



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**AquiParameter: Base de datos abierta de
parámetros hidrogeológicos. Compilación y
análisis estadístico usando R**

TESIS

Que para obtener el título de
Ingeniero Geólogo

P R E S E N T A

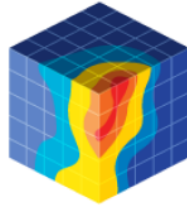
Héctor Eduardo Báez Reyes

DIRECTOR DE TESIS

Dr. José Antonio Hernández Espriú



Ciudad Universitaria, Cd. Mx., 2021



The
HYDRO
geology
Group
FACULTY OF ENGINEERING
U N A M



CONTRIBUCIÓN DEL GRUPO DE HIDROGEOLOGÍA DE LA FACULTAD DE INGENIERÍA, UNAM

Este documento representa la Tesis No. 27 DEL Grupo de Hidrogeología de la Facultad de Ingeniería (UNAM) como parte de su contribución científica y de formación de recursos humanos especializados en la evaluación cuantitativa de acuíferos.

This bachelor's dissertation represents the Thesis No. 27 OF THE Hydrogeology Group, Faculty of Engineering (UNAM) as part of our scientific contribution and human resources development in quantitative groundwater hydrology.

<http://www.ingenieria.unam.mx/hydrogeology/>

Twitter: @hydrogeologymx

Agradecimientos del proyecto

Agradezco a la Dirección General de Cómputo y de Tecnologías de Información y Comunicación, DGTIC, UNAM por proporcionarnos el servidor web para alojar la página “AquiParameter” y darnos la oportunidad de crear un proyecto libre y de fácil acceso a través de internet.

A la Mtra. Rebeca Valenzuela Argüelles y al Act. Mario Alberto Hernández Mayorga les agradezco por todo su apoyo y tiempo para hacer posible que la base de datos AquiParameter esté alojada en la nube.

Agradezco a la comunidad “Open Source” de desarrolladores por todas las herramientas que hacen posible trabajar con software de código abierto.

A mi director de tesis, Dr. José Antonio Hernández Espriú, y a mis sinodales: Dr. Enrique Alejandro González Torres, Ing. Gabriel Salinas Calleros, Ing. Martín Carlos Vidal García y Dr. Josue Tago Pacheco, les agradezco por aconsejarme para hacer de este un proyecto un trabajo de calidad.

Agradecimientos personales

A mis padres que me han brindado todo lo que un hijo puede pedir; por apoyar y respetar mis decisiones y forjarme para ser un buen profesional. Además, de enseñarme todos los valores que regirán mi vida.

A mi director de tesis, Dr. José Antonio Hernández Espriú, por brindarme todo su apoyo, tiempo y enseñarme nuevas herramientas que me serán de gran ayuda en mi vida profesional. Agradezco su disposición y ayuda para impulsarme a explotar mis habilidades.

Agradezco la Facultad de Ingeniería y a la Universidad Nacional Autónoma de México por los conocimientos, enseñanzas y vivencias que me ha brindado y por darme la oportunidad de conocer personas fantásticas que han cambiado mi manera de ver la vida.

A mis profesores de la Facultad por darme los conocimientos y habilidades necesarias que necesito para ser un profesional respetable.

Dedicatoria

A mi familia, por todo el amor y los buenos deseos que me brindan cada día y por siempre estar conmigo y apoyar mis decisiones. Una familia de la que estoy muy orgulloso. Porque sin ellos no lo habría logrado.

A mi madre Andrea Reyes González, por siempre brindarme su invaluable amor y por inculcarme buenos hábitos y valores que me han llevado a cumplir importantes logros como este.

A mi padre Librado Báez Herrera, por su amor y enseñanzas que hoy me han llevado a completar uno de los pasos más importantes en mi carrera profesional. Además de enseñarme a superarme y enfrentar cualquier reto.

A mis hermanos, por ser un ejemplo de dedicación y superación y porque sé que siempre estarán allí para apoyarme.

A mis amigos, con lo que he pasado momentos memorables porque sé que son un parte impórtate de mi vida.

Resumen

Los parámetros hidrogeológicos son un conjunto de propiedades físicas que rigen el movimiento del agua subterránea. En el campo de la hidrogeología existen una infinidad de estudios y experimentos en donde se caracterizan propiedades hidrogeológicas de diversos materiales, publicados en diferentes libros y revistas. Una base de datos de parámetros hidrogeológicos es un recurso que brinda a la comunidad científica y académica, una herramienta de consulta.

En el presente trabajo se expone una base de datos de los principales parámetros hidrogeológicos para cada una de las litologías más representativas. La base de datos presentada es de libre acceso para la comunidad científica, profesional y académica. La base de datos, referencias y el código de programación, pueden consultarse en la siguiente dirección: aquiparameters.rua.unam.mx.

Existen muchas herramientas para el procesado de datos, no obstante, se optó por trabajar con el lenguaje de programación R, un lenguaje diseñado para la computación estadística. En este trabajo toda la estadística, así como todas las etapas de la ciencia de datos, se desarrollaron con esta herramienta.

La base de datos, así como la estadística pueden consultarse en la página de internet. Por su parte en el presente documento se muestran los códigos en R, así como los pasos detallados que se llevaron a cabo para la preparación, visualización y comunicación de datos para producir un producto final.

Abstract

Hydrogeological parameters are a set of physical properties that govern the movement of groundwater. In the field of hydrogeology, there is an infinity of studies and experiments where hydrogeological properties of different materials are characterized, published in different books and magazines. A database of hydrogeological parameters is a resource that provides the scientific and academic community with a consultation tool.

In the present work, a database of the main hydrogeological parameters for each of the most representative lithologies is presented. The database present in this work is freely accessible to the scientific, professional, and academic community. The database, references and programming code can be consulted at the following address: aquiparameters.rua.unam.mx.

There are many tools for data processing, however, in this work, we chose to work with the R programming language, a language designed for statistical computing. In this work, all statistics, as well as all stages of data science, were developed with this tool.

Índice general

1	Introducción	20
1.1	Antecedentes	20
1.2	Justificación	21
1.3	Objetivos	22
1.3.1	Objetivo General	22
1.3.2	Objetivos Específicos	23
2	Marco Teórico	24
2.1	Unidades hidrogeológicas	24
2.1.1	Parámetros Hidrogeológicos	25
2.2	Estadística y ciencia de datos	36
2.2.1	Conceptos Generales	36
2.2.2	Métodos Gráficos	37
2.2.3	Métodos Numéricos	41
2.2.4	Correlaciones	44
2.3	Introducción al lenguaje de programación R	45
2.4	Bases de datos	49
3	Métodos	50
3.1	Obtención de datos	50
3.2	Generación de la base de datos	51
3.3	Procesado de datos con R	52
3.3.1	Creación de un nuevo proyecto	53
3.3.2	Instalación de paqueterías y lectura de datos	53
3.3.3	Generación de gráficos con la paquetería ggplot	56
3.3.4	Generación de gráficos interactivos con la paquetería plotly	63
3.3.5	Generación de tablas con medidas numéricas	64
3.3.6	Generación de tabla interactiva con la paquetería DT	67
3.3.7	Creación de una aplicación con la paquetería shiny	67
4	Discusión y resultados	78
4.1	Base de datos: detalles y descripción	78
4.2	Aplicación de consulta	79

4.3	Análisis estadístico	83
4.3.1	Porosidad Total	84
4.3.2	Porosidad Efectiva	104
4.3.3	Conductividad Hidráulica	123
4.3.4	Coefficiente de almacenamiento	145
4.3.5	Rendimiento específico	163
4.3.6	Compresibilidad	182
5	Conclusiones	190
6	Referencias citadas en el texto	192

Índice de figuras

2.1	Modelo de un ambiente kárstico (Tomado de Gilli et al. 2012).	25
2.2	Ejemplos de rocas fracturadas. A) Rocas Metamórficas (meta argilitas) en “Khetri Cooper Belt”, India, con diversas familias de fracturas rellenas de minerales arcillosos. B) Areniscas de “Vindhyan Group”, India, cuyos planos de estratificación representan las principales discontinuidades (Tomado de Singhal y Grupta, 2010).	25
2.3	Tipos de porosidad: a) Roca sedimentaria bien clasificada con alta porosidad; b) Roca sedimentaria mal clasificada con baja porosidad; c) Roca sedimentaria bien clasificada cuyos componentes tienen porosidad por si mismas, por lo que todo el sistema tiene alta porosidad; d) Roca sedimentaria bien clasificada cuya porosidad disminuyó por procesos de sedimentación en los intersticios; e) Roca soluble con porosidad producto de la disolución; f) Roca cristalina cuya porosidad se debe al fracturamiento (Tomado de Hiscock y Bense, 2006).	27
2.4	Imagen ilustrativa de los experimentos de Henry Darcy (Obtenido de Leap, 2006).	29
2.5	Representación esquemática del coeficiente de almacenamiento de un acuífero confinado (Obtenido de LEap, 2006).	32
2.6	Representación de un medio poroso donde se presenta el esfuerzo total, esfuerzo efectivo y presión de fluido en un plano arbitrario (Obtenido de Freeze y Cherry, 1979).	33
2.7	Esquema de la compactación de un acuífero producto del bombeo de agua subterránea (Obtenido de Freeze y Cherry, 1979).	35
2.8	Histogramas suavizados: a) unimodal simétrico; b) bimodal; c) con sesgo positivo; d) con sesgo negativo.	39
2.9	Distintas distribuciones con la misma media (Obtenido de Mendenhall et a. 1994).	43
2.10	Procesado de datos (Obtenido de Wickham y Grolemond, 2016).	47
2.11	Entorno de trabajo del lenguaje R con su interfaz en Rstudio.	48
2.12	Procesado de una base de datos (Obtenido de Özsü y Valdúriez, 1999).	49

3.1	Nuevo proyecto en Rstudio.	53
3.2	Script en Rstudio	55
3.3	Diagrama de rangos de la conductividad hidráulica (m/d) con ggplot2.	57
3.4	Diagramas de caja de la porosidad total (%) con ggplot2.	58
3.5	Diagramas de caja de la conductividad hidráulica (m/d) en escala logarítmica con ggplot2.	59
3.6	Histograma del coeficiente de almacenamiento (-) con ggplot2.	60
3.7	Histogramas del coeficiente de almacenamiento (-) con ggplot2.	61
3.8	Diagrama de correlación de la conductividad hidráulica (m/d) vs Porosidad Total (%) con ggplot2.	62
3.9	Diagramas de caja interactivos de la conductividad hidráulica.	64
3.10	Interfaz interactiva de la tabla "database" con la paquetería DT.	67
4.1	Página principal de la plataforma web (Apartado "Data")	81
4.2	Apartado "Boxplot" de la plataforma	82
4.3	Apartado "Histogram" de la plataforma	82
4.4	Apartado "Correlation" de la plataforma	83
4.5	Apartado "Download" de la plataforma	83
4.6	Apartado "About" de la plataforma	84
4.7	Diagrama de rangos de la porosidad total (%) para los diferentes tipos de litología.	85
4.8	Diagrama de rangos de la porosidad total (%) para diferentes litologías.	85
4.9	Diagramas de caja de la porosidad total (%) para las diferentes clasificaciones litológicas.	87
4.10	Histogramas de la porosidad total (%) para las diferentes clasificaciones litológicas. Histogramas contruidos a partir de 15 clases	89
4.11	Diagramas de correlación de la porosidad total en función de las diferentes propiedades hidrogeológicas. A: Porosidad total vs Porosidad Efectiva. B: Porosidad Total vs Conductividad Hidráulica (transformación logarítmica en base 10 del eje x). C: Porosidad Total vs Coeficiente de Almacenamiento (transformación logarítmica en base 10 del eje x). D: Porosidad Total vs Rendimiento Específico (transformación logarítmica en base 10 del eje x). E: Porosidad Total vs Compresibilidad (transformación logarítmica en base 10 del eje x)	90
4.12	Diagramas de caja de la porosidad total (%) en diferentes rocas ígneas.	91
4.13	Histogramas de la porosidad total (%) en diferentes rocas ígneas. Histogramas contruidos a partir de 10 clases	93
4.14	Diagramas de caja de la porosidad total (%) en diferentes rocas metamórficas.	94
4.15	Histogramas de la porosidad total (%) en diferentes rocas metamórficas. Histogramas contruidos a partir de 10 clases	97

4.16 Diagramas de caja de la porosidad total (%) en diferentes rocas sedimentarias.	98
4.17 Histogramas de la porosidad total (%) en diferentes rocas sedimentarias. Histogramas contruidos a partir de 10 clases . . .	100
4.18 Diagramas de caja de la porosidad total (%) en diferentes sedimentos no consolidados.	101
4.19 Histogramas de la porosidad total (%) en diferentes sedimentos no consolidados. Histogramas contruidos a partir de 10 clases . . .	104
4.20 Diagrama de rangos de la porosidad efectiva (%) para los diferentes tipos de litología.	105
4.21 Diagrama de rangos de la porosidad efectiva (%) para diferentes litologías.	106
4.22 Diagramas de caja de la porosidad efectiva (%) para las diferentes clasificaciones litológicas.	107
4.23 Histogramas de la porosidad efectiva (%) para las diferentes clasificaciones litológicas. Histogramas contruidos a partir de 15 clases	109
4.24 Diagramas de correlación de la porosidad efectiva en función de las diferentes propiedades hidrogeológicas. A: Porosidad Efectiva vs Porosidad Total. B: Porosidad Efectiva vs Conductividad Hidráulica (transformación logarítmica en base 10 del eje x). C: Porosidad Efectiva vs Coeficiente de Almacenamiento (transformación logarítmica en base 10 del eje x). D: Porosidad Efectiva vs Rendimiento Específico (transformación logarítmica en base 10 del eje x). E: Porosidad Efectiva vs Compresibilidad (transformación logarítmica en base 10 del eje x)	110
4.25 Diagramas de caja de la porosidad efectiva (%) en diferentes rocas ígneas.	111
4.26 Histogramas de la porosidad efectiva (%) en diferentes rocas ígneas. Histogramas contruidos a partir de 10 clases	113
4.27 Diagramas de caja de la porosidad efectiva (%) en diferentes rocas metamórficas.	114
4.28 Histogramas de la porosidad efectiva (%) en diferentes rocas metamórficas. Histogramas contruidos a partir de 10 clases . . .	116
4.29 Diagramas de caja de la porosidad efectiva (%) en diferentes rocas sedimentarias.	117
4.30 Histogramas de la porosidad efectiva (%) en diferentes rocas sedimentarias. Histogramas contruidos a partir de 10 clases . . .	120
4.31 Diagramas de caja de la porosidad efectiva (%) en diferentes sedimentos no consolidados.	121
4.32 Histogramas de la porosidad efectiva (%) en diferentes sedimentos no consolidados. Histogramas contruidos a partir de 10 clases . . .	123
4.33 Diagrama de rangos en escala logarítmica en base 10 de la conductividad hidráulica (m/d) para los diferentes tipos de litología.	124

