presión medida en el pozo con respecto al logaritmo del tiempo), también tiene que ser ajustada ya que indica varias características del yacimiento. Sin embargo, cómo la derivación numérica es un problema mal planteado, el ruido de los datos es altamente propagado, lo que dificulta la obtención de una buena estimación de los parámetros del modelo.

Para la caracterización de yacimientos, el modelo utilizado para el problema directo es un modelo de Triple Porosidad-Doble Permeabilidad (TP-DP), en el que es necesario identificar hasta 10 parámetros para un caso real (ver [Gomez et al. \(2014a,b\)](#) y [Minutti et al. \(2016\)](#) en el que se describe la caracterización de yacimientos utilizando el modelo TP-DP). La solución del modelo para un conjunto dado de parámetros, produce la curva de la presión en el pozo en estudio. Debido al número de parámetros a identificar, este modelo puede dar más información del medio poroso que el modelo convencional de doble porosidad de Warren y Root ([Warren and Root, 1963](#)). Por lo tanto, es necesario un método de reducción de ruido para poder producir las llamadas curvas tipo de la presión y de la log-derivada, que tienen que ser ajustadas para producir una estimación precisa de los parámetros del modelo.

El algoritmo de Bourdet ([D. Bourdet and Pirard, 1989](#)), basado en un enfoque de diferencia finita, ha sido el método estándar para calcular la log-derivada de los datos de pruebas de pozos, pero ha demostrado ser ineficiente cuando el nivel de ruido es alto.

[H. S. Lane and Watson \(1991\)](#) abordan el problema de la reducción de ruido en

los datos de pruebas de pozo para un modelo más simple, utilizando splines en los que se optimizan los parámetros de suavizado. Más tarde [F. H. Escobar and Losada \(2004\)](#) también abordan el problema y comparan diferentes métodos de reducción de ruido, que se basan en esquemas de diferencias finitas (principalmente el algoritmo de Bourdet con diferentes valores de L y la derivada de polinomios interpolados) y splines, concluyendo que los splines son el método más efectivo.

El enfoque de utilizar diferencias finitas de varios pasos y/o splines, para reducir el ruido en los datos, está altamente extendido en muchas áreas diferentes. Sin embargo, las diferencias finitas basadas en la interpolación de polinomios en varios puntos, consideradas por Escobar *et al.* en realidad podrían aumentar el ruido de los datos.

En este estudio, comparamos tres métodos de reducción de ruido: Diferencias finitas de varios pasos, splines y mezcla de curvas sub-óptimas. Tanto en los splines cómo en las diferencias finitas multipaso, se propone un enfoque de *machine-learning*, que utiliza la información del modelo del problema directo, para determinar los parámetros óptimos de cada uno de los métodos. Siendo los splines un método ampliamente utilizado en diferentes áreas para reducir el ruido en los datos, también se estudia el efecto de otro tipo de splines: los splines de regresión.

En la primera parte del trabajo se presenta la metodología de cómo los algoritmos clásicos de diferencias finitas multipaso y de splines, se adaptan a un enfoque de *machine-learning* y se describe el método propuesto de mezcla de curvas sub-óptimas. En la segunda parte se presentan los resultados obtenidos al aplicar estos métodos de reducción de ruido a un amplio conjunto de datos, midiendo el error del método y también su efecto en el error de estimación de parámetros del problema inverso.

3.1.1. Metodología

Considerando un enfoque de *machine-learning*, utilizamos el modelo del problema directo (el modelo TP-DP) para generar un primer grupo de 1,000 conjuntos de datos sintéticos aleatorios, para los que posteriormente se agregó ruido Gaussiano. En cada uno de los métodos a estudiar, este conjunto de datos se utiliza para entrenar los algoritmos, es decir, para determinar el valor óptimo de los parámetros que minimizan el error entre los datos con y sin ruido. Posteriormente se genera un segundo grupo de 1,000 conjuntos de datos sintéticos aleatorios, para los que también se añade ruido. Este conjunto de datos es el conjunto de datos de validación cruzada y utilizamos

estos datos para comparar el rendimiento de los algoritmos propuestos, con el enfoque clásico y también con los parámetros optimizados. El propósito de determinar los parámetros óptimos de cada algoritmo con un conjunto de datos de entrenamiento, es ayudar al método de reducción de ruido a reproducir mejor el tipo de datos que provienen del modelo del problema directo.

El error del algoritmo de reducción de ruido en cada dato, se mide por:

$$Error = \left| \frac{d'_{algorithm} - d'_{noise-free}}{d'_{noise-free}} \right| \quad (3.1)$$

donde $d'_{noise-free}$ es la log-derivada de los datos sin ruido (en los ejemplos sintéticos) y $d'_{algorithm}$ es la log-derivada de los datos obtenidos después de aplicar el algoritmo de reducción de ruido a los datos con ruido.

Para comparar el rendimiento de cada método se utiliza la media y la mediana de $\log_{10}(Error)$ ya que, al estar midiendo un error relativo, estamos más interesados en el orden de magnitud del error. También debido a que la media es más sensible a la existencia de valores atípicos que la mediana, cuando se utilizan errores relativos, $\log_{10}(Error)$ tiene una distribución más simétrica, por lo que se ve menos afectada por los valores atípicos, siendo así la media y mediana más representativas de la muestra.

3.1.2. Diferencias finitas de varios pasos

En el caso de datos con ruido, no tiene sentido utilizar esquemas de diferencias finitas basados en la interpolación de polinomios (cómo es el caso de los esquemas presentados por [Fornberg \(1988\)](#)), ya que aumentan el ruido de la derivada numérica (esto ocurre en [Escobar et al.](#)). En su lugar, utilizamos los esquemas presentados por [Savitzky and Golay \(1964\)](#), donde el polinomio base, ajusta a los datos en el sentido de mínimos cuadrados, siguiendo una idea similar a la derivada de Lanczos. Basándonos en esto, elegimos un esquema de 9 puntos debido a que este es el número máximo de puntos que mostraron una mejora relevante en la reducción del ruido, para un conjunto de datos muestra. La fórmula de diferencia finita de 9 puntos de

Savitzky y Golay es:

$$f'_0 \approx \frac{(f_1 - f_{-1}) + 2(f_2 - f_{-2}) + 3(f_3 - f_{-3}) + 4(f_4 - f_{-4})}{60h} \quad (3.2)$$

Proponemos aquí otro método de diferencia finita de 9 puntos:

$$f'_0 \approx \frac{1}{4} \left(\omega_1 \frac{f_1 - f_{-1}}{2h} + \omega_2 \frac{f_2 - f_{-2}}{4h} + \omega_3 \frac{f_3 - f_{-3}}{6h} + \omega_4 \frac{f_4 - f_{-4}}{8h} \right) \quad (3.3)$$

donde $\sum_{i=1}^4 \omega_i = 1$. Los valores de ponderación ω_i se determinan minimizando el error de la Ecuación 3.1, obtenido con el conjunto de datos de entrenamiento. Posteriormente, los errores de ambos métodos se comparan utilizando el conjunto de datos de validación cruzada.