4.34	Diagrama de rangos en escala logarítmica en base 10 de la conductividad hidráulica (m/d) para diferentes litologías.	125
4.35	Diagramas de caja en escala logarítmica en base 10 de la conductividad hidráulica (m/d) para las diferentes clasificaciones litológicas.	126
4.36	Histogramas en escala logarítmica en base 10 de la conductividad hidráulica (m/d) para las diferentes clasificaciones litológicas. Histogramas construidos a partir de 30 clases	129
4.37	Diagramas de correlación de la conductividad hidráulica en función de las diferentes propiedades hidrogeológicas. A: Conductividad hidráulica vs Porosidad Total (transformación logarítmica en base 10 del eje y). B: Conductividad Hidráulica vs Porosidad Efectiva (transformación logarítmica en base 10 del eje y). C: Conductividad hidráulica vs Coeficiente de Almacenamiento (transformación logarítmica en base 10 del eje y, y el eje x). D: Conductividad hidráulica vs Rendimiento Específico (transformación logarítmica en base 10 del eje y y el eje x); E. Conductividad hidráulica vs Compresibilidad (transformación logarítmica en base 10 del eje y y eje x)	130
4.38	Diagramas de caja en escala logarítmica en base 10 de la conductividad hidráulica (m/d) de diferentes rocas ígneas.	131
4.39	Histogramas en escala logarítmica en base 10 de la conductividad hidráulica (m/d) en diferentes rocas ígneas. Histogramas construidos a partir de 20 clases	134
4.40	Diagramas de caja en escala logarítmica en base 10 de la conductividad hidráulica (m/d) de diferentes rocas metamórficas. 135	
4.41	Histogramas en escala logarítmica en base 10 de la conductividad hidráulica (m/d) en diferentes rocas metamórficas. Histogramas construidos a partir de 20 clases	137
4.42	Diagramas de caja en escala logarítmica en base 10 de la conductividad hidráulica (m/d) de diferentes rocas sedimentarias. 138	
4.43	Histogramas en escala logarítmica en base 10 de la conductividad hidráulica (m/d) en diferentes rocas sedimentarias. Histogramas construidos a partir de 20 clases	141
4.44	Diagramas de caja en escala logarítmica en base 10 de la conductividad hidráulica (m/d) de diferentes sedimentos no consolidados.	142
4.45	Histogramas en escala logarítmica en base 10 de la conductividad hidráulica (m/d) en diferentes sedimentos no consolidados. Histogramas construidos a partir de 20 clases	145
4.46	Diagrama de rangos en escala logarítmica en base 10 del coeficiente de almacenamiento (-) para los diferentes tipos de litología.	146
4.47	Diagrama de rangos en escala logarítmica en base 10 del coeficiente de almacenamiento (-) para diferentes litologías.	147

4.48	Diagramas de caja en escala logarítmica del coeficiente de almacenamiento (-) para las diferentes clasificaciones litológicas. .	148
4.49	Histogramas en escala logarítmica en base 10 del coeficiente de almacenamiento (-) para las diferentes clasificaciones litológicas. Histogramas contruidos a partir de 30 clases	150
4.50	Diagramas de correlación del coeficiente de almacenamiento en función de las diferentes propiedades hidrogeológicas. A: Coeficiente de Almacenamiento vs Porosidad Total (transformación logarítmica en base 10 del eje y). B: Coeficiente de Almacenamiento vs Porosidad Efectiva (transformación logarítmica en base 10 del eje y). C: Coeficiente de Almacenamiento vs Conductividad Hidráulica (transformación logarítmica en base 10 del eje y y el eje x). D: Coeficiente de Almacenamiento vs Rendimiento Específico (transformación logarítmica en base 10 del eje y, y el eje x). E: Coeficiente de Almacenamiento vs Compresibilidad (transformación logarítmica en base 10 del eje y y eje x)	151
4.51	Diagramas de caja en escala logarítmica en base 10 del coeficiente de almacenamiento (-) de diferentes rocas ígneas.	152
4.52	Histogramas en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en rocas ígneas. Histogramas contruidos a partir de 20 clases	154
4.53	Diagramas de caja en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en diferentes rocas metamórficas.	155
4.54	Histogramas en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en rocas metamórficas. Histogramas contruidos a partir de 20 clases	157
4.55	Diagramas de caja en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en diferentes rocas sedimentarias.	158
4.56	Histogramas en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en rocas sedimentarias. Histogramas contruidos a partir de 20 clases	160
4.57	Diagramas de caja en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en diferentes sedimentos no consolidados.	161
4.58	Histogramas en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en sedimentos no consolidados. Histogramas contruidos a partir de 15 clases	164
4.59	Diagrama de rangos en escala logarítmica en base 10 del rendimiento específico (-) para los diferentes tipos de litología.	165
4.60	Diagrama de rangos en escala logarítmica en base 10 del rendimiento específico (-) para diferentes litologías.	166
4.61	Diagramas de caja en escala logarítmica en base 10 del rendimiento específico para las diferentes clasificaciones litológicas.	167
4.62	Histogramas en escala logarítmica en base 10 del rendimiento específico (-) para las diferentes clasificaciones litológicas. Histogramas contruidos a partir de 20 clases	169

- 4.63 Diagramas de correlación del rendimiento específico en función de las diferentes propiedades hidrogeológicas. A: Rendimiento Específico vs Porosidad Total (transformación logarítmica en base 10 del eje y); B: Rendimiento específico vs Porosidad Efectiva (transformación logarítmica en base 10 del eje y); C: Rendimiento Específico vs Conductividad Hidráulica (transformación logarítmica en base 10 del eje y, y el eje x); D: Rendimiento Específico vs Coeficiente de Almacenamiento (transformación logarítmica en base 10 del eje y, y el eje x); E: Rendimiento específico vs Compresibilidad (transformación logarítmica en base 10 del eje y, y eje x) 170
- 4.64 Diagramas de caja en escala logarítmica en base 10 del rendimiento específico (-) de diferentes rocas ígneas. 171
- 4.65 Histogramas en escala logarítmica en base 10 del rendimiento específico (-) en diferentes rocas ígneas. Histogramas construidos a partir de 20 clases 173
- 4.66 Diagramas de caja en escala logarítmica en base 10 del rendimiento específico (-) de diferentes rocas metamórficas. . . . 174
- 4.67 Histogramas en escala logarítmica en base 10 del rendimiento específico (-) en diferentes rocas metamórficas. Histogramas construidos a partir de 20 clases 176
- 4.68 Diagramas de caja en escala logarítmica en base 10 del rendimiento específico (-) de diferentes rocas sedimentarias. . . . 177
- 4.69 Histogramas en escala logarítmica en base 10 del rendimiento específico (-) en diferentes rocas sedimentarias. Histogramas construidos a partir de 20 clases 179
- 4.70 Diagramas de caja en escala logarítmica en base 10 del rendimiento específico (-) de diferentes sedimentos no consolidados. 180
- 4.71 Histogramas en escala logarítmica en base 10 del rendimiento específico (-) en diferentes sedimentos no consolidados. Histogramas construidos a partir de 20 clases 183
- 4.72 Diagrama de rangos en escala logarítmica en base 10 de la compresibilidad (m^2/N) para los diferentes tipos de litología 184
- 4.73 Diagramas de caja en escala logarítmica en base 10 de la compresibilidad (m^2/N) para las diferentes clasificaciones litológicas. 185
- 4.74 Histogramas en escala logarítmica en base 10 de la compresibilidad (m^2/N) para las diferentes clasificaciones litológicas. Histogramas construidos a partir de 15 clases 187

- 4.75 Diagramas de correlación de la compresibilidad en función de las diferentes propiedades hidrogeológicas. A: Compresibilidad vs Porosidad Total (transformación logarítmica en base 10 del eje y). B: Compresibilidad vs Porosidad Efectiva (transformación logarítmica en base 10 del eje y). C: Compresibilidad vs Conductividad Hidráulica (transformación logarítmica en base 10 del eje y, y el eje x). D: Compresibilidad vs Coeficiente de Almacenamiento (transformación logarítmica en base 10 del eje y, y el eje x). E: Compresibilidad vs Rendimiento Específico (transformación logarítmica en base 10 del eje y, y eje x) 188

Índice de tablas

2.1	Tipos de datos en lenguaje de programación R	47
3.1	Encabezado de la base de datos	52
4.1	Número de observaciones por propiedad hidrogeológica para cada tipo de litología	79
4.2	Número de observaciones por propiedad hidrogeológica para cada litología	80
4.3	Valores mínimos, máximos y cuartiles de la porosidad total (%)	86
4.4	Comparación de las medidas de tendencia central de la porosidad total (%)	87
4.5	Medidas de dispersión de la porosidad total (%)	88
4.6	Valores mínimos, máximos y cuartiles de la porosidad total (%)	92
4.7	Comparación de las medidas de tendencia central de la porosidad total (%)	92
4.8	Medidas de dispersión de la porosidad total (%)	93
4.9	Valores mínimos, máximos y cuartiles de la porosidad total (%)	95
4.10	Comparación de las medidas de tendencia central de la porosidad total (%)	95
4.11	Medidas de dispersión de la porosidad total (%)	96
4.12	Valores mínimos, máximos y cuartiles de la porosidad total (%)	97
4.13	Comparación de las medidas de tendencia central de la porosidad total (%)	99
4.14	Medidas de dispersión de la porosidad total (%)	99
4.15	Valores mínimos, máximos y cuartiles de la porosidad total (%)	102
4.16	Comparación de las medidas de tendencia central de la porosidad total (%)	102
4.17	Medidas de dispersión de la porosidad total (%)	103
4.18	Valores mínimos, máximos y cuartiles de la porosidad efectiva (%)	106
4.19	Comparación de las medidas de tendencia central de la porosidad efectiva (%)	107
4.20	Medidas de dispersión de la porosidad efectiva (%)	108
4.21	Valores mínimos, máximos y cuartiles de la porosidad efectiva (%)	111

4.22 Comparación de las medidas de tendencia central de la porosidad efectiva (%)	112
4.23 Medidas de dispersión de la porosidad efectiva (%)	113
4.24 Valores mínimos, máximos y cuartiles de la porosidad efectiva (%)	115
4.25 Comparación de las medidas de tendencia central de la porosidad efectiva (%)	115
4.26 Medidas de dispersión de la porosidad efectiva (%)	116
4.27 Valores mínimos, máximos y cuartiles de la porosidad efectiva (%)	118
4.28 Comparación de las medidas de tendencia central de la porosidad efectiva (%)	118
4.29 Medidas de dispersión de la porosidad efectiva (%)	119
4.30 Valores mínimos, máximos y cuartiles de la porosidad efectiva (%)	121
4.31 Comparación de las medidas de tendencia central de la porosidad efectiva (%)	122
4.32 Medidas de dispersión de la porosidad efectiva (%)	122
4.33 Valores mínimos, máximos y cuartiles de la conductividad hidráulica (m/d)	127
4.34 Comparación de las medidas de tendencia central de la conductividad hidráulica (m/d)	127
4.35 Medidas de dispersión de la conductividad hidráulica (m/d) . . .	128
4.36 Valores mínimos, máximos y cuartiles de la conductividad hidráulica (m/d)	132
4.37 Comparación de las medidas de tendencia central de la conductividad hidráulica (m/d)	132
4.38 Medidas de dispersión de la conductividad hidráulica (m/d) . . .	133
4.39 Valores mínimos, máximos y cuartiles de la conductividad hidráulica (m/d)	135
4.40 Comparación de las medidas de tendencia central de la conductividad hidráulica (m/d)	136
4.41 Medidas de dispersión de la conductividad hidráulica (m/d) . . .	137
4.42 Valores mínimos, máximos y cuartiles de la conductividad hidráulica (m/d)	139
4.43 Comparación de las medidas de tendencia central de la conductividad hidráulica (m/d)	140
4.44 Medidas de dispersión de la conductividad hidráulica (m/d) . . .	140
4.45 Valores mínimos, máximos y cuartiles de la conductividad hidráulica (m/d)	143
4.46 Comparación de las medidas de tendencia central de la conductividad hidráulica (m/d)	143
4.47 Medidas de dispersión de la conductividad hidráulica (m/d) . . .	144
4.48 Valores mínimos, máximos y cuartiles del coeficiente de almacenamiento (-)	147
4.49 Comparación de las medidas de tendencia central del coeficiente de almacenamiento (-)	148
4.50 Medidas de dispersión del coeficiente de almacenamiento (-) . . .	149