3.1.3. Splines

Para los splines, consideramos tres casos. En primer lugar, un enfoque clásico en el que se utiliza un spline cúbico y el parámetro de suavizado se determina mediante validación cruzada, es decir, para cada uno de los 1,000 conjuntos de datos, se utiliza una parte de los datos para determinar los coeficientes óptimos del spline, y otra parte para determinar el parámetro de suavizamiento, seleccionando en cada caso el valor que minimiza el error cuadrado medio.

En el segundo enfoque, utilizamos optimización para determinar un único valor del parámetro de suavizamiento, que minimiza el error de todo el conjunto de datos de entrenamiento y el mismo valor se utilizará en el conjunto de datos de validación cruzada.

En el tercer enfoque utilizamos splines de regresión (ver [Monahan \(2001\)](#)) donde, a diferencia de los splines cúbicos, es posible utilizar diferentes tipos de funciones base para aproximar los datos. Para obtener un buen resultado, las funciones base deben ser similares o estar relacionadas con las funciones subyacentes tras el conjunto de datos a reproducir, en este caso, el modelo del problema directo que se utiliza para ajustar los datos medidos. En el modelo TP-DP, las aproximaciones asintóticas del modelo en tiempos tempranos y tardíos (que aparecen en [Camacho-Velázquez et al. \(2005\)](#)), se utilizan como funciones base. El parámetro de suavizamiento se

determina de la misma manera que en el caso anterior, utilizando optimización para obtener el valor que minimiza el error de todo el conjunto de datos de entrenamiento.

El rendimiento de cada algoritmo se determina midiendo el error de la Ecuación 3.1 en el conjunto de datos de validación cruzada.

3.1.4. Mezcla de curvas sub-óptimas

Proponemos aquí un método que llamamos *mezcla de curvas sub-óptimas*, que consiste en utilizar los datos con ruido (presión medida en el pozo) para realizar una serie de optimizaciones rápidas (con pocas iteraciones del algoritmo) para resolver el problema inverso con diferentes puntos iniciales, con el fin de obtener aproximaciones a óptimos locales, es decir, obtener diferentes valores de los parámetros del modelo directo, que puedan reproducir los datos de forma aproximada. Estos datos aproximados son las curvas sub-óptimas.

Estas curvas sub-óptimas se combinan en una única curva usando una combinación lineal, donde el peso de cada curva se determina minimizando el error cuadrático entre la curva combinada y los datos con ruido. Basándonos en el modelo de regresión lineal múltiple, podemos obtener el vector de los pesos (b) usando la ecuación normal de regresión:

$$b = (X'X)^{-1}X'y \quad (3.4)$$

donde, en este modelo, X es una matriz cuyas columnas son los diferentes valores de las curvas sub-óptimas, X' es su traspuesta y y es un vector que contiene los datos con ruido.

Se añade un parámetro de regularización (λ) al modelo para evitar problemas de matrices singulares (debido a curvas sub-óptimas similares) y también para evitar el problema de sobre-ajuste, que puede llevar a reproducir el ruido en los datos. Así, aplicando la regularización de Tikhonov, el modelo final para estimar los pesos de cada curva es:

$$b = (X'X + I\lambda)^{-1}X'y \quad (3.5)$$

donde I es la matriz de identidad.

La curva de datos con ruido reducido y su log-derivada, se obtienen con $b'X$ y $b'X_d$, donde X_d es la matriz cuyas columnas son la log-derivada de las columnas de X .

Tanto el número de curvas sub-óptimas a utilizar cómo el valor del parámetro de

regularización se obtienen optimizando (minimizando) el error del conjunto de datos de entrenamiento. Para este caso, se obtuvieron 40 curvas sub-óptimas del modelo TP-DP para cada uno de los 1,000 casos del conjunto de datos de entrenamiento.

Una vez determinados estos parámetros, el proceso se aplica al conjunto de datos de validación cruzada, para estimar el error del método propuesto.

3.1.5. Proceso de entrenamiento y comparativo de los métodos

El siguiente es un resumen del proceso de entrenamiento para cada algoritmo y cómo se realiza la comparación de la reducción de ruido alcanzada por cada método.

1. Se crean conjuntos de datos de entrenamiento y validación cruzada. Cada conjunto posee 1,000 conjuntos de datos sintéticos que se generan al resolver el problema directo, utilizando valores de parámetros aleatorios y añadiendo ruido aleatorio a los datos.
2. Para el método de splines cúbicos, el parámetro de regularización se estima dinámicamente para cada conjunto de datos mediante validación cruzada, donde parte de los datos se utilizan para encontrar el valor óptimo del parámetro de regularización minimizando el error cuadrático.
3. Para los algoritmos entrenados, el conjunto de datos de entrenamiento se utiliza para obtener los valores de los parámetros estáticos de cada método (*e.g.*, el suavizamiento), minimizando el error entre los datos con ruido y los libres de ruido, utilizando optimización.
 - En el caso de diferencias finitas de varios pasos, se utiliza optimización para estimar los pesos óptimos de cada diferencia finita.
 - En el caso de los algoritmos basados en splines, la optimización se utiliza para estimar el valor óptimo de los parámetros de regularización.
 - En el caso de la mezcla de curvas sub-óptimas, la optimización es utilizada para estimar el número de curvas sub-óptimas a usar en la combinación lineal, así como para el valor del parámetro de regularización.

4. El conjunto de datos de validación cruzada se utiliza para estimar el error de cada método, aplicando la Ecuación 3.1, donde se utilizan las ponderaciones y los parámetros de regularización (previamente estimados con el conjunto de datos de entrenamiento). Este conjunto de datos también se utiliza para estimar el error de los métodos clásicos.

3.1.6. Error de los parámetros del modelo del problema directo

Para medir el efecto de los diferentes métodos de reducción de ruido en la estimación de los parámetros del modelo del problema directo, resolvemos el problema inverso mediante optimización siguiendo el procedimiento descrito en [Minutti et al. \(2016\)](#), para así obtener una estimación de los parámetros del modelo para los datos sintéticos en el conjunto de datos de validación cruzada.

Adicionalmente, proporcionamos cómo punto inicial para el algoritmo de optimización, el valor real de los parámetros con los que se generaron los datos libres de ruido (casos sintéticos). Debido a que el algoritmo de optimización tiene cómo punto de partida el valor real de los parámetros, cualquier error en las estimaciones de los parámetros se debe al ruido en los mismos.

Para medir el error de cada parámetro se utiliza la siguiente fórmula:

$$err(\hat{\theta}) = \frac{|\theta - \hat{\theta}|}{\text{máx}_{\theta} - \text{mín}_{\theta}} \quad (3.6)$$

donde θ es el valor verdadero del parámetro, $\hat{\theta}$ es el valor estimado dado por el algoritmo de optimización, y $\text{máx}_{\theta} - \text{mín}_{\theta}$ es la longitud del intervalo de valores posibles que puede tomar el parámetro. Así, la Ecuación 3.6 mide el error absoluto expresado entre 0 y 1.

3.2. Resultados

Usando el conjunto de datos de validación cruzada, la media y la mediana de $\log_{10}(\text{Error})$ se determinan para cada uno de los métodos descritos anteriormente,

obteniendo los valores del Cuadro 3.1, siendo los métodos clásicos el esquema de 9 puntos de Savitzky y Golay (FD-9_CLS) y los splines cúbicos (SPL_CLS).

Cuadro 3.1: Media y mediana de $\log_{10}(Error)$ para los algoritmos probados.

	Diferencias finitas Multi-paso		Smoothing Splines			Mezcla de curvas sub-óptimas
	FD-9_CLS	FD-9	SPL_CLS	SPL	REG_SPL	MIX_OPT_C
Media	-1.6257	-1.7406	-2.1575	-2.1846	-2.2604	-2.5310
Mediana	-1.5021	-1.6342	-2.1716	-2.1875	-2.2846	-2.5974

Puede observarse que, al igual que en las diferencias finitas de varios pasos y también en los splines, los métodos en los que los parámetros de los algoritmos se determinaron utilizando un conjunto de datos de entrenamiento (FD-9, SPL, REG_SPL), son más eficaces que el uso de un método clásico (FD-9_CLS, SPL_CLS). También se observa que el método de splines de regresión (REG_SPL) es más eficiente que los otros dos métodos de splines, y el método de mezcla de curvas sub-óptimas (MIX_OPT_C) es mejor que cualquier otro de los algoritmos analizados. Específicamente, la mezcla de curvas sub-óptimas tiene sólo el 53.6 % del ruido que tienen los splines de regresión, el segundo mejor método ($\frac{10^{-2.5310}}{10^{-2.2604}} \approx 0.5363$), lo que muestra que el uso de la información del modelo del problema directo, produce un método más eficiente para la reducción del ruido.

Las Figuras 3.1 y 3.2 muestran los boxplots e histogramas de los métodos en los que se entrenaron los algoritmos: Diferencias finitas multi-paso (FD-9), splines cúbicos con parámetros de suavizamiento optimizados (SPL), splines de regresión con parámetros de suavizamiento optimizados (REG_SPL) y la mezcla de curvas sub-óptimas (MIX_OPT_C).

En las Figuras 3.3-3.5 se presentan tres ejemplos tomados del conjunto de datos de validación cruzada, donde la presión y su log-derivada se grafican en una escala log-log, usando los tres mejores métodos de reducción de ruido: splines entrenados (splines cúbicos y de regresión), y mezcla de curvas sub-óptimas. También se presentan los datos con ruido y su derivada, calculados con un método de diferencias finitas (Bourdet).

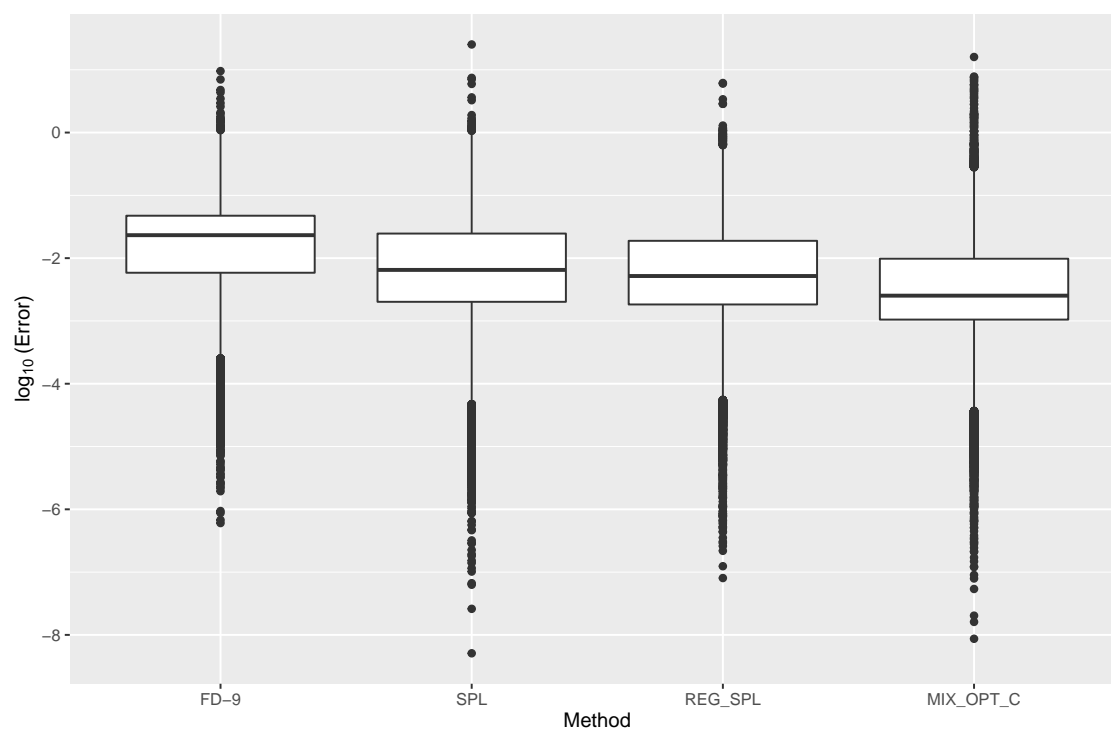


Figura 3.1: Boxplot de $\log_{10}(\text{Error})$ para los algoritmos entrenados.

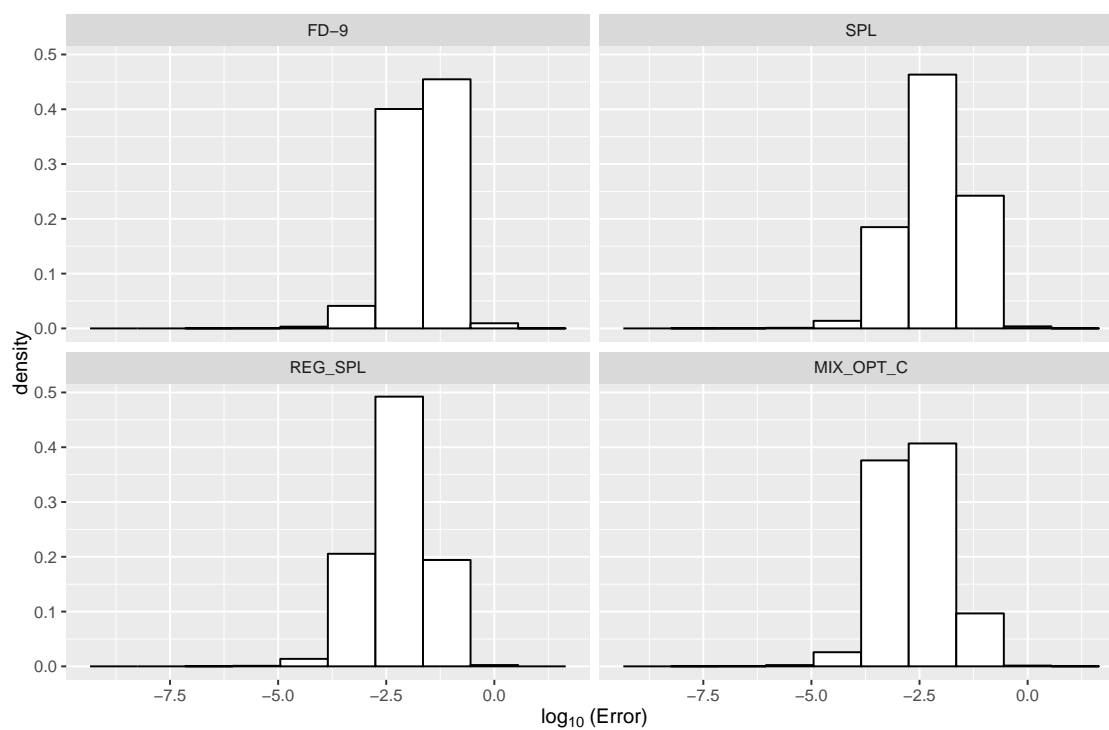


Figura 3.2: Histograma de $\log_{10}(\text{Error})$ para los algoritmos entrenados.