4.51	Valores mínimos, máximos y cuartiles del coeficiente de almacenamiento (-)	152
4.52	Comparación de las medidas de tendencia central del coeficiente de almacenamiento (-)	153
4.53	Medidas de dispersión del coeficiente de almacenamiento (-)	153
4.54	Valores mínimos, máximos y cuartiles del coeficiente de almacenamiento (-)	155
4.55	Comparación de las medidas de tendencia central del coeficiente de almacenamiento (-)	156
4.56	Medidas de dispersión del coeficiente de almacenamiento (-)	156
4.57	Valores mínimos, máximos y cuartiles del coeficiente de almacenamiento (-)	158
4.58	Comparación de las medidas de tendencia central del coeficiente de almacenamiento (-)	159
4.59	Medidas de dispersión del coeficiente de almacenamiento (-)	159
4.60	Valores mínimos, máximos y cuartiles del coeficiente de almacenamiento (-)	162
4.61	Comparación de las medidas de tendencia central del coeficiente de almacenamiento (-)	162
4.62	Medidas de dispersión del coeficiente de almacenamiento (-)	163
4.63	Valores mínimos, máximos y cuartiles del rendimiento específico (-)	165
4.64	Comparación de las medidas de tendencia central del rendimiento específico (-)	167
4.65	Medidas de dispersión del rendimiento específico (-)	168
4.66	Valores mínimos, máximos y cuartiles del rendimiento específico (-)	171
4.67	Comparación de las medidas de tendencia central del rendimiento específico (-)	172
4.68	Medidas de dispersión del rendimiento específico (-)	172
4.69	Valores mínimos, máximos y cuartiles del rendimiento específico (-)	174
4.70	Comparación de las medidas de tendencia central del rendimiento específico (-)	175
4.71	Medidas de dispersión del rendimiento específico (-)	175
4.72	Valores mínimos, máximos y cuartiles del rendimiento específico (-)	177
4.73	Comparación de las medidas de tendencia central del rendimiento específico (-)	178
4.74	Medidas de dispersión del rendimiento específico (-)	178
4.75	Valores mínimos, máximos y cuartiles del rendimiento específico (-)	181
4.76	Comparación de las medidas de tendencia central del rendimiento específico (-)	181
4.77	Medidas de dispersión del rendimiento específico (-)	182
4.78	Valores mínimos, máximos y cuartiles de la compresibilidad (\bar{m}/N)	184

4.79 Comparación de las medidas de tendencia central de la compresibilidad (\bar{n}_i/N)	185
4.80 Medidas de dispersión de la compresibilidad (\bar{n}_i/N)	186

Capítulo 1

Introducción

1.1 Antecedentes

El avance de las nuevas tecnologías hace posible una fácil captura, almacenamiento, proceso y distribución de la información. Esta información es de importancia crucial para la producción científica, profesional y académica. Por esta razón surge la necesidad de crear herramientas para almacenar, localizar, seleccionar, coleccionar y procesar información (Rivera, 1994).

Desde los primeros experimentos de Henry Darcy publicados en 1856, se han llevado a cabo una infinidad de estudios y experimentos con el fin de caracterizar propiedades hidráulicas de materiales geológicos. Todos estos experimentos y sus interpretaciones han sido publicados en: libros, revistas, bases de datos e informes; en diferentes campos de investigación como: la hidrogeología, ingeniería petrolera e ingeniería civil (Comunian and Renard, 2009).

Algunas agencias gubernamentales y otros organismos a cargo del suministro de agua cuentan con información general del agua subterránea y superficial dentro de sus informes; así como catálogos de información. No obstante, estas bases de datos se limitan a datos climatológicos o de escorrentía, por lo que surge la necesidad de desarrollar un sistema de gestión de bases de datos con el fin de mejorar el manejo de agua subterránea (Carrera-Hernández and Gaskin, 2007).

Algunos Centros Geológicos Nacionales cuentan con sus propias bases de datos hidrogeológicos, sin embargo, no están liberados al público en general. Algunos de estos recursos se obtienen mediante peticiones (Allen et al., 1997; Jones et al., 2000), o mediante páginas Web como: “Canadian geospatial data infrastructure” (CGDI, 2007); “ADES groundwater national portal in France” (ADES, 2008); o el “Ground Water Atlas of the United States by the United States Geological Survey” (USGS, 2008), son algunos otros ejemplos de bases de datos (Comunian

and Renard, 2009).

La mayoría de las bases de datos hidrogeológicas se inclinan a recolectar información centralizada, como los trabajos de Newell et al. (1990), donde se desarrolla una base de datos en colaboración con los miembros de la “National Water Well Association (NWWA)”, donde se presenta la información hidrogeológica de 400 campos de trabajo a lo largo de todo el país de Estados Unidos.

Otro ejemplo son los trabajos de Wösten et al. (1999), con un proyecto de recolección de datos hidráulicos de suelos para generar un catálogo central con el fin de resolver la falta de datos que limita la simulación de agua en el suelo; los datos están disponibles en instituciones y servicios de Europa. En este proyecto participaron un total de 20 instituciones de 12 países europeos, establecimiento la base de datos de las propiedades hídras de los suelos europeos “HYPRES”.

Además, existen iniciativas para administrar conjuntos de datos específicos en un formato estandarizado para garantizar su disponibilidad a largo plazo y facilitar la colaboración entre varios equipos de investigación. Un ejemplo de una base de datos orientada geográficamente que provee una interfaz entre los experimentadores de campo y los modeladores se expone en los trabajos de (De Dreuzy et al., 2006).

La recopilación de datos hidrogeológicos se obtiene a partir de estudios y análisis de núcleos y pruebas de bombeo. Las principales propiedades que suelen registrarse en bases de datos de estudios hidrogeológicos corresponden con: porosidad, conductividad hidráulica, transmisividad, almacenamiento. De acuerdo con Jones et al. (2000), estas son las propiedades fundamentales de un acuífero, en un contexto hidrogeológico; ya que suelen ser invariantes en el tiempo.

Los trabajos de Comunian and Renard (2009) son un ejemplo documentado de una base de datos de parámetros hidrogeológicos de todo el mundo; como un proyecto colaborativo disponible vía internet y el cual está respaldado por una comunidad amplia la cual tiene acceso libre a la información, así como la libertad de proveer información.

El principal objetivo de una base de datos de parámetros hidrogeológicos de todo el mundo Comunian and Renard (2009) es ayudar a la comunidad científica, así como a los profesionales hidrólogos e ingenieros, con un catálogo de valores y distribuciones estadísticas de las diferentes propiedades físicas de rocas para ofrecer un complemento de información bien documentado.

1.2 Justificación

La incertidumbre sobre las condiciones y variables en la ingeniería geológica es una característica peculiar en comparación con otras ingenierías. Por lo que para desarrollar un análisis racional se involucra la teoría de la probabilidad y

estadística, para propiciar una condición sistemática en la toma de decisiones (Einstein and Baecher, 1983).

En la mayoría de los casos dentro de un estudio hidrogeológico solo se muestrea directamente, menos del 1% del volumen total de un acuífero y en el mejor de los casos de un 20 a un 40% (Anderson et al., 1999). Los profesionales de la hidrogeología hacen frente a esta incertidumbre mediante el uso de técnicas estocásticas con el fin de gestionar proyectos hidrogeológicos (Freeze et al., 1990).

Sin embargo, las técnicas probabilísticas siguen teniendo limitaciones por la falta de datos, por lo que algunos autores como: Dagan (2002) y de De Marsily et al. (2005) sugieren un compendio de diferentes propiedades de las diversos componentes presentes en los medios geológicos.

En un proyecto de agua subterránea se realizan predicciones utilizando un puñado de datos disponibles para parametrizar modelos computacionales, los cuales tienen muchos grados de libertad y presentan muchos órdenes de magnitud. Esto a menudo genera incertidumbre en los resultados. Las herramientas computacionales pueden tratar con el problema de la falta de datos más no resolverlos (Tartakovsky and Larrabee, 2008).

El problema continúa siendo la falta de información, por lo que contar con una base de datos hidrogeológicos es fundamental; por una parte, proporciona un registro o una copia segura de los datos, así como un acceso rápido y controlado; además de poder llevar un monitoreo de datos de agua subterránea. Las bases de datos de agua subterránea actuales son poco aptas cuando se tiene como objetivo la caracterización de acuíferos (De Dreuzy et al., 2006).

Actualmente se cuenta con una extensa cantidad de información sobre hidráulica de acuíferos (porosidad total, porosidad efectiva conductividad hidráulica, transmisividad, coeficiente de almacenamiento, rendimiento específico, compresibilidad) para diferentes litologías; publicada en libros, artículos y reportes. De manera que un compendio se vuelve una tarea necesaria.

En general no se cuenta con una base de datos hidráulicos accesible y completa, con datos suficientes para un análisis estadístico, de donde se pueda obtener información de interés como: valores representativos, promedios, valores extremos y valores anómalos de las litologías más representativas.

1.3 Objetivos

1.3.1 Objetivo General

- Generar una base de datos de parámetros hidrogeológicos, para su posterior análisis con herramientas de computación estadística como Rstudio, que incluya las siguientes variables: Porosidad total, porosidad efectiva, conductividad hidráulica, coeficiente de almacenamiento,

rendimiento específico y compresibilidad; teniendo en cuenta las litologías más representativas dentro de cada grupo de rocas: Ígneas, Sedimentarias y Metamórficas, así como sedimentos no consolidados.

1.3.2 Objetivos Específicos

- Recopilar información geológica e hidrogeológica mediante la revisión detallada de bibliografía.
- Obtener datos de parámetros hidrogeológicos a partir del análisis y depuración de información.
- Generar una base de datos de parámetros hidráulicos por litología.
- Realizar un análisis estadístico de la base de datos con el fin de obtener información de interés, aplicable a estudios hidrogeológicos, mediante herramientas computacionales como Rstudio.
- Generar información y material didáctico de interés hidrogeológico como: diagramas y tablas con medidas numéricas.
- Publicar la base de datos y los resultados mediante una plataforma interactiva y de fácil acceso.

Capítulo 2

Marco Teórico

2.1 Unidades hidrogeológicas

La presencia de agua subterránea dentro de la corteza terrestre y la aparición de manantiales en la superficie de la tierra están determinados por la litología de los diversos materiales geológicos, las estructuras geológicas regionales, la geomorfología de los accidentes geográficos y la disponibilidad de fuentes de recarga (Hiscock, 2009).

El empaquetamiento y la clasificación de los granos influyen en la porosidad. El material clástico bien clasificado tiene una alta porosidad independientemente del tamaño de grano. En el material mal clasificado, la porosidad es menor ya que los granos de pequeño tamaño ocupan espacios porosos entre los granos más grandes. Además, la compactación y la cementación reducen la porosidad (Singhal and Gupta, 2010).

El agua pura es capaz de disolver minerales como la halita y el yeso. De la misma forma el agua que contiene dióxido de carbono (CO_2) es capaz de disolver calcita, aragonita y dolomita (Gilli et al., 2012). Por lo que *acuíferos kársticos* (figura 2.1) se presentan en yeso, calizas y dolomitas generalmente (White, 2016).

En general todos los tipos de rocas tienen discontinuidades, como ejemplo las fracturas, las cuales facilitan el almacenamiento de fluidos y su movimiento; sin embargo, discontinuidades, como las fallas y los diques, también pueden actuar como fronteras para el flujo de agua (figura 2.2). El entendimiento de las propiedades hidrogeológicas en medios fracturados como: porosidad, permeabilidad, y en general el movimiento de agua subterránea es poco conocido (Singhal and Gupta, 2010).

Las rocas fracturadas se componen de bloques (matriz) y fracturas. Los bloques tienen baja permeabilidad y alta capacidad de almacenamiento, mientras que las fracturas tienen alta permeabilidad y baja capacidad de almacenamiento. A

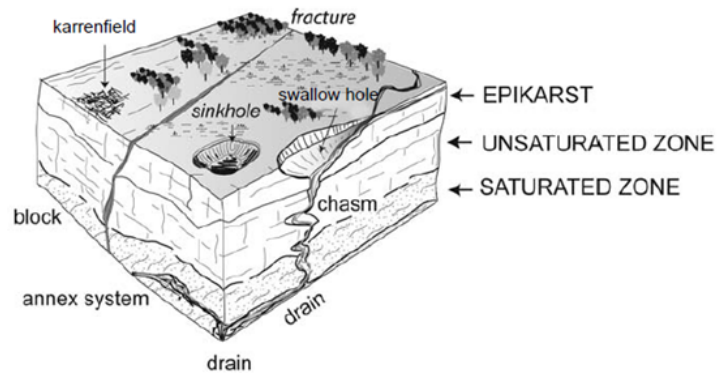


Figura 2.1: Modelo de un ambiente kárstico (Tomado de Gilli et al. 2012).

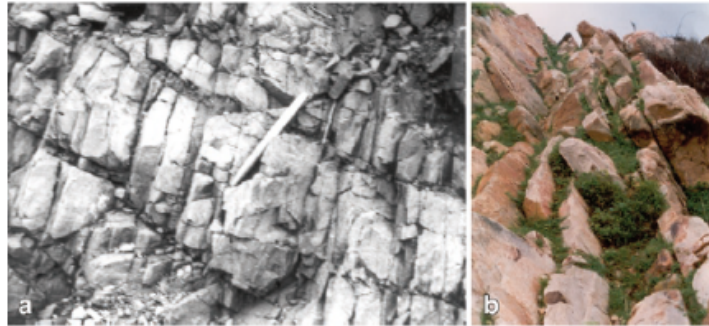


Figura 2.2: Ejemplos de rocas fracturadas. A) Rocas Metamórficas (meta argilitas) en “Khetri Cooper Belt”, India, con diversas familias de fracturas rellenas de minerales arcillosos. B) Areniscas de “Vindhyan Group”, India, cuyos planos de estratificación representan las principales discontinuidades (Tomado de Singhal y Gupta, 2010).

este tipo de acuíferos se les denomina de doble porosidad (Singhal and Gupta, 2010).

2.1.1 Parámetros Hidrogeológicos

La infiltración de la lluvia en la capa freática y el flujo de agua subterránea en un acuífero hacia un área de descarga, se rigen por leyes físicas que describen los cambios en la energía y el movimiento del agua subterránea (Hiscock, 2009). Es así como las propiedades hidráulicas de las formaciones geológicas gobiernan sus características de almacenamiento y transmisión de agua en el subsuelo (Singhal and Gupta, 2010).

2.1.1.1 Porosidad

La porosidad es la fracción de un volumen de material que está ocupado por espacios vacíos. Esta porosidad representada por el símbolo n o ϕ se define como: “la relación del volumen de huecos y el volumen unitario total de un material”. Sus unidades son en porcentaje o fracción decimal y está expresada como (Hiscock y Bense, 2006):

$$\phi = \frac{V_v}{V_t}. \quad (2.1)$$

Donde:

ϕ = Porosidad

V_v = Volumen de huecos

V_t = Volumen unitario total

Singhal and Gupta (2010) definen que la porosidad de una roca o suelo está determinada en gran medida por:

- Forma y disposición de los granos constituyentes
- Grado de clasificación
- Compactación y cementación
- Fracturas
- Solución

Durante el proceso de génesis de una roca ya sea por precipitación o por enfriamiento de un magma; sedimentación de materiales sueltos; o durante la generación de un suelo; se adquiere cierta porosidad inherente conocida como *porosidad primaria*. Esta porosidad puede disminuir por procesos de precipitación, sedimentación o compactación posteriores a la diagénesis. Por su parte, procesos posteriores a la diagénesis como fracturación, disolución y compactación generan *porosidad secundaria* (Delleur L, 2006). Ejemplos de porosidad en diversos materiales geológicos se presentan en la figura 2.3.

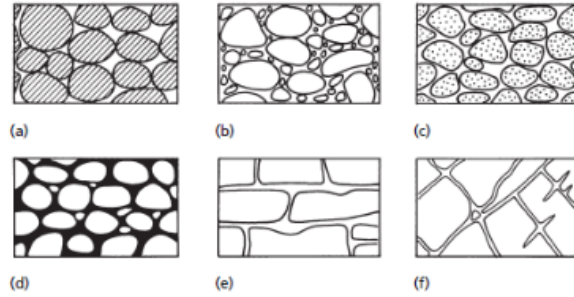


Figura 2.3: Tipos de porosidad: a) Roca sedimentaria bien clasificada con alta porosidad; b) Roca sedimentaria mal clasificada con baja porosidad; c) Roca sedimentaria bien clasificada cuyos componentes tienen porosidad por si mismas, por lo que todo el sistema tiene alta porosidad; d) Roca sedimentaria bien clasificada cuya porosidad disminuyó por procesos de sedimentación en los intersticios; e) Roca soluble con porosidad producto de la disolución; f) Roca cristalina cuya porosidad se debe al fracturamiento (Tomado de Hiscock y Bense, 2006).

De acuerdo con Delleur L (2006), la *Porosidad Total* (ϕ_T) es la suma de la porosidad primaria y la porosidad secundaria de un material. La *Porosidad Efectiva* (ϕ_{ef}) es la fracción de la porosidad total o fracción de poros que permite el paso de agua (Delleur L, 2006).

La porosidad total se relaciona con la capacidad de almacenamiento del material, mientras que la porosidad efectiva se relaciona con la capacidad de transmisión de un medio (Hiscock, 2009).

Darcy (1856) describe el flujo de agua en medios porosos a partir de la aplicación de los principios de flujo. La ley de Darcy se expresa como:

$$v = ks. \quad (2.2)$$

Donde:

v = La velocidad de flujo

s = La pendiente del gradiente hidráulico

k = Coeficiente de permeabilidad que tiene unidades de velocidad.

Suponiendo que el caudal es el producto del área de la sección transversal y la velocidad; y el área efectiva es el área total multiplicada por la porosidad del medio. Se obtiene:

$$q = kpAs = K_p As. \quad (2.3)$$

Donde:

q = Caudal

K_p = Coeficiente de permeabilidad

A = Sección transversal

p = porosidad

s = La pendiente del gradiente hidráulico

El coeficiente K_p se denomina coeficiente de permeabilidad o conductividad hidráulica. Depende de las propiedades del fluido y del medio poroso y se puede expresar como:

$$K_p = k \frac{w}{\mu} = C d^2 \frac{w}{\mu}. \quad (2.4)$$

Donde:

k = La permeabilidad intrínseca del medio

w = El peso específico del fluido

μ = La viscosidad absoluta

C = Factor que describe la forma, empaquetamiento, porosidad y otras características del medio

d = El tamaño promedio de los poros del material.

Por lo tanto, la permeabilidad intrínseca se expresa en darcys y tiene dimensiones de área. En conclusión, la *permeabilidad* intrínseca está en función del tamaño de las aberturas de los poros a través de las cuales atraviesa un fluido, se expresa como (Darcy, 1856):

$$k = C d^2 \quad (2.5)$$

2.1.1.2 Conductividad Hidráulica

La permeabilidad intrínseca es una propiedad que describe la transmisión de agua en un medio poroso, no obstante, no describe la facilidad con la que un líquido puede fluir a través de un medio. Por lo que surge el concepto de *Conductividad Hidráulica* (K), el cual combina propiedades del fluido y del medio. Se describe como (Delleur L, 2006):

$$K = \frac{k \rho g}{\mu}. \quad (2.6)$$

Donde:

k = Permeabilidad intrínseca

ρ = Densidad del fluido

μ = Viscosidad

g = gravedad

La Conductividad Hidráulica tiene dimensión de velocidad (cm/seg o ft/day). Su significado físico es el siguiente: “El volumen de un líquido fluyendo perpendicular a una unidad de área de un medio poroso por influencia de un gradiente hidráulico unitario” (Delleur L, 2006). Darcy (1856) experimentó con filtros de arena para un suministro de agua para la ciudad de Dijon, Francia, y determinó los factores que controlan el caudal de agua a través de la arena (figura 2.4). Proponiendo la ley de Darcy:

$$q = -K \left(\frac{dh}{dl} \right). \quad (2.7)$$

Donde:

K = La conductividad hidráulica

q = La descarga específica o el caudal volumétrico por unidad de área de un medio poroso, perpendicular a la dirección del flujo

dh/dl = El gradiente hidráulico a lo largo de la trayectoria del flujo. El signo negativo implica que el flujo está en la dirección del gradiente decreciente

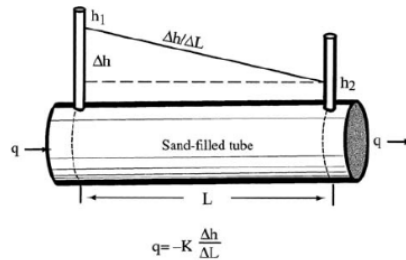


Figura 2.4: Imagen ilustrativa de los experimentos de Henry Darcy (Obtenido de Leap, 2006).

La conductividad hidráulica depende tanto de las propiedades del medio (material de roca) como del fluido (Singhal and Gupta, 2010).

2.1.1.3 Transmisividad

Suele utilizarse por conveniencia el término de transmisividad para describir la facilidad con que el agua se mueve a través de un medio poroso. Se expresa como el producto de la conductividad hidráulica y el espesor saturado de un acuífero (Delleur L, 2006):

$$T = Kb. \quad (2.8)$$

Tiene unidades de L^2/T y se define como: “el volumen de agua por unidad de tiempo que pasa a través de una unidad de área de un acuífero con un determinado espesor bajo un gradiente hidráulico” (Delleur L, 2006). La transmisividad generalmente se reporta en unidades de pies cuadrados por día o metros cuadrados por día.

La tasa total de flujo (Q) a través de cualquier área (A) del acuífero perpendicular a la dirección del flujo bajo el gradiente (I) se da como (Delleur L, 2006):

$$Q = TIA. \quad (2.9)$$

2.1.1.4 Almacenamiento Específico

Para un acuífero saturado, el almacenamiento *específico* (S_s) se define como el volumen de agua que una unidad de volumen de acuífero libera del almacenamiento bajo una unidad de disminución en la carga hidráulica. El agua que se libera del almacenamiento por una disminución de la carga hidráulica se produce por dos mecanismos: la compactación del acuífero causadas por el aumento del esfuerzo efectivo (σ_e), y por la expansión del agua causada por la disminución de presión del fluido en los poros (Freeze and RA, 1979).

Se considera que el volumen de agua producida de una unidad de volumen de un acuífero durante la compactación es igual a la reducción en el volumen de una unidad de volumen del acuífero por lo que:

$$dV_w = -dV_T = \alpha V_T d\sigma_e. \quad (2.10)$$

Donde:

V_w = Volumen de agua

V_T = Unidad de volumen

α = Compresibilidad

σ_e = Esfuerzo específico

Para una unidad de volumen $V_T = 1$ y considerando $d\sigma_e = -\rho g dh$, se tiene:

$$dV_w = \alpha \rho g. \quad (2.11)$$

Si el volumen de agua producida por la expansión de agua es:

$$dV_w = -\beta V_w dp. \quad (2.12)$$

Donde:

β = Compresión del agua

p = Presión del fluido

Ahora suponiendo que el volumen de agua (V_w) en el volumen total (V_T) es igual a nV_T donde n es la porosidad. Y considerando $V_T = 1$ y $dp = -\rho g dz = \rho g d(h-z) = \rho g dh$, con $dh=1$. Se tiene la ecuación del almacenamiento específico (Freeze and RA, 1979):

$$Ss = pg(\alpha + n\beta). \quad (2.13)$$

El *almacenamiento específico* (Ss) tiene unidades de $[L^{-1}]$.

Existe otro término conocido como *retención específica* (Sr) el cual se le asigna al volumen de agua que no es drenado de la roca (Fetter, 2000).

2.1.1.5 Coeficiente de Almacenamiento

El *coeficiente de almacenamiento* (S) de un acuífero confinado de espesor “b” se puede definir como el volumen de agua que un acuífero libera de su almacenamiento por unidad de superficie cuando el nivel potenciométrico disminuye (figura 2.5). Sus unidades son adimensionales (-).

Para un acuífero confinado de espesor “b” el coeficiente de almacenamiento está dado por (Freeze and RA, 1979):

$$S = Ssb. \quad (2.14)$$

Si $Ss = pg(\alpha + nb)$, la ecuación se puede expresar como:

$$S = \rho gb(\alpha + n\beta). \quad (2.15)$$

2.1.1.6 Rendimiento Específico

En acuíferos libres el concepto de almacenamiento se precisa como *rendimiento específico* (Sy) y se define como: “El volumen de agua que un acuífero no confinado libera del almacenamiento por unidad de superficie cuando el nivel freático desciende en una unidad”. En general el almacenamiento de un acuífero libre es mucho más alto que el de un acuífero confinado (Freeze and RA, 1979).

Los valores de Sy se expresan como una fracción decimal o como un porcentaje. Singhal and Gupta (2010) aclaran que el rendimiento específico depende de la duración del drenaje, la temperatura, la composición mineral del agua, el tamaño de grano y otras características de textura de los materiales del acuífero.

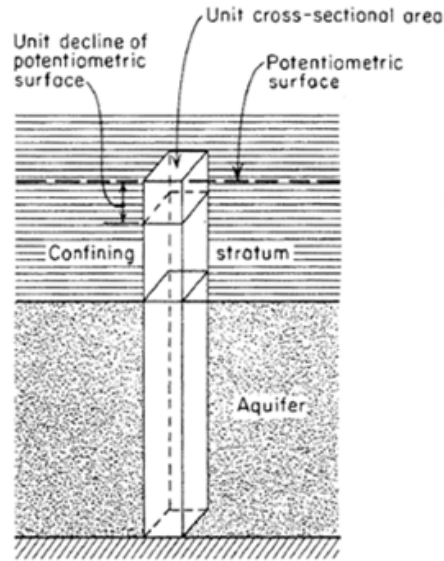


Figura 2.5: Representación esquemática del coeficiente de almacenamiento de un acuífero confinado (Obtenido de LEap, 2006).

Los valores altos de liberación del almacenamiento en un acuífero libre representan el drenado de los poros de un medio, mientras que la liberación del almacenamiento de acuíferos confinados representa solo los efectos secundarios de la expansión del agua y la compactación del acuífero, causados por cambios en la presión del fluido (Freeze and RA, 1979).

Es posible establecer que la porosidad efectiva y el rendimiento específico son equivalentes si se entiende que el volumen de agua que se extrae depende de la porosidad efectiva (Heath, 2004). Partiendo de la ecuación de porosidad total:

$$\phi_T = \phi_{ef} + \phi_{noef} = Sy + Sr. \quad (2.16)$$

Por lo que:

$$Sy = \phi_{ef}. \quad (2.17)$$

Donde:

Sy = Rendimiento específico

Sr = Retención específica

ϕ_T = Porosidad total

ϕ_{ef} = Porosidad efectiva

ϕ_{noef} = Porosidad no efectiva

2.1.1.7 Compresibilidad

Rocas y suelos que contienen agua dentro de sus poros se ven afectadas por esfuerzos externos como la presión del suelo o roca suprayacente y por esfuerzos internos como la presión del agua de los poros (Delleur L, 2006).

La compresibilidad es una propiedad que describe un cambio de volumen, en un material bajo un esfuerzo aplicado. Para la compresibilidad de un medio poroso existen tres mecanismos que llevan a la reducción de volumen, los cuales son: compresión del agua en los poros, compresión de los granos de arena individuales, y por un reacondicionamiento de los granos de arena (Freeze and RA, 1979).

El primero de estos mecanismos está controlado por la compresibilidad, el segundo se considera o tiende a ser insignificante (Freeze and RA, 1979). Para entender el tercero se considera el principio de estrés efectivo propuesto por Terzaghi (1925). Se simula un equilibrio de tensiones (figura 2.6) en un plano en una formación geológica saturada en profundidad (Freeze and RA, 1979).

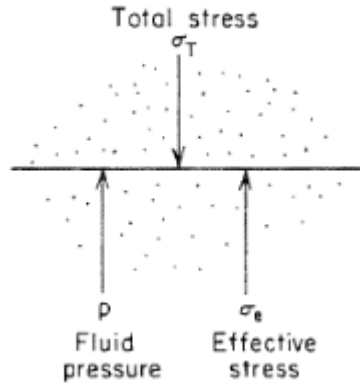


Figura 2.6: Representación de un medio poroso donde se presenta el esfuerzo total, esfuerzo efectivo y presión de fluido en un plano arbitrario (Obtenido de Freeze y Cherry, 1979).