La mayor reducción de ruido se obtiene con la mezcla de curvas sub-óptimas. Los splines cúbicos y los splines de regresión tienen problemas de oscilaciones, especialmente en los valles de las curvas de las log-derivadas.

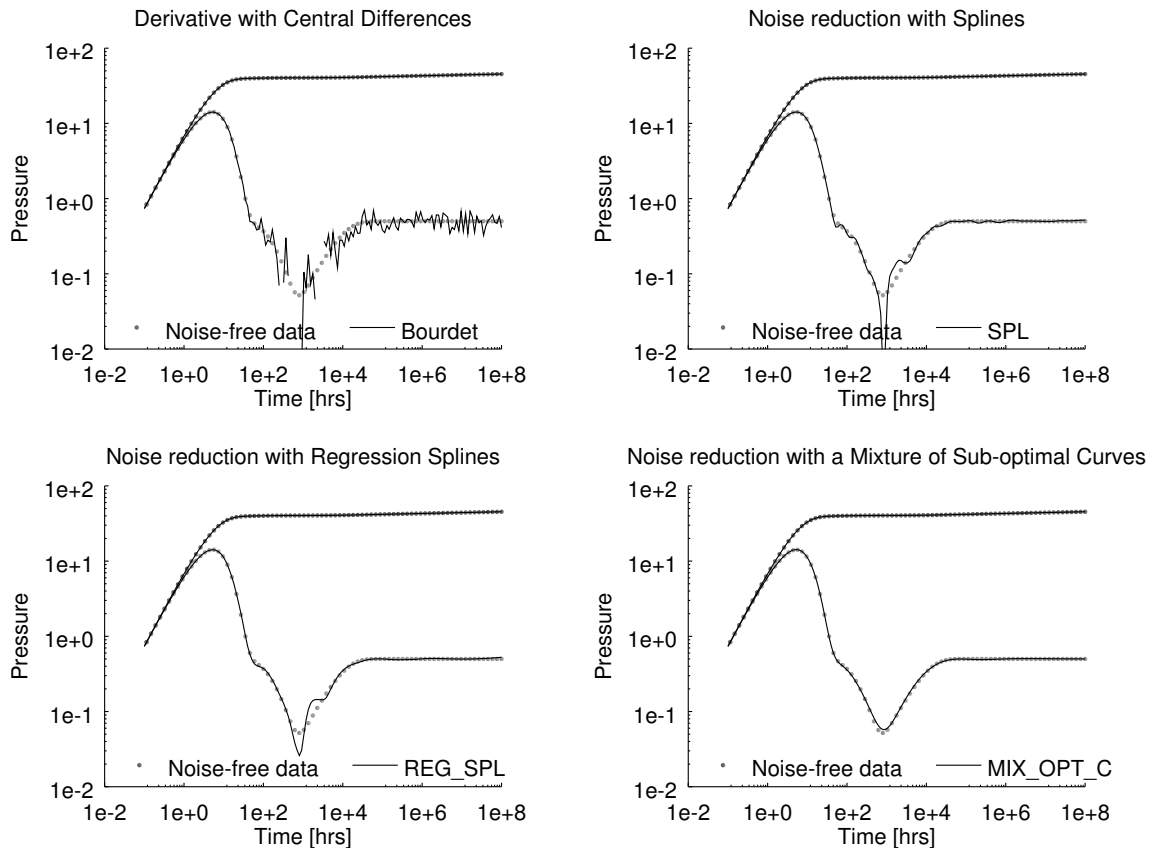


Figura 3.3: Ejemplo 1: Reducción de ruido con los tres mejores métodos.

En el Cuadro 3.2 se muestra el error medio de estimación (utilizando la Ecuación 3.6) de los parámetros del modelo del problema directo, cuando se resuelve el problema inverso utilizando los datos con ruido en el conjunto de datos de validación cruzada y también cuando se utiliza los datos con ruido reducido generados mediante los algoritmos entrenados. Al proporcionar al algoritmo de optimización el valor real de los parámetros como punto de partida, el error de estimación se debe al ruido en los datos.

Para cada parámetro se puede observar una reducción significativa en el error de

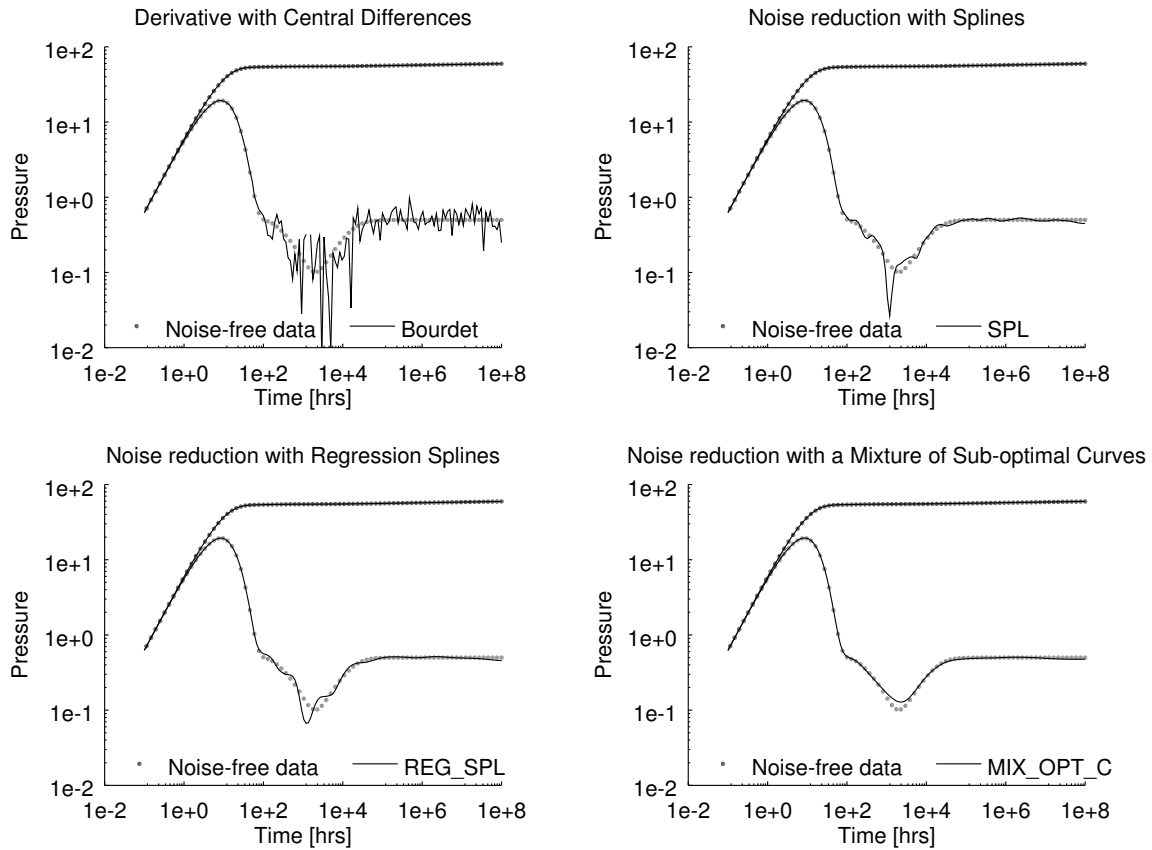


Figura 3.4: Ejemplo 2: Reducción de ruido con los tres mejores métodos.

Cuadro 3.2: Error medio de estimación de los parámetros del modelo con diferentes métodos de reducción de ruido.

Método	ω_f	ω_v	λ_{mf}	λ_{mv}	λ_{vf}	κ	s	C_D
NOISY DATA	0.1862	0.1844	0.1008	0.0966	0.1615	0.1567	0.0148	0.0011
SPL	0.0955	0.0962	0.0403	0.0355	0.0775	0.0670	0.0012	0.0003
REG_SPL	0.0923	0.0945	0.0395	0.0365	0.0800	0.0639	0.0011	0.0003
MIX_OPT_C	0.0666	0.0623	0.0360	0.0340	0.0719	0.0488	0.0005	0.0001

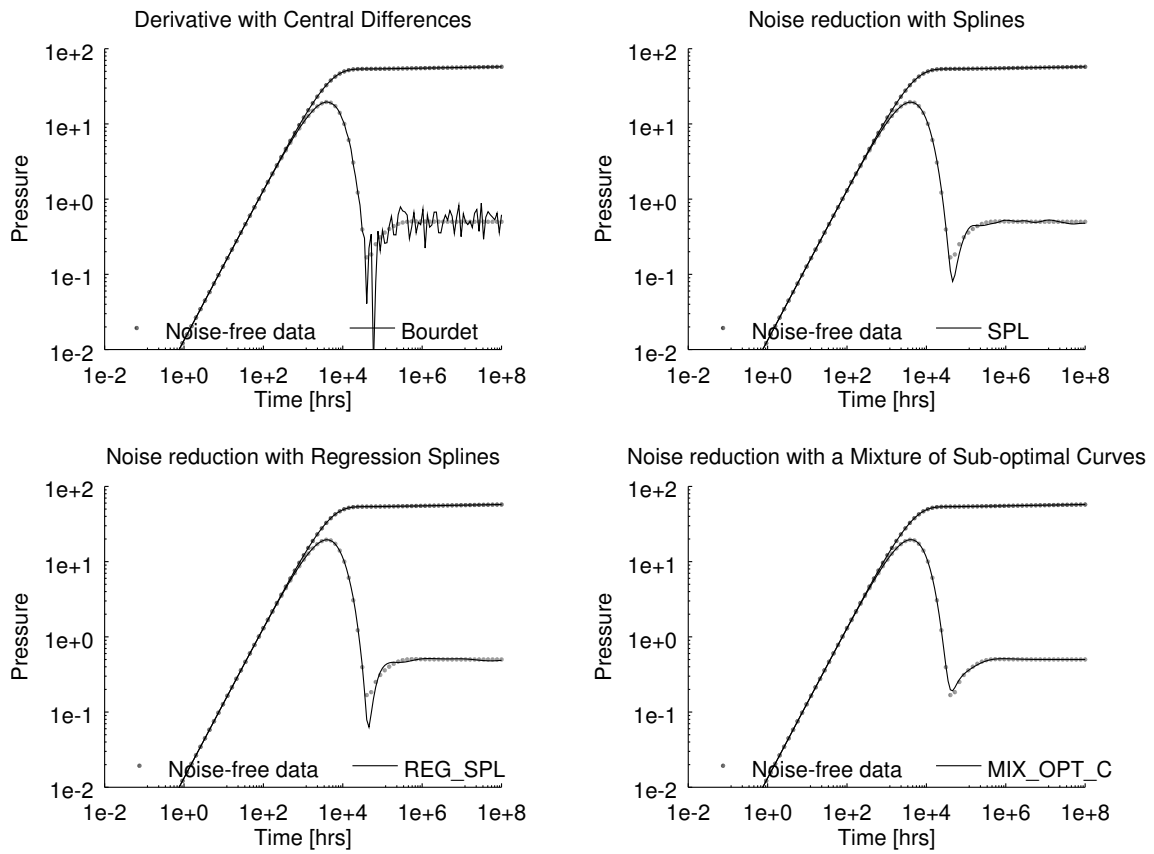


Figura 3.5: Ejemplo 3: Reducción de ruido con los tres mejores métodos.

estimación, comparado con los resultados de los datos con ruido. Para parámetros como ω 's, λ_{mf} , λ_{mv} y κ , el error medio de estimación es un tercio del error de los datos con ruido, cuando se utilizó la mezcla de curvas sub-óptimas.

Aunque los splines de regresión fueron un mejor método de reducción de ruido que los splines cúbicos, no se presentó diferencia significativa en el error medio de estimación de los parámetros.

Los resultados del Cuadro 3.2 muestran que:

- Si no se utiliza un método de reducción de ruido adecuado, incluso con una buena estimación inicial de los parámetros (o un buen algoritmo de optimización), debido al ruido, sólo podemos esperar una precisión limitada en la estimación de los parámetros del problema inverso.
- Incluso si un método de reducción de ruido funciona mejor (es decir, un error menor entre los datos con ruido y los datos libres de ruido), puede no tener un efecto significativo en el problema inverso, este es el caso de splines cúbicos y de regresión, donde los de regresión tiene un error menor que el los cúbicos (Cuadro 3.1), pero la mejora en el error medio de estimación de los parámetros es casi la misma. Esto puede significar que parte de la mejora realizada por el algoritmo podría estar en un área (un intervalo de tiempo en la curva de presión) donde la solución del modelo del problema directo, puede tener poca sensibilidad a los parámetros.
- El método de curvas sub-óptimas no tiene el problema mencionado anteriormente, debido a que se realizan optimizaciones rápidas, resolviendo el problema inverso muchas veces y es por ello que la forma de la curva cambia principalmente en áreas que son sensibles a los parámetros.

3.3. Conclusiones

Entrenando los algoritmos con un conjunto de datos sintéticos se puede ver un error menor entre los datos con y sin ruido, en comparación con un enfoque clásico. Especialmente cuando se utiliza un spline de regresión (con funciones base derivadas del modelo del problema directo), se obtienen mejores resultados que con un spline cúbico clásico.