Donde σ_T es el esfuerzo total que actúa hacia abajo en el plano por efecto del peso de las rocas y el agua superpuestas. Este estrés es soportado en parte por el esqueleto granular del medio poroso y en parte por la presión del fluido en los poros (p). La componente del esfuerzo total que no es soportada por el fluido se llama esfuerzo efectivo (σ_e) (Freeze and RA, 1979). La reorganización de los granos del suelo es causada por cambios en el esfuerzo efectivo, y no por cambios en el esfuerzo total. El equilibrio de esfuerzos se expresa como:

$$\sigma_T = p + \sigma_e. \quad (2.18)$$

En términos del gradiente se puede definir de la siguiente manera:

$$d\sigma_T = dp + d\sigma_e. \quad (2.19)$$

En general los problemas de agua subterránea no implican un cambio en el esfuerzo total, ya que el peso de roca y agua suprayacente permanece constante, por lo que:

$$dp = -d\sigma_e. \quad (2.20)$$

El esfuerzo efectivo en cualquier punto del sistema, así como las deformaciones volumétricas resultantes, están controlados por las presiones del fluido en ese punto. Suponiendo que $p = \rho g \psi$ y $\psi = h - z$ los cambios en el esfuerzo efectivo en un punto se rigen por los cambios en la carga hidráulica en ese punto (Freeze and RA, 1979):

$$d\sigma_e = -\rho g d\psi = \rho g dh. \quad (2.21)$$

Donde:

σ_e = El esfuerzo efectivo

ρ = Densidad

g = gravedad

ψ = Carga hidráulica

h = Altura hidráulica

La *compresibilidad* de un medio poroso está definida como (Freeze and RA, 1979):

$$\alpha = \frac{dV_T/V_T}{d\sigma_e}. \quad (2.22)$$

Donde:

V_T = Volumen total del medio

σ_e = Estrés efectivo.

El volumen total (V_T) se expresa como:

$$V_T = V_S + V_V. \quad (2.23)$$

Un aumento en el esfuerzo efectivo lleva a una reducción del volumen total que se traduce en un reacomodo de los materiales granulares ya que la compresión de granos individuales se considera insignificante. Para muchos casos prácticos los cambios en los esfuerzos de tensión tienen a ser muy pequeños en la mayoría de los casos que son insignificantes por lo que la compresibilidad del acuífero se toma como un único parámetro isotrópico (Freeze and RA, 1979).

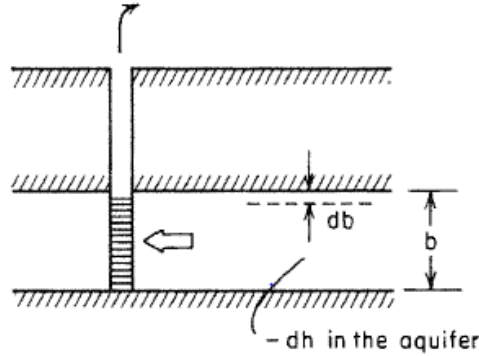


Figura 2.7: Esquema de la compactación de un acuífero producto del bombeo de agua subterránea (Obtenido de Freeze y Cherry, 1979).

Los materiales vecinos proporcionan un confinamiento horizontal al acuífero por lo que la compresibilidad es unidimensional. En la figura 2.7 se representa la naturaleza de la deformación que ocurre en un acuífero, en donde si el esfuerzo producto del material suprayacente permanece constante y la altura hidráulica en el acuífero disminuye, la ecuación de compresibilidad está dada por (Freeze and RA, 1979):

$$db = -abd\sigma_e = -ab\rho gdh. \quad (2.24)$$

El signo menos indica que el decremento de la carga hidráulica produce un decremento en el espesor

El bombeo induce gradientes hidráulicos horizontales hacia el pozo en el acuífero y, como resultado, la altura hidráulica disminuye en cada punto cerca del pozo. En respuesta, las tensiones efectivas aumentan en estos puntos, dando como resultado la compactación del acuífero (figura 2.7). Por el contrario, bombear agua a un acuífero aumenta la altura hidráulica, lo que disminuye las tensiones efectivas y provoca la expansión del acuífero. Si la compactación de un sistema acuífero-acuífero debido al bombeo de agua subterránea se propaga a la superficie del suelo, el resultado es el hundimiento de la tierra (Freeze and RA, 1979).

2.2 Estadística y ciencia de datos

2.2.1 Conceptos Generales

La *ciencia de datos* permite convertir datos crudos en entendimiento, comprensión y conocimiento de interés para una mejor toma de decisiones; entre sus herramientas se utilizan técnicas de ordenamiento, transformación, visualización y modelado de datos, así como la presentación de resultados (Wickham and Grolemund, 2016). La ciencia se relaciona con las matemáticas, la estadística y las tecnologías de la información,

Stuart et al. (1991) definen que: “Estadística es la rama del método científico que estudia los datos obtenidos por contar o medir las propiedades de poblaciones”. Por su parte Mendenhall et al. (1986) afirma que descripciones de diferentes autores para la definición de estadística coinciden en esta como: la recolección de información, seleccionando un subconjunto de un gran conjunto de datos, con el fin de inferir las características del conjunto completo.

Mediante la estadística se puede razonar de manera lógica, así como tomar decisiones en presencia de incertidumbre y variación. En todas las ciencias se trabaja con conjuntos de datos o hechos; y la estadística proporciona los métodos para organizar, resumir y obtener conclusiones (Devore, 2008).

Examinar todo un grupo de datos (población) puede ser imposible o poco práctico, es por eso que se examina sólo una pequeña parte del grupo denominado muestra (Spiegel and Stephens, 2009). Para (Devore, 2008) una *población* se define como una colección bien definida de objetos en una investigación. Mientras que un subconjunto de esta población se denomina *muestra*. Por lo regular, el interés se centra sólo en ciertas características o variables de los objetos de una población.

Una *variable* es cualquier elemento que puede tomar cualquiera de los valores de determinado conjunto al que se le conoce como dominio de la variable (Spiegel and Stephens, 2009). Devore (2008) define que una variable es *discreta* si su conjunto de valores posibles es finito o se puede enumerar en una sucesión infinita y una variable es *continua* si sus valores posibles consisten en un intervalo completo en la recta numérica.

Un conjunto de *datos univariados* consiste en observaciones de una sola variable, mientras que un conjunto de *datos multivariados* es cuando se hacen observaciones en más de dos variables (Devore, 2008).

El método más común para asegurar una selección representativa de datos es tomar una muestra aleatoria simple; una en la que cualquier subconjunto particular de tamaño específico tiene la misma probabilidad de ser elegida (Devore, 2008). Los datos en bruto que se recolectan en una investigación necesitan ser organizados, lo más común distribuirlos en clases o categorías para su mejor análisis (Spiegel y Stephens, 2009).

La representación tabular de los datos en clases con sus respectivas frecuencias se le conoce como distribución de frecuencias o tabla de frecuencias. A los datos organizados y resumidos en una tabla de frecuencias se les denomina datos agrupados (Spiegel and Stephens, 2009).

La estadística descriptiva permite resumir y explicar peculiaridades importantes de datos en una investigación mediante métodos gráficos y medidas numéricas (Devore, 2008). Algunos métodos de estadística descriptiva son los siguientes:

Métodos gráficos

- Histogramas
- Diagramas de Caja
- Correlaciones

Métodos Numéricos

- Medias de tendencia central
- Medidas de dispersión

Existe otra rama de la estadística denominada estadística inferencial la cual aplica técnicas para hacer una generalización en toda la población a partir de una muestra (Devore, 2008). Hacer una inferencia implica describir en forma parcial o completa un fenómeno o un objeto físico (Mendenhall et al., 1986).

2.2.2 Métodos Gráficos

2.2.2.1 Histogramas

Un histograma es una representación gráfica de la distribución de frecuencias y consiste en un conjunto de rectángulos cuyas bases se encuentran en un eje horizontal x , en donde sus centros coinciden con las marcas de clase y su base tiene una longitud igual al intervalo de clase. El área del rectángulo es proporcional a la frecuencia de la clase (Spiegel and Stephens, 2009).

Construir un histograma para datos continuos, requiere (Devore, 2008): subdividir el eje de las mediciones en una cantidad adecuada de intervalos de clase o clases, de modo que cada observación esté contenida en exactamente una clase. Añadir un dígito de centésimos a los límites de clase evita que las observaciones caigan en los límites resultantes. La forma de construir un histograma es el siguiente:

- Determine la frecuencia y la frecuencia relativa para cada clase.
- Marque los límites de clase en un eje de medición horizontal.
- Arriba de cada intervalo de clase, trace un rectángulo cuya altura sea la frecuencia relativa (o frecuencia) correspondiente

No hay reglas en relación con el número de clases o la elección de las clases. Entre 5 y 20 clases son satisfactorias para la mayor parte de conjuntos de datos. Una regla empírica razonable del número de clases está dado por (Devore, 2008):

$$\text{No. de clases} = \sqrt{\text{No. de observaciones}}. \quad (2.25)$$

El tamaño o la *amplitud* de un intervalo de clase es la diferencia entre sus fronteras superior e inferior y si todos los intervalos de clase tienen la misma amplitud, esta amplitud común se denota como “*c*” (Spiegel and Stephens, 2009).

Los *límites de clase* o fronteras de clase se obtienen sumando el límite superior de un intervalo de clase al límite inferior del intervalo de clase inmediato superior y dividiendo entre 2 (Spiegel and Stephens, 2009).

La *frecuencia* de cualquier valor particular de x es la cantidad de veces que se presenta ese valor en una clase en cuestión. La *frecuencia relativa* de un valor es la fracción o proporción de las veces que se presenta ese valor (Devore, 2008).

$$\text{Frecuencia relativa de un valor} = \frac{\text{No. de veces que aparece el valor}}{\text{No. de observaciones en el conjunto de datos}}. \quad (2.26)$$

Las *marcas de clase* es el punto medio de un intervalo de clase y se obtiene sumando los límites de clase inferior y superior y dividiendo entre 2 (Spiegel and Stephens, 2009).

Un *polígono de frecuencias* es una representación gráfica que simula las frecuencias de clase contra las marcas de clase. Se genera a partir de unir los puntos medios de las partes superiores de los rectángulos con una línea.

Medidas de forma Un *histograma unimodal* es el que sube hasta un solo pico y después baja; es decir que una sola clase es la que presenta mayor frecuencia. Un *histograma multimodal* tiene picos distintos; es decir que hay dos o más clases distintas que tienen altas frecuencias (figura 2.8). En cuanto a la simetría los histogramas se pueden clasificar en (Devore, 2008):

- *Simétrico* si la mitad izquierda es imagen especular de la mitad derecha
- *Sesgado positivo* si el extremo derecho o superior se extiende comparado con el extremo izquierdo o inferior
- *Sesgo negativo* si se extiende a la izquierda.

Momentos Los momentos se obtienen a partir de la medida de las variables y su frecuencia, con el fin de caracterizar las distribuciones de tal forma que, si los momentos coinciden en dos distribuciones, diremos que son iguales (Muñoz, 2000).

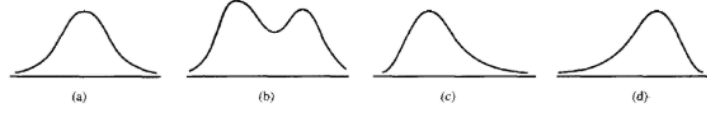


Figura 2.8: Histogramas suavizados: a) unimodal simétrico; b) bimodal; c) con sesgo positivo; d) con sesgo negativo.

Dados N valores X_1, X_2, \dots, X_N que toma la variable X , se define la cantidad [@(Spiegel and Stephens, 2009)]:

$$\overline{X^r} = \frac{X_1^r + X_2^r + \dots + X_N^r}{N} = \frac{\sum_{j=1}^N X_j^r}{N} = \frac{\sum X^r}{N}. \quad (2.27)$$

Está definido como r -ésimo momento. El primer momento, en el que $r = 1$ es la media aritmética X . El r -ésimo momento respecto a la media X se define como:

$$m_r = \frac{\sum_{j=1}^N (X_j - \overline{X})^r}{N} = \overline{(X - \overline{X})^r}. \quad (2.28)$$

Si $r = 1$, entonces $m_1 = 0$. Si $r = 2$, entonces m_2 es la varianza. El r -ésimo momento respecto a cualquier origen A se define de la manera siguiente:

$$m'_r = \frac{\sum_{j=1}^N (X_j - A)^r}{N} = \frac{\sum (X - A)^r}{N} = \frac{\sum d^r}{N} = \overline{(X - A)^r}. \quad (2.29)$$

Donde $d = X - A$ son las desviaciones de las X respecto de A . Si $A = 0$, la ecuación m'_r se reduce a la ecuación de $(\overline{X^r})$. Debido a esto, a la ecuación $(\overline{X^r})$ suele llamársele el r -ésimo momento respecto de cero (Spiegel and Stephens, 2009).

Sesgo El sesgo es un grado en el que una distribución se aleja de la simetría. Esta medida se obtiene de la diferencia de la mediana y la moda; dividida entre una medida de dispersión (Spiegel and Stephens, 2009):

$$\text{Sesgo} = \frac{\text{Mediana} - \text{Moda}}{\text{Desviación Estandar}} = \frac{\overline{X} - \text{Moda}}{s}. \quad (2.30)$$

El tercer momento respecto de la media puede emplearse para medir el sesgo, tal medida expresada en forma adimensional es:

$$\text{Coeficiente momento de sesgo} = a_3 = \frac{m_3}{s^3} = \frac{m_3}{(\sqrt{m_2})^3} = \frac{m_3}{\sqrt{m_2^3}}. \quad (2.31)$$

De manera gráfica si una curva de distribución (polígono de frecuencias suavizado) tiene una cola más larga hacia la derecha del máximo central que hacia la izquierda, se dice que la distribución es sesgada a la derecha, o que tiene un sesgo positivo. Si ocurre lo contrario, se dice que es sesgada a la izquierda o que tiene un sesgo negativo, ver fig. 2.8.