El método propuesto en este trabajo, mezcla de curvas sub-óptimas, combina las ideas de utilizar funciones base que provienen del propio modelo del problema directo y de entrenar el algoritmo para obtener un conjunto de parámetros optimizados para el problema. Este método fue el más eficiente para la reducción de ruido, teniendo sólo el 53.6 % del ruido que tiene las splines de regresión, que fue el segundo mejor método.

Cuando se resolvió el problema inverso para el conjunto de datos de validación cruzada, observamos el impacto de los diferentes métodos de reducción de ruido en el problema de estimación de parámetros, siendo la mezcla de curvas sub-óptimas también el mejor método, teniendo para muchos parámetros, hasta un tercio del error de estimación, en comparación con el error de los datos con ruido. Esta mejora en la estimación de los parámetros es probablemente la característica más relevante del método propuesto, ya que, independientemente de la mejora que se pueda tener en la representación de los datos al disminuir su ruido, es la estimación de los parámetros el principal interés en este tipo de problemas inversos.

Debido a que en muchos problemas inversos existe un modelo matemático que se utiliza para aproximar o reproducir un fenómeno real (problema directo), los métodos presentados en este trabajo pueden ser aplicados para resolverlos, ya que el modelo del problema directo puede ser utilizado para entrenar los algoritmos para reducir el ruido en los datos observados. Especialmente mezclar diferentes soluciones sub-óptimas del problema inverso es un enfoque que permite reducir el ruido en los datos de forma más consistente con los fenómenos subyacentes que modela el problema directo, y por ello pueden ser mejor utilizados en un problema inverso.

Estos métodos, basados en el modelo del problema directo, son especialmente importantes cuando hay un alto número de parámetros a identificar, ya que en estos casos aumentan la probabilidad de ajustar la solución del problema inverso al ruido en los datos. Por lo tanto, un método de reducción de ruido que sea consistente con el modelo del problema directo, ayuda a mejorar la precisión de los parámetros estimados.

En la aplicación estudiada, caracterización de yacimientos de petróleo, las estimaciones de parámetros pueden ser utilizadas para tomar decisiones sobre pronósticos de producción. Es por esto que una estimación más precisa de los parámetros puede tener un gran impacto en la producción de un pozo.

Conclusiones

En este trabajo se abordaron problemas inversos cuyas soluciones presentan diferentes problemáticas, cómo lo son la existencia de múltiples óptimos locales, altos tiempos de computo y la necesidad de estimar un gran número de parámetros, con alta precisión.

Para dar solución a estos problemas se proponen enfoques cuya base es una interpretación estadística del problema, la cual permite entender el efecto de cada uno de los parámetros en el problema a resolver, ya sea debido a que el nuevo enfoque es el que proporciona la re-interpretación de los parámetros (el caso de tomografía hidráulica) o porque se estudia el efecto global de cada uno de los parámetros para poder determinar el método óptimo de su estimación (el caso del análisis de pruebas de pozo).

Aunque ambos enfoques requieren de un estudio profundo del problema a resolver, la metodología propuesta permite mejorar significativamente la precisión de los parámetros a estimar, así cómo el tiempo de computo necesario.

En todos los casos se realizaron múltiples experimentos, los cuales permiten explorar el espacio de parámetros de cada uno de los problemas, para así asegurar que los resultados son robustos y representativos. Adicionalmente, los métodos desarrollados se aplicaron a casos reales.

De los resultados de los experimentos se obtienen las siguientes conclusiones:

- Al estudiar las necesidades de cada uno de los problemas se puede plantear un enfoque estadístico de re-interpretación de los parámetros, que permita dar solución al problema. En el caso de tomografía hidráulica, se observa que gran parte de la problemática en su solución, es debida al alto número de parámetros a estimar, por lo que se sugiere una re-parametrización del

problema, la cual es consistente con el fenómeno físico que representa, por lo que la nueva parametrización permite mantener una interpretación física de los parámetros, a la vez que mantiene una relación con métodos previos de solución del problema.

- El caso de la incertidumbre en la solución se aborda mediante el uso de una solución por consenso de las diferentes soluciones (el estimador de Bayes), que muestra mejorar significativamente las predicciones del modelo, ya que la falta de suficientes datos para el modelado inverso puede provocar que diferentes soluciones reproduzcan los datos, siendo más eficiente el uso de un estimador que toma en cuenta esta incertidumbre en la solución, resultando en errores mucho menores en las predicciones generadas entre datos observados y simulados.
- Para el caso de análisis de pruebas de pozo, el análisis de sensibilidad global de los parámetros nos permite conocer los parámetros que más influyen en la respuesta del modelo del problema directo, dando información de que parámetros pueden ser estimados de forma independiente debido a su alta sensibilidad, que parte de los datos es más importante para esa estimación, parámetros que requieren una estimación conjunta.
- Lo anterior simplifica el problema de optimización a resolver, mediante una optimización *ad-hoc*, que determina los parámetros más sensibles al inicio (para que estos no afecten la estimación de los menos sensibles de manera posterior).
- Los análisis propuestos también muestran el nivel de error de cada uno de los parámetros del modelo, así como el rango en que será mayor su error, permitiendo verificar si los errores obtenidos son consistentes con los esperados, permitiendo diagnosticar y validar la eficiencia del modelado inverso.
- En ambos problemas se observa que la forma de las funciones objetivo se puede mejorar significativamente mediante la inclusión de derivadas en la función objetivo (primer caso) o mediante la asignación de distintos pesos en el conjunto de datos a reproducir, dando prioridad a los datos que favorecen la estimación de los parámetros menos sensibles.
- El enfoque propuesto para abordar el ruido en los datos, donde las funciones base que son usadas para filtrar los datos provienen del modelo directo, permite

mantener una consistencia entre los datos filtrados y el problema inverso, lo cual mejora significativamente la estimación de los parámetros con respecto a otras estrategias de filtrado, que es el principal objetivo en la solución de los problemas inversos.

Bibliografía

- Abdassah, D. (1986). Triple-porosity systems for representing naturally fractured reservoirs. *Society of Petroleum Engineers*, 1.
- Agarwal, R. G., Al-Hussainy, R., and Jr, H. R. (1970). An investigation of wellbore storage and skin effect in unsteady liquid flow: I. analytical treatment. *Society of Petroleum Engineers Journal*, 10:279 – 290.
- Balankin, A. S. and Elizarraraz, B. E. (2012). Map of fluid flow in fractal porous medium into fractal continuum flow. *Phys. Rev. E*, 85:056314.
- Barenblatt, G. I., Zheltov, I. P., and Kochina, I. N. (1960). Basic concepts in the theory of seepage of homogeneous liquids in fissured rocks. *Journal of Applied Mathematics and Mechanics*, pages 1286 – 1303.
- Berg, S. J. and Illman, W. A. (2011). Capturing aquifer heterogeneity: Comparison of approaches through controlled sandbox experiments. *Water Resources Research*, 47(9).
- Berg, S. J. and Illman, W. A. (2015). Comparison of hydraulic tomography with traditional methods at a highly heterogeneous site. *Groundwater*, 53(1):71–89.
- Bohling, G. C., Zhan, X., Butler Jr., J. J., and Zheng, L. (2002). Steady state analysis of tomographic pumping tests for characterization of aquifer heterogeneities. *Water Resources Research*, 38(12):60–1–60–15.
- Bonalde, I. and Ramones, M. (1994). A robust algorithm for parameter estimation in well tests. *Society of Petroleum Engineers*, 2.

- Bourgeois, M. and Horne, R. (1993). Well-test-model recognition with laplace space. *SPE Formation Evaluation*, 8(1):17 – 25.
- Bourgeois, M., of Petroleum Engineering, S. U. D., and Acquitaine, E. (1992). *Well Test Interpretation Using Laplace Space Type Curves*. Stanford University.
- Brauchler, R., Liedl, R., and Dietrich, P. (2003). A travel time based hydraulic tomographic approach. *Water Resources Research*, 39(12).
- Camacho-Velázquez, R., Vásquez-Cruz, M., Castrejon-Aivar, R., and Arana-Ortiz, V. (2005). Pressure transient and decline curve behavior in naturally fractured-vuggy carbonate reservoirs. *SPE Reservoir Evaluation & Engineering* 8, pages 95 – 111.
- Cardiff, M., Barrash, W., Kitanidis, P., Malama, B., Revil, A., Straface, S., and Rizzo, E. (2009). A potential-based inversion of unconfined steady-state hydraulic tomography. *Groundwater*, 47(2):259–270.
- Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., and Slooten, L. J. (2005). Inverse problem in hydrogeology. *Hydrogeology Journal*, 13(1):206–222.
- Castagna, M. and Bellin, A. (2009). A bayesian approach for inversion of hydraulic tomographic data. *Water Resources Research*, 45(4).
- Chen, Z., Huan, G., and Ma, Y. (2006). *Computational methods for multiphase flows in porous media*. Society for Industrial and Applied Mathematics.
- D. Bourdet, J. A. A. and Pirard, Y. M. (1989). Use of pressure derivative in well test interpretation. *Society of Petroleum Engineers*, 4.
- Da Prat, G., Cinco-Ley, H., and Ramey, H. J. (1981). Decline curve analysis using type curves for two-porosity systems. *SPEJ*, pages 354 – 362.
- Dastan, A. and Horne, R. N. (2010). A new look at nonlinear regression in well-test interpretation. In *SPE Annual Technical Conference and Exhibition, 19-22 September, Florence, Italy*.
- Dehghani, K., Harris, P., Edwards, K., and Dees, W. (1999). Modeling a vuggy carbonate reservoir. *McElroy Field, West Texas, AAPG Bulletin* 83, pages 19 – 42.

- Everdingen, A. V. (1953). The skin effect and its influence on the productive capacity of a well. *Society of Petroleum Engineers*, 5.
- F. H. Escobar, J. M. N. and Losada, H. D. (2004). Evaluation of pressure derivative algorithms for well-test analysis. In *SPE International Thermal Operations and Heavy Oil Symposium and Western Regional Meeting, 16-18 March, Bakersfield, California*.
- Fair, W. B. (1981). Pressure buildup analysis with wellbore phase redistribution. *Society of Petroleum Engineers*, 21.
- Fornberg, B. (1988). Generation of finite difference formulas on arbitrarily spaced grids. *Mathematics of Computation*, 51(184):699–699.
- Genrich, J. and Sommer, F. (1989). Novel approach to sensitivity analysis. *Society of Petroleum Engineers*, 41.
- Gomez, S., Camacho, R., Vásquez, M., Ramos, G., del Castillo, N., and Mesejo, A. (2014a). Well test characterization of naturally fractured vuggy reservoirs, with a global optimization method. In *OTC-24762-MS. Offshore Technology Conference-Asia*.
- Gomez, S., Gosselin, O., and Barker, J. W. (2001). Gradient-based history matching with a global optimization method. *Society of Petroleum Engineers*, 6.
- Gomez, S. and Levy, A. (1982). The tunnelling method for solving the constrained global optimization problem with several non-connected feasible regions. *Lecture Notes in Mathematics*, 909.
- Gomez, S., Ramos, G., A. Mesejo, R. C., Vásquez, M., and del Castillo, N. (2014b). Well test analysis of naturally fractured vuggy reservoirs with an analytical triple porosity-double permeability model and a global optimization method. *Oil & Gas Science and Technology*, 69(4):653 – 671.
- Gomez, S., Ramos, G., Mesejo, A., Camacho, R., Vásquez, M., and Castillo, N. (2013). Study of the characterization of naturally fractured vuggy reservoirs, with totally penetrated wells using global optimization. synthetic and real data. Technical Report 156, IIMAS-UNAM.

- Gottlieb, J. and Dietrich, P. (1995). Identification of the permeability distribution in soil by hydraulic tomography. *Inverse Problems*, 11(2):353.
- Gringarten, A., Bourdet, D., Landel, P., and Kniazeff, V. (1979). A comparison between different skin and wellbore storage type-curves for early-time transient analysis. In *SPE 8205*, Las Vegas.
- Gurpinar, O., Kalbus, J., and List, D. (2000). Numerical modeling of a large, naturally fractured oil complex. In *SPE International Petroleum Conference and Exhibition in Mexico*, Villahermosa, Mexico.
- H. S. Lane, W. J. L. and Watson, A. T. (1991). An algorithm for determining smooth, continuous pressure derivatives from well-test data. *Society of Petroleum Engineers*, 6.
- Harbaugh, A. W. and Geological Survey (U.S.) (2005). *MODFLOW-2005, the U.S. Geological Survey modular ground-water model the ground-water flow process*. U.S. Dept. of the Interior, U.S. Geological Survey Reston, Va.
- Heath, M. T. (1996). *Scientific Computing: An Introductory Survey*. McGraw-Hill Higher Education, 2nd edition.
- Hoeksema, R. J. and Kitanidis, P. K. (1984). An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. *Water Resources Research*, 20(7):1003–1020.
- Illman, W. A., Craig, A. J., and Liu, X. (2008). Practical issues in imaging hydraulic conductivity through hydraulic tomography. *Groundwater*, 46(1):120–132.
- Illman, W. A., Liu, X., and Craig, A. (2007). Steady-state hydraulic tomography in a laboratory aquifer with deterministic heterogeneity: Multi-method and multiscale validation of hydraulic conductivity tomograms. *Journal of Hydrology*, 341(3):222 – 234.
- Illman, W. A. and Neuman, S. P. (2001). Type curve interpretation of a cross-hole pneumatic injection test in unsaturated fractured tuff. *Water Resources Research*, 37(3):583–603.