Curtosis La curtosis de manera gráfica indica qué tan puntiaguda es una distribución por lo que una distribución que tiene un pico relativamente alto se le llama leptocúrtica, en tanto que si es relativamente aplastada se dice platicúrtica; mientras que una distribución normal, que no es ni puntiaguda ni muy aplastada se llama mesocúrtica. Para medirla se necesita el cuarto momento respecto a la media (Spiegel and Stephens, 2009).

$$\text{Coeficiente momento de curtosis} = a_4 = \frac{m_4}{s^4} = \frac{m_4}{m_2^2}. \quad (2.32)$$

Cuartiles y Percentiles Un conjunto de datos ordenados de acuerdo a su magnitud puede dividirse en varias partes de tamaños iguales. “En términos generales, los cuartiles dividen el conjunto de datos en cuatro partes iguales, donde las observaciones arriba del tercer cuartil son el cuarto superior del conjunto de datos, el segundo cuartil es idéntico a la mediana y el primer cuartil separa el cuarto inferior de los tres cuartos superiores” (Devore, 2008).

Los valores de los cuartiles se denotan como: Q_1 , Q_2 y Q_3 correspondientes con primero, segundo y tercer cuartiles, respectivamente; el valor Q_2 coincide con la mediana. De la misma forma se puede dividir en conjuntos de diez partes iguales (deciles) denotados como D_1, D_2, \dots, D_9 y los valores que dividen al conjunto en 100 partes iguales son los percentiles y se les denota P_1, P_2, \dots, P_{99} (Spiegel and Stephens, 2009).

2.2.2.2 Diagramas de Caja

Los diagramas de caja son representaciones gráficas del conjunto de datos, que entre sus características muestran (Devore, 2008):

- Centro
- Dispersión
- Naturaleza y magnitud de cualquier desviación respecto a la simetría
- Identificación de “valores atípicos”.

El diagrama de caja más sencillo se basa en el resumen de cinco cantidades:

- Mínima (xi)
- Cuarto inferior
- Mediana

- Cuarto superior
- Máxima (xi)

Para hacer un diagrama de caja (Devore, 2008):

- Primero, se traza una escala horizontal de medición.
- Después, se coloca un rectángulo sobre este eje; el lado izquierdo del rectángulo está en el cuarto inferior y el derecho en el cuarto superior (así, ancho de la caja =f).
- Después se traza un segmento de recta vertical o algún otro símbolo dentro del rectángulo en el lugar de la mediana
- Posteriormente se trazan unos “bigotes” desde cualquier extremo del rectángulo hasta las observaciones mínima y máxima.
- Si existen valores atípicos. Cada valor atípico moderado se representa con un círculo lleno.

Los valores atípicos son valores tan marcadamente diferentes del resto de la muestra que levantan la sospecha de que pueden ser de una población diferente o de que pueden estar en un error.

Tukey (1977) define que un valor es atípico si se encuentra 1.5IQR alejado del cuarto más cercano. Un valor atípico extremo está a más de 3 IQR alejado del cuarto más cercano. Se trata de un valor atípico moderado en cualquier otro caso. Donde el rango intercuartílico (IQR) está definido como el intervalo entre los cuartiles Q_3 y Q_1 .

Existen otros métodos para calcular valores atípicos (Olea, 2009; Seo, 2006; Jeong et al., 2017).

2.2.3 Métodos Numéricos

Las medidas descriptivas numéricas de un conjunto de datos buscan información numérica de interpretación significativa que pueda describir la distribución de las muestras. Las medidas son las siguientes: medidas de tendencia central y medidas de dispersión o variación (Mendenhall et al., 1986).

2.2.3.1 Medidas de tendencia central

Media La medida de tendencia central más común empleada en estadística es la media aritmética.

La *media aritmética* de una muestra de n respuestas medidas y_1, y_2, \dots, y_n está dada por:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (2.33)$$

Un problema con la media es que su valor puede ser afectado en gran medida incluso por la presencia de un solo valor atípico (Devore, 2008).

La *media geométrica* (G) de N números positivos $X_1, X_2, X_3, \dots, X_N$ es la raíz n -ésima del producto de los números (Spiegel and Stephens, 2009).

$$G = \sqrt[N]{X_1 X_2 X_3 \dots X_N}. \quad (2.34)$$

La *media armónica* (H) de un conjunto de N números $X_1, X_2, X_3, \dots, X_N$ es el recíproco de la media aritmética de los recíprocos de los números (Spiegel and Stephens, 2009):

$$H = \frac{1}{\frac{1}{N} \sum_{j=1}^N \frac{1}{X_j}} = \frac{N}{\sum \frac{1}{X}}. \quad (2.35)$$

Mediana Esta medida de tendencia central es el valor medio cuando las observaciones se ordenan de menor a mayor (Devore, 2008). Se obtiene al ordenar todas las observaciones (n) de menor a mayor incluyendo los valores repetidos.

Único valor medio si n es impar

$$\bar{X} = \left(\frac{n+1}{e} \right)^{simo}. \quad (2.36)$$

Si n es par la media es el promedio de los valores ordenados.

$$\left(\frac{n}{2} \right)^{simo} \text{ y } \left(\frac{n}{2} + 1 \right)^{simo}. \quad (2.37)$$

En el tratamiento de valores atípicos de los datos; la media y la mediana están en los extremos opuestos, ya que la media es sensible incluso a uno de estos valores, mientras que la mediana necesita una gran cantidad de valores atípicos para verse afectada.

Una solución para obtener la tendencia central en una muestra con datos atípicos es mediante una media recortada. Por ejemplo, eliminando el 10% más pequeño y 10% más grande de la muestra y luego se promedia el resto (Devore, 2008).

Moda La moda de un conjunto de números es el valor que se presenta con más frecuencia. Puede no haber moda y cuando la hay, puede no ser única. Se puede expresar como (Spiegel and Stephens, 2009):

$$Moda = L_1 + \left(\frac{\nabla_1}{\nabla_1 + \nabla_2} \right) c. \quad (2.38)$$

Donde:

L_1 = La frontera inferior de la clase modal (clase que contiene la moda)

∇_1 = Exceso de frecuencia modal sobre la frecuencia en clase inferior inmediata

∇_2 = Exceso de frecuencia modal sobre la frecuencia en la clase superior inmediata

c = Amplitud del intervalo de la clase modal

2.2.3.2 Medidas de dispersión

Dos conjuntos de mediciones podrían tener distribuciones de frecuencia muy diferentes pero iguales medias (figura 2.9). La diferencia entre las distribuciones se encuentra en la variación o dispersión de las mediciones que están a lado y lado de la media: y la variabilidad de datos (Mendenhall et al., 1986).

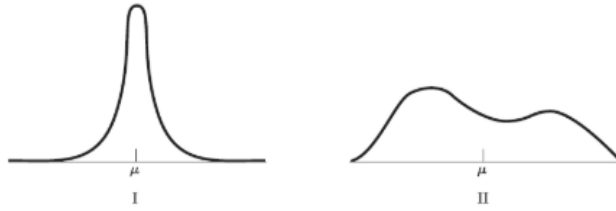


Figura 2.9: Distintas distribuciones con la misma media (Obtenido de Mendenhall et a. 1994).

Rango: El rango o recorrido es una medida de variabilidad simple la cual se define como la diferencia entre los valores máximo y mínimo en la muestra. Una desventaja del recorrido es que depende solo de las dos observaciones más extremas, y no toma en cuenta las posiciones de los valores restantes (Devore, 2008).

Desviación Estándar: La desviación respecto a la media es una medida de variabilidad que se obtiene restando la media de cada una de las “n” observaciones de la muestra (Devore, 2008):

$$\text{suma de las desviaciones} = \sum_{i=1}^n (x_i - \bar{x}). \quad (2.39)$$

Desviación respecto a la mediana: La desviación mediana consiste en tomar el promedio de las distancias con respecto a la mediana y se denota por (Bañuelos and Manzanares, 2016):

$$DMd = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X}). \quad (2.40)$$

Varianza: Es una función de las desviaciones (o distancias) de las mediciones muestrales desde la media. La varianza de una muestra de mediciones y_1, y_2, \dots, y_n es la suma del cuadrado de las diferencias entre las mediciones y su media, dividida entre $n - 1$ (Mendenhall et al., 1986):

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (2.41)$$

Cuanto mayor sea la varianza de un conjunto de mediciones, mayor será la cantidad de variación dentro del conjunto. La varianza puede determinar información acerca de la variación de un solo conjunto de datos cuando se interpreta en términos de la desviación estándar (Mendenhall et al., 1986).

Coefficiente de Variación: El coeficiente de variación es la medida de la dispersión en relación con la media de los datos. Se denota por V y está dado por (Spiegel and Stephens, 2009):

$$\text{coeficiente de variación}(V) = \frac{\sigma}{\bar{x}}. \quad (2.42)$$

Donde σ es la desviación de las unidades que se empleen por lo que es útil cuando se trata de comparar la desviación estándar y la media \bar{x} . El coeficiente de variación es adimensional o en porcentaje.

Rango semi-intercuartil El rango semi-intercuartil es otra medida de dispersión de un conjunto, se denota Q y se expresa como (Spiegel and Stephens, 2009):

$$Q = \frac{Q_3 - Q_1}{2}. \quad (2.43)$$

Donde Q_1 y Q_3 son el primero y tercer cuartiles en los datos.

2.2.4 Correlaciones

La correlación busca determinar qué tan bien la relación entre variables es explicada por una ecuación lineal, o de otro tipo. Cuando solo se intervienen dos variables se trata de un caso de *correlación simple* y *regresión simple*, y cuando intervienen más de dos variables se habla de un caso de *correlación múltiple* y *regresión múltiple* (Spiegel and Stephens, 2009).

Existen *correlaciones lineales* en las cuales los puntos tienden a formar una recta, mientras que si los puntos forman una curva, se trata de una *correlación no lineal*.

La representación de los puntos de datos en un plano x - y se le denomina *diagrama de dispersión*. Para determinar que tan bien una recta o una curva, describe los datos muestrales, se puede usar una técnica cuantitativa, mientras que la manera cualitativa es mediante las medidas de correlación. Las medidas de correlación son las siguientes (Bañuelos and Manzanares, 2016):

Covarianza de dos conjuntos de datos

La covarianza de dos conjuntos de datos es una media de la dispersión promedio de los datos respecto a sus medias, definida como:

$$Cov = \frac{SS_{xy}}{n}. \quad (2.44)$$

donde:

$$SS_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \quad (2.45)$$

Coefficiente de correlación

El coeficiente de correlación r está dado por (Spiegel and Stephens, 2009):

$$r = \pm \sqrt{\frac{\text{Variación Explicada}}{\text{Variación Real}}} = \pm \sqrt{\frac{\sum (Y_{est} - \bar{Y})^2}{\sum (Y - \bar{Y})^2}}. \quad (2.46)$$

En donde Y_{est} se calcula a partir de una ecuación de regresión lineal o no lineal.

Por la relación que guardan las variables existen tres tipos de correlación (Bañuelos and Manzanares, 2016):

- **Correlación directa o positiva:** Cuando al aumentar el valor de la variable dependiente, aumenta también el valor de la variable independiente o viceversa. Si la correlación es igual a la unidad, se tiene una correlación positiva perfecta.
- **Correlación inversa o negativa:** Cuando al aumentar el valor de la variable independiente, disminuye la variable dependiente. Si la correlación es igual a -1 se le denomina una correlación negativa perfecta.
- **Correlación nula:** Se da cuando no existe una relación lineal entre variables.

El *coeficiente de determinación* (r^2) proporciona el porcentaje de explicación de una variable en términos de la otra. Cuando r^2 es muy cercano a uno, el modelo explica en un mayor porcentaje el comportamiento de la variable independiente. Por su parte si r^2 es cercano a cero entonces el modelo no explica el comportamiento de las variables (Bañuelos and Manzanares, 2016).

En ocasiones la relación entre x y y no está dada por una recta. Sin embargo, la relación no lineal que existe entre x y y puede ser linealizada, mediante transformaciones lineales.

2.2.4.1 Modelos linealizables

La teoría estadística parte del supuesto de una distribución normal para desarrollar técnicas estadísticas clásicas como las medidas de tendencia central. Sin embargo, cuando no existe una distribución normal, se puede optar por transformar los datos a una escala en la que podamos sacar inferencias válidas de nuestros resultados (Olivier et al., 2008).

Un tipo común es la transformación logarítmica usando cualquier base para el logaritmo; no obstante, utilizar el logaritmo natural resulta ser más convencional. Generalmente se dice que una muestra se distribuye de forma log-normal cuando el logaritmo de la variable aleatoria se distribuye normalmente (Olivier et al., 2008). Existen otros tipos de transformaciones que involucran logaritmos, desarrollados en los trabajos de John and Draper (1980).

2.3 Introducción al lenguaje de programación R

R es un programa diseñado para computación estadística y gráficos, equipado con un lenguaje de programación, gráficos de alto nivel, interfaces a otros idiomas de programación y facilidades de depuración (Team, 2000). El programa está disponible en internet bajo la Licencia Pública General (GPL), lo que permite el libre, uso, distribución y lucro, siempre y cuando el receptor tenga los mismos derechos y el código fuente esté disponible gratuitamente (Dalgaard, 2005).

El *lenguaje R* es un dialecto de *S* que fue creado en la década de 1980 por su diseñador principal John M. Chambers. La sintaxis del lenguaje R tiene similitudes al lenguaje C, sin embargo, R es de la variedad de FPL o lenguaje de programación funcional con afinidades al Lisp y APL; R es un lenguaje de programación bien desarrollado, simple y efectivo que incluye condicionales, bucles, funciones recursivas definidas por el usuario e instrucciones de entrada y salida (Team, 2000).