- Illman, W. A., Zhu, J., Craig, A. J., and Yin, D. (2010). Comparison of aquifer characterization approaches through steady state groundwater model validation: A controlled laboratory sandbox study. *Water Resources Research*, 46(4).
- Joseph, J. A. and Koederitz, L. F. (1983). Applied pressure transient analysis using nonlinear regression methods: Case studies. *Society of Petroleum Engineers*.
- Kazemi, H. (1969). Pressure transient analysis of naturally fractured reservoirs with uniform fracture distribution. *SPE Journal* 9, pages 451 – 462.
- Liu, J., Bodvarsson, G., and Wu, Y.-S. (2002). Analysis of flow behavior in fractured lithophysal reservoirs. *Journal of Contaminant Hydrology*, 62-63.
- Liu, X., Zhou, Q., Birkholzer, J., and Illman, W. A. (2013). Geostatistical reduced-order models in underdetermined inverse problems. *Water Resources Research*, 49(10):6587–6600.
- Mao, D., Yeh, T.-C. J., Wan, L., Wen, J.-C., Lu, W., Lee, C.-H., and Hsu, K.-C. (2013). Joint interpretation of sequential pumping tests in unconfined aquifers. *Water Resources Research*, 49(4):1782–1796.
- Menekse, K., Onur, M., and Zeybek, M. (1995). Analysis of well tests from naturally fractured reservoirs by automated type-curve matching. In *Middle East Oil Show, 11-14 March, Bahrain*.
- Minutti, C., Gomez, S., and Ramos, G. (2018). A machine-learning approach for noise reduction in parameter estimation inverse problems, applied to characterization of oil reservoirs. *Journal of Physics: Conference Series*, 1047:012010.
- Minutti, C., Illman, W. A., and Gomez, S. (2019). An algorithm for hydraulic tomography based on a mixture model. *Lecture Notes in Computer Science*, 11538:471–486.
- Minutti, C., Ramos, G., Gomez, S., Camacho, R., Vazquez, M., and del Castillo, N. (2016). Robust characterization of naturally fractured carbonate reservoirs through sensitivity analysis and noise propagation reduction. In *SPE Latin America and Caribbean Heavy and Extra Heavy Oil Conference, 19-20 October, Lima, Peru*.

- Monahan, J. F. (2001). *Numerical Methods of Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- Nair, N. (2008). *Measurement and modeling of multiscale flow and transport through large-vug Cretaceous carbonates*. Proquest, Umi Dissertation Publishing.
- Onur, M. (1998). Numerical laplace transformation of sampled data for well-test analysis. *SPE Reservoir Evaluation & Engineering*, 4:268 – 277.
- Qiu, P., Hu, R., Hu, L., Liu, Q., Xing, Y., Yang, H., Qi, J., and Ptak, T. (2019). A numerical study on travel time based hydraulic tomography using the sirt algorithm with cimmino iteration. *Water*, 11(5).
- Renshaw, C. E. (1996). Estimation of fracture zone geometry from steady-state hydraulic head data using iterative sequential cokriging. *Geophysical Research Letters*, 23(19):2685–2688.
- Savitzky, A. and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36:1627–1639.
- Schiozer, D., Netto, S. A., Ligerio, E., and Maschio, C. (2005). Sensitivity analysis of two-phase reservoir history matching. *Petroleum Society of Canada*, 44.
- Stehfest, H. (1970). Algorithm 386: Numerical inversion of laplace transforms. *Communications of the ACM*, 13(1):47 – 49.
- Ugray, Z., Lasdon, L. S., Plummer, J. C., Glover, F. W., Kelly, J. P., and Martí, R. (2007). Scatter search and local nlp solvers: A multistart framework for global optimization. *INFORMS Journal on Computing*, 19(3):328–340.
- Vieira, P. and Rosa, A. (1996). A comparison of several methods for automated well test analysis. *Society of Petroleum Engineers*, 1.
- Vik, B., Sylta, K. E., and Skauge, A. (2012). Connectivity in vuggy carbonates, new experimental methods and applications. *Transport in Porous Media*, 93:561 – 575.
- Wang, H. and Anderson, M. (1995). *Introduction to Groundwater Modeling: Finite Difference and Finite Element Methods*. Elsevier Science.

- Warren, J. and Root, P. (1963). The behavior of naturally fractured reservoirs. *Society of Petroleum Engineers*, 3.
- Watson, A. (1989). Sensitivity analysis of two-phase reservoir history matching. *Society of Petroleum Engineers*, 4.
- Wilkinson, D. (1992). Pressure transient parameter estimation in the laplace domain. In *67th Annual Technical Conference and Exhibition of the SPE*, pages 601 – 611, Washington, DC.
- Wu, Y.-H., Liu, H., and Bodvarsson, G. (2004). A triple-continuum approach for modeling flow and transport processes in fractured rock. *Journal of Contaminant Hydrology* 73, pages 145 – 179.
- Wu, Y.-S., Qin, G., Ewing, R. E., Efendiev, Y., and Kang, Z. (2006). A multiple-continuum approach for modeling multiphase flow in naturally fractured vuggy petroleum reservoirs. In *International Oil & Gas Conference and Exhibition in China, 5-7 December, Beijing, China*.
- Xiang, J., Yeh, T.-C. J., Lee, C.-H., Hsu, K.-C., and Wen, J.-C. (2009). A simultaneous successive linear estimator and a guide for hydraulic tomography analysis. *Water Resources Research*, 45(2).
- Yeh, T.-C. J., Jin, M., and Hanna, S. (1996). An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields. *Water Resources Research*, 32(1):85–92.
- Yeh, T.-C. J. and Liu, S. (2000). Hydraulic tomography: Development of a new aquifer test method. *Water Resources Research*, 36(8):2095–2105.
- Zha, Y., Yeh, T.-C. J., Illman, W. A., Onoe, H., Mok, C. M. W., Wen, J.-C., Huang, S.-Y., and Wang, W. (2017). Incorporating geologic information into hydraulic tomography: A general framework based on geostatistical approach. *Water Resources Research*, 53(4):2850–2876.
- Zha, Y., Yeh, T.-C. J., Illman, W. A., Zeng, W., Zhang, Y., Sun, F., and Shi, L. (2018). A reduced-order successive linear estimator for geostatistical inversion and

- its application in hydraulic tomography. *Water Resources Research*, 54(3):1616–1632.
- Zhang, L., Bryant, S., Jennings, J., Arbogast, T., and Paruchuri, R. (2004). Multiscale flow and transport in highly heterogeneous carbonates. In *SPE Annual Technical Conference and Exhibition*, pages 26 – 29, Houston, Texas.
- Zhang, L., Nair, N., Jennings, J., and Bryant, S. (2005). Models and methods for determining transport properties of touching-vug carbonates. In *SPE Annual Technical Conference and Exhibition*, Dallas, Texas.
- Zhao, Z. and Illman, W. A. (2017). On the importance of geological data for three-dimensional steady-state hydraulic tomography analysis at a highly heterogeneous aquifer-aquitard system. *Journal of Hydrology*, 544:640 – 657.
- Zhao, Z. and Illman, W. A. (2018). Three-dimensional imaging of aquifer and aquitard heterogeneity via transient hydraulic tomography at a highly heterogeneous field site. *Journal of Hydrology*, 559:392 – 410.
- Zhao, Z., Illman, W. A., and Berg, S. J. (2016). On the importance of geological data for hydraulic tomography analysis: Laboratory sandbox study. *Journal of Hydrology*, 542:156 – 171.
- Zhu, J. and Yeh, T.-C. J. (2005). Characterization of aquifer heterogeneity using transient hydraulic tomography. *Water Resources Research*, 41(7).