El tipo de datos básicos en R es el *vector*, definido como una colección ordenada de valores del mismo tipo. Estos valores pueden ser numéricos, caracteres o lógicos. Otros tipos de datos incluyen listas (vectores con diferentes tipos de datos) y funciones. Las matrices, los marcos de datos y los objetos se crean a

Tabla 2.1: Tipos de datos en lenguaje de programación R

objeto	tipos	Multiples tipos
vector	numérico, caracter, complejo o lógico	No
factor	numérico o factor	No
arreglo	numérico, caracter, complejo o lógico	No
matriz	numérico, caracter, complejo o lógico	No
data.frame	numérico, caracter, complejo o lógico	Sí
ts	numérico, caracter, complejo o lógico	Sí
lista	numérico, caracter, complejo, lógico, expresión, función	Sí

Nota:

Tipos de datos del lenguaje de programación R en donde se muestran los diferentes objetos y tipos de datos que pueden guardar esos objetos. Además se indican si los objetos pueden guardar múltiples tipos de datos (Modificado de Paradis, 2002)

partir de vectores (Morandat et al., 2012). Los tipos de datos que usa R de acuerdo con Paradis (2002) se pueden observar en la tabla 2.1.

Se diferencia mucho con otros softwares de análisis de datos ya que posee un “*environment*” el cual tiene la intención de caracterizarlo como un sistema totalmente planificado y coherente, en lugar de una acumulación de herramientas muy específicas e inflexibles; permitiendo que se puedan desarrollar nuevos métodos de análisis de datos interactivos (Team, 2000).

Entre los procesos (figura 2.10) más básicos que puede realizar R se encuentran (Wickham and Grolemund, 2016):

- La visualización de datos mediante la creación de temas personalizados e informativos, así como técnicas poderosas para convertir datos en gráficos
- Transformación de datos mediante un conjunto de expresiones clave para seleccionar variables importantes, filtrar información, crear nuevas variables y computar compendios.
- Combinar información de la visualización y la transformación de datos para obtener conclusiones de los datos

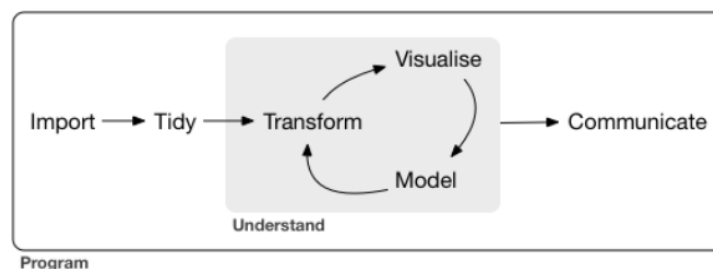


Figura 2.10: Procesado de datos (Obtenido de Wickham y Grolemund, 2016).

El sistema R se ha desarrollado rápidamente; y se ha ampliado con una extensa colección de paqueterías (packages). Un aspecto importante es la existencia de una lista de correo, además de diversos foros para la comunicación de informes de errores y debates sobre el desarrollo de R, por lo que R está en constante evolución (Dalgaard, 2005).

Hay cerca de 25 paquetes suministrados con R (llamados paquetes “estándar” y “recomendados”) y muchos más están disponibles a través de la familia de sitios de Internet CRAN (a través de <https://CRAN.R-project.org>) y en otros lugares (Team, 2000).

RStudio es un entorno de desarrollo integrado (IDE) para R (figura 2.11), que incluye una consola, editor de sintaxis que soporta la ejecución de un código directo, el ambiente donde están declaradas las variables, así como herramientas para el trazado, la historia de comandos ingresados, la depuración y la gestión del espacio de trabajo.

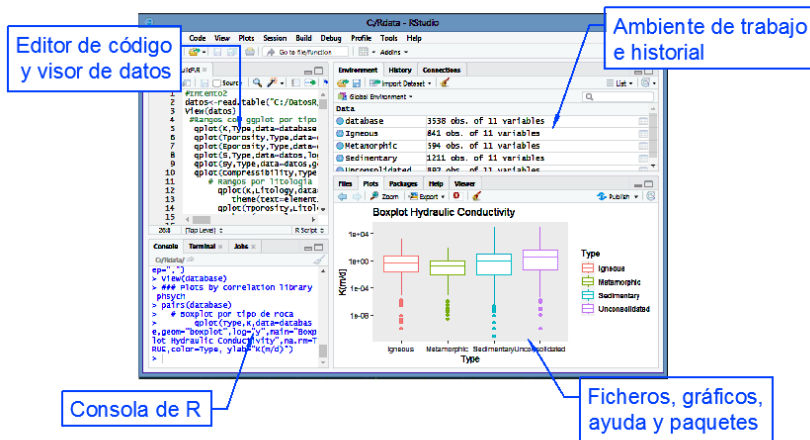


Figura 2.11: Entorno de trabajo del lenguaje R con su interfaz en Rstudio.

R posee varias paqueterías para hacer gráficos, siendo *ggplot2* una de las más elegantes y versátiles. Esta paquetería implementa la gramática de gráficos, así como un sistema coherente para describir y construir gráficos (Wickham and Grolemund, 2016). En general permite una mejor visualización de resultados, sobre todo enfocados en el momento de divulgación de resultados de investigación (Field et al., 2012). Puede complementarse la herramienta de *ggplot* con paqueterías como *plotly* permitiendo crear gráficos interactivos para el usuario.

Uno de los objetivos más importantes de un análisis estadístico es la

comunicación eficiente de la información a un público determinado. *Rstudio* resuelve este problema con paqueterías para el desarrollo de plataformas; un ejemplo es la paquetería *shiny* la cual facilita la creación de aplicaciones web interactivas directamente desde R (Team, 2020). Tutoriales y material didáctico para la creación de aplicaciones con la paquetería *shiny* pueden consultarse en el siguiente enlace: <https://shiny.rstudio.com/tutorial/>.

2.4 Bases de datos

Una base de datos es una colección de un conjunto de datos interrelacionados; organizados y presentados para satisfacer las necesidades de información de una comunidad determinada de usuarios (Byers et al., 1987). En una base de datos cada unidad de información almacenada está compuesta por datos elementales, cada uno de los cuales representa características particulares de la entidad que se describe (Rivera, 1994).

Las bases de datos constituyen una valiosa herramienta para el almacenamiento y procesamiento de la información (figura 2.12). Son, asimismo, un acceso fácil, eficiente, oportuno y preciso a diferentes acervos del conocimiento. Así como fuentes secundarias, ya que son la llave para llevarnos a la fuente original de la información (Rivera, 1994).

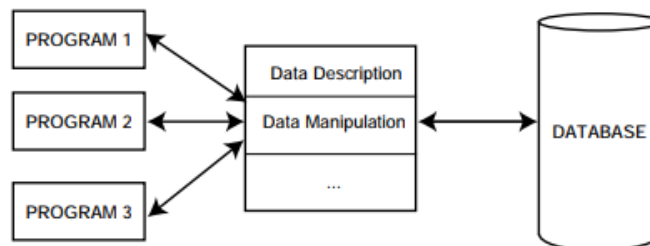


Figura 2.12: Procesado de una base de datos (Obtenido de Özsu y Valduriez, 1999).

Dentro de una base de datos es muy importante tener en cuenta que existen omisiones de información por pérdida o inexistencia de la misma por lo que en algunas ocasiones no se cuenta con el 100% de los datos. Los datos faltantes dentro de una base de datos suelen denotarse como *NA* (Cordero Sánchez, 2013 por ser siglas de “*Not Available*”).

Un aspecto importante en el uso de RStudio es el fácil manejo de bases de datos; ya que permite guardar la base en una variable para poder aplicar modificaciones posteriores. R proporciona una gran cantidad de herramientas estadísticas como pueden ser modelos lineales y no lineales, así como test estadísticos y análisis de series temporales. También permite de una manera muy sencilla crear gráficas (Cordero Sánchez, 2013).

Capítulo 3

Métodos

El método que se llevó a cabo para realizar el proyecto consiste en cuatro etapas, la primera de ellas fue una búsqueda y revisión minuciosa de bibliografía, en la cual se recopiló información geológica e hidrogeológica de interés. La segunda etapa consistió en la depuración y selección de datos para el diseño y generación de una base de datos de parámetros hidrogeológicos. En la tercera etapa la base de datos fue tratada con *Rstudio*, una herramienta de computación estadística para el análisis estadístico. Finalmente se programó una aplicación para la visualización de los resultados mediante la paquetería *shiny* de *Rstudio*. Cada etapa se describe a continuación.

3.1 Obtención de datos

Se revisaron más de 800 recursos bibliográficos entre libros, artículos, informes, revistas y boletines, para la obtención de información. Algunos ejemplos de las principales fuentes consultadas corresponden con: Hydrogeology Journal; Bulletin of the Geological Society of America; Journal of the American Water Resources Association; Journal of Geochemical Exploration; Journal of Geophysical Research; Journal of Structural Geology; Journal of Contaminant Hydrology; Journal of the Geological Society of India; Arabian Journal of Geosciences; así como revistas, informes y boletines de dependencias de diferentes países.

La consulta de libros físicos fue una parte importante en la etapa de obtención de datos. Para la indagación de bibliografía en internet se implementaron búsquedas con operadores booleanos para mejorar los resultados. Las búsquedas se realizaron dentro de potentes buscadores, en plataformas como: Elsevier, Springer, Mendeley, Google Scholar, etc.

Posteriormente se procedió a la revisión rápida y detallada de cada recurso, ya fueran libros, artículos, informes, etc. para la obtención de la mayor cantidad

de información geológica e hidrogeológica que estos pudieran brindar. Siempre buscando a grandes rasgos, obtener los siguientes datos:

- Datos de parámetros hidráulicos: *Porosidad Total* (ϕ_T), *Porosidad Efectiva* (ϕ_{ef}), *Conductividad Hidráulica* (K), *Coefficiente de Almacenamiento* (S), *Rendimiento Específico* (S_y) y *Compresibilidad* (α).
- Tipo o tipos de litología; o sedimentos no consolidados asociados a los valores de los parámetros hidráulicos.
- Breves comentarios de las litologías o sedimentos no consolidados.
- Locación, así como nombre de la formación o del acuífero.
- Bibliografía a partir de la cual se obtuvieron los datos.

Algunas de las consideraciones para el registro de datos fueron las siguientes:

- Para los casos en los que se presentaran rangos de valores parámetros hidráulicos se optó por registrar el valor mínimo y máximo.
- Si se mostraban valores de transmisividad y de espesor saturado; se optó por efectuar la operación para obtener la conductividad hidráulica en caso de que esta no estuviera expuesta por el autor.
- En los casos en que la Conductividad Hidráulica presentara dimensiones diferentes a m/d como ejemplo: ft/day, cm/seg, m/seg ; se procedió a realizar la conversión correspondiente.
- En los casos en que se presentaran valores de almacenamiento específico y el espesor del acuífero, se procedió a realizar el cálculo correspondiente para obtener el coeficiente de almacenamiento, si este no era mostrado por el autor.

3.2 Generación de la base de datos

La base de datos creada es de tipo estático, donde se almacena y registra información. La información tiene una estructura rectangular, es decir que cada renglón corresponde con observaciones, mientras que cada columna corresponde con una determinada variable.

Se generó una base de datos en formato (.csv), de información geológica e hidrogeológica a partir de la información recolectada. La base de datos se diseñó de tal forma que se mostrara la siguiente información:

- *Type*: Clasificación de la roca (Ígnea, Sedimentaria o Metamórfica) o si corresponde con sedimentos no consolidados.
- *Lithology*: Nombre de la litología.
- *Tporosity*: Valor de Porosidad Total en porcentaje (%).

Tabla 3.1: Encabezado de la base de datos

Type	Lithology	Porosity	Eporosity	K	S	Sy	Compressibility	Note	Location	Reference_ID
Igneous	Andesite	0.50	NA	NA	NA	NA	NA	NS	Doenstedt Germany	77
Igneous	Andesite	NA	NA	3.600000	9.23e-04	NA	NA	Fractured	Taipei basin Taiwan	78
Igneous	Andesite	NA	NA	8.380000	1.25e-03	NA	NA	Fractured	Taipei basin Taiwan	78
Igneous	Andesite	NA	NA	6.350000	5.29e-04	NA	NA	Fractured	Taipei basin Taiwan	78
Igneous	Andesite	NA	NA	1.911510	2.66e-03	NA	NA	Fractured	Taipei basin Taiwan	78
Igneous	Andesite	NA	NA	3.280000	7.78e-04	NA	NA	Fractured	Taipei basin Taiwan	78
Igneous	Andesite	NA	NA	NA	3.01e-04	NA	NA	NS	Taipei basin Taiwan	78
Igneous	Andesite	NA	NA	0.000864	NA	NA	NA	Dense	Zacapu basin MichoacAn	111
Igneous	Andesite	NA	NA	8.640000	NA	NA	NA	Fractured	Zacapu basin MichoacAn	111
Igneous	Andesite	NA	0.27	NA	NA	NA	NA	Fractured	LaLumbrebris-FlowtrackatVolcAndeColima MAExico	113
Igneous	Andesite	0.03	0.26	NA	NA	NA	NA	Fractured	LaLumbrebris-FlowtrackatVolcAndeColima MAExico	113
Igneous	Andesite	0.30	0.29	NA	NA	NA	NA	Fractured	LaLumbrebris-FlowtrackatVolcAndeColima MAExico	113
Igneous	Andesite	NA	NA	0.200000	NA	NA	NA	Fractured	Jeju Island Korea	412
Igneous	Andesite	NA	NA	124.900000	NA	NA	NA	Fractured	Jeju Island Korea	412
Igneous	Andesite	NA	5.00	0.020000	NA	NA	NA	Fractured	Valle Central northern Costa Rica	457
Igneous	Andesite	NA	0.25	0.500000	NA	NA	NA	Fractured	Valle Central northern Costa Rica	457
Igneous	Andesite	0.20	NA	0.043200	2.50e-05	NA	NA	Fractured	The aquifer of the Puebla Valley MAExico	461
Igneous	Andesite	0.18	NA	0.432000	2.00e-07	NA	NA	NS	The aquifer of the Puebla Valley MAExico	461
Igneous	Andesite	NA	NA	0.691200	NA	NA	NA	Fractured	The aquifer of the Puebla Valley MAExico	461
Igneous	Andesite	0.16	NA	NA	NA	NA	NA	NS	Maya CheliAbinsk en Rusia	545

- *Eporosity*: Valor de Porosidad Efectiva en porcentaje (%).
- *K*: Valor de Conductividad Hidráulica en unidades (m/d).
- *S*: Valor del Coeficiente de Almacenamiento adimensional (-).
- *Sy*: Valor del Rendimiento Específico adimensional (-).
- *Compressibility*: Valor de la compresibilidad en unidades (Pa^{-1} o m^2/N).
- *Note*: Información adicional importante a ser considerada.
- *Location*: Lugar donde se tomaron las muestras o medidas de acuerdo con el autor.
- *Reference_ID*: Bibliografía consultada.

3.3 Procesado de datos con R

Para el estudio de la estadística descriptiva de la base de datos se utilizó una herramienta de computación estadística denominada *Rstudio*, la cual provee un entorno de desarrollo para lenguaje de programación R.

Con esta herramienta se realizó el siguiente proceso:

- Lectura de la base de datos.
- División en grupos y subgrupos de la información de la base de datos.
- Generación de gráficos: Histogramas, Diagramas de caja (Boxplot), Diagramas de Rangos y correlaciones.
- Elaboración de tablas con información estadística de interés: Medidas de tendencia central y medias de dispersión.
- Creación de una aplicación web con la paquetería shiny para la consulta de la información.

3.3.1 Creación de un nuevo proyecto

Para iniciar con el proceso se creó un nuevo proyecto. Dentro de Rstudio se seleccionó la pestaña *File* -> *New Project* -> *New Directory* -> *New Project* , y se guardó el proyecto con el nombre “Rdata” en la carpeta *C:/* .

Dentro del nuevo proyecto se abrió un nuevo *script* para comenzar a escribir el código. Para crear un nuevo script se selecciona la pestaña *File* -> *New File* -> *Rscript*. Este *script* también se guardó dentro de la carpeta del proyecto con el nombre *script1*.

De la misma manera la base de datos *hydraulic_parameters.csv* también se guardó dentro del proyecto en la dirección: *C:/Rdata/*.

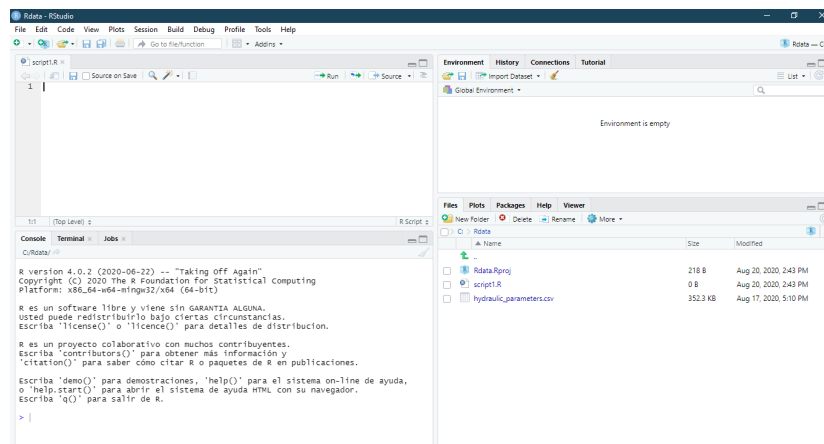


Figura 3.1: Nuevo proyecto en Rstudio.

3.3.2 Instalación de paqueterías y lectura de datos

El procedimiento que se llevó a cabo para realizar la estadística descriptiva de los datos es el siguiente:

Primero se instalaron las paqueterías que servirán de ayuda para optimizar funciones. Las paqueterías principales para un análisis preliminar son las siguientes: *ggplot2*, *dplyr*. Las paqueterías usadas se instalaron con el comando *install.packages()*.

```
# Instala las paqueterías necesarias
install.packages(c(ggplot2, dplyr))
```

```
# Habilita las paqueterías para que R pueda utilizarlas
library(ggplot2)
library(dplyr)
```

```
# Asigna un directorio de trabajo
setwd("C:/Rdata")

# Carga la base de datos desde la dirección "C:/Rdata/"
#y la guarda en una variable llamada "database"
database<-read.csv("C:/Rdata/hydraulic_parameters.csv",
                  header=TRUE, sep=",")

# Permite visualizar los datos de la variable database
View(database)
```

Después se procedió a habilitar las paqueterías necesarias con el comando `library()`.

Para conocer las funciones básicas y especializadas de cada paquetería se utilizó el comando `help()` dentro de la consola. Esta función permite el acceso a una hoja de referencia con toda la información relacionada a la paquetería especificada.

```
# Información de la paquetería "ggplot2"
help("ggplot2")
```

Dentro del programa Rstudio se asigna un directorio de trabajo con la instrucción `setwd()`, el directorio corresponde con la dirección y carpeta donde se guardarán todos los resultados y de donde se leerán los archivos. En este caso una carpeta en la unidad *C*: llamada *Rdata*:

Para cargar la base de datos se creó una nueva variable llamada *database*; para asignarle un valor a la variable se utiliza el signo “<-”. El comando `read.csv` permite leer el archivo (base de datos), ubicado en *C:/Rdata/*.

El comando `read.csv` utiliza los siguientes parámetros:

- El nombre del archivo, con la extensión *.csv* entre comillas, que está alojado en el directorio de trabajo.
- El segundo parámetro es la instrucción `header=TRUE`, el cual indica que el primer renglón corresponde con los nombres de cada columna (encabezado).
- El tercer parámetro `sep=","` le indica al programa el tipo de tabulación. En este caso al ser un *.csv delimitado por comas*, el separador es una coma.

Es importante tener en cuenta que R tiene diferentes entradas para la lectura de datos por lo que se puede leer archivos con extensión *.xlsx* o *.txt* entre otros. R carga los datos como un objeto de tipo *data.frame* el cual es una estructura rectangular de datos compuesta por datos heterogéneos. Para verificar la estructura interna de los datos se utiliza la instrucción `str()`.

```
str(database)
```

```
## 'data.frame': 3651 obs. of 11 variables:
## $ Type : chr "Igneous" "Igneous" "Igneous" "Igneous" ...
## $ Lithology : chr "Andesite" "Andesite" "Andesite" "Andesite" ...
## $ Tporosity : num 0.5 NA NA NA NA NA NA NA NA ...
## $ Eporosity : num NA NA NA NA NA NA NA NA 0.27 ...
## $ K : num NA 3.6 8.38 6.35 1.91 ...
## $ S : num NA 0.000923 0.00125 0.000529 0.00266 0.000778 0.000301 NA NA ...
## $ Sy : num NA NA NA NA NA NA NA NA ...
## $ Compressibility: num NA NA NA NA NA NA NA NA ...
## $ Note : chr "NS" "Fractured" "Fractured" "Fractured" ...
## $ Location : chr "Doonstedt_Germany." "Taipei basin_Taiwan" "Taipei basin_Taiwan" ...
## $ Reference_ID : chr "77" "78" "78" "78" ...
```

Con el fin de verificar que los datos se hayan cargado correctamente, el comando `View()` permite visualizar e inspeccionar los datos en una nueva pestaña dentro del visualizador de datos.

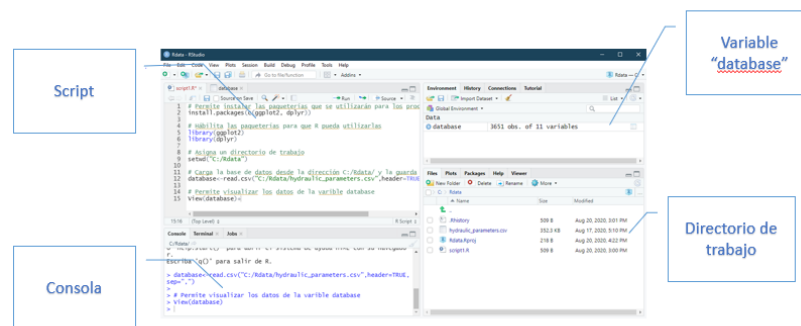


Figura 3.2: Script en Rstudio

Una vez cargada la base de datos, el siguiente paso es la división en subconjuntos para una mejor manipulación de la información. Para describir este proceso se toma como ejemplo la creación de subconjuntos de datos que contengan la información de parámetros hidráulicos exclusivamente de cada tipo de roca, como se muestra en el siguiente script:

```
# Pasos descritos previamente
library(ggplot2)
library(dplyr)

setwd("C:/Rdata")

database<-read.csv("hydraulic_parameters.csv",header=TRUE,sep=",")

# Crea subconjuntos y los guarda en sus respectivas variables
Igneous<-database[database$Type=="Igneous", ]

Metamorphic<-database[database$Type=="Metamorphic", ]

Sedimentary<-database[database$Type=="Sedimentary", ]
```



```
Unconsolidated<-database[database$Type=="Unconsolidated", ]
```

Es decir que creamos nuevas variables cuyo nombre es el de cada tipo de roca en donde se guarda la información de los subconjuntos de la variable creada con el nombre *database*. La instrucción de extracción es mediante corchetes `[[`, donde se debe indicar, que renglones y que columnas extraer, separados por una coma dentro de los corchetes. Por ejemplo, para extraer los parámetros hidrogeológicos de rocas ígneas, dentro de los corchetes se especifica los renglones de la columna en este caso los renglones de la columna *Type* (`database$Type` en el cual los valores sean iguales a *Igneous* ; mientras que para declarar que se guarden todas las columnas, se deja un espacio en blanco después de la coma.

3.3.3 Generación de gráficos con la paquetería ggplot

Rangos

Los diagramas de rangos se realizaron con el siguiente script:

```
# Pasos descritos previamente
library(ggplot2)

database<-read.csv("hydraulic_parameters.csv",header=TRUE,sep=",")

# Plot con ggplot2
ggplot(data=database, aes(x=K,y=Type,color=Type))+
  geom_line(na.rm=TRUE)+scale_x_log10()+
  theme_bw()
```

En donde la instrucción *geom()* permite declarar los datos, en este caso la variable *database* . En la instrucción *aes()* se declaran las variables *x* y *y* que se van a graficar, así como de los valores a partir de los cuales generarán colores. El siguiente comando es *geom_line()* el cual da la instrucción del tipo de gráfico que se quiere generar, en este caso líneas, dentro de esta instrucción se declara *na.rm=TRUE* para que se eliminen los valores *NA* al momento de graficar. Finalmente se agrega la instrucción *scale_x_log10()* para indicar una escala logarítmica en el eje *x*.

El comando *theme_bw()* es opcional y permite cambiar el estilo del gráfico.

Se generaron diagramas de rangos para cada propiedad hidrogeológica: Porosidad Total, Porosidad Efectiva, Conductividad Hidráulica, Coeficiente de Almacenamiento, Rendimiento Específico y Compresibilidad; en función de su clasificación por tipo de roca y en función de su litología.

Boxplot

Se elaboraron un conjunto de diagramas de caja o boxplot con la paquetería *ggplot2*. A continuación se presenta el script:

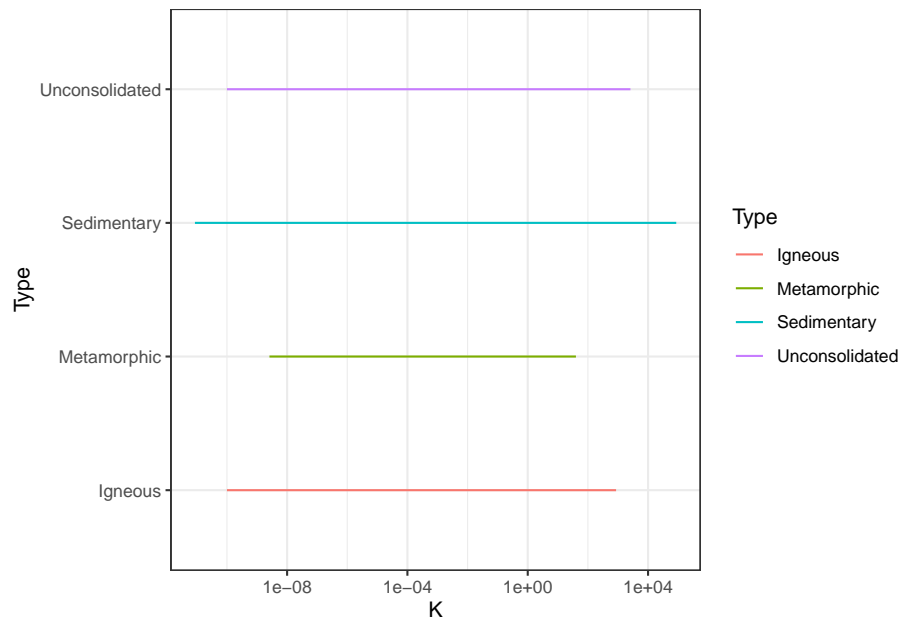


Figura 3.3: Diagrama de rangos de la conductividad hidráulica (m/d) con ggplot2.

```
library(ggplot2)

database<-read.csv("hydraulic_parameters.csv",header=TRUE,sep="," )

# Boxplot con ggplot2
ggplot(data=database, aes(x=Type,y=Tporosity,color=Type))+
  geom_boxplot(na.rm=TRUE)+theme_bw()
```

El script genera diagramas de caja para cada clasificación de rocas (ígneas, sedimentarias, metamórficas, no consolidados). Para poder realizar este diagrama se utiliza la función `ggplot()`, indicando los datos y la función `aes()` con las variables que se van a graficar, seguido de la instrucción `geom_boxplot()` que indica el tipo de gráfico, en este caso un boxplot; dentro de esta función se declara `na.rm=TRUE` para eliminar los valores faltantes o *NA*. El parámetro `theme_bw()` es opcional y permite cambiar el estilo del gráfico.

El programa Rstudio utiliza por default el método de Turkey (1977) para discriminar valores atípicos en donde: se define que un valor es atípico si se encuentra 1.5IQR alejado del cuarto más cercano.

Se generaron diagramas de caja para cada propiedad hidrogeológica: *Porosidad Total*, *Porosidad Efectiva*, *Conductividad Hidráulica*, *Coefficiente de*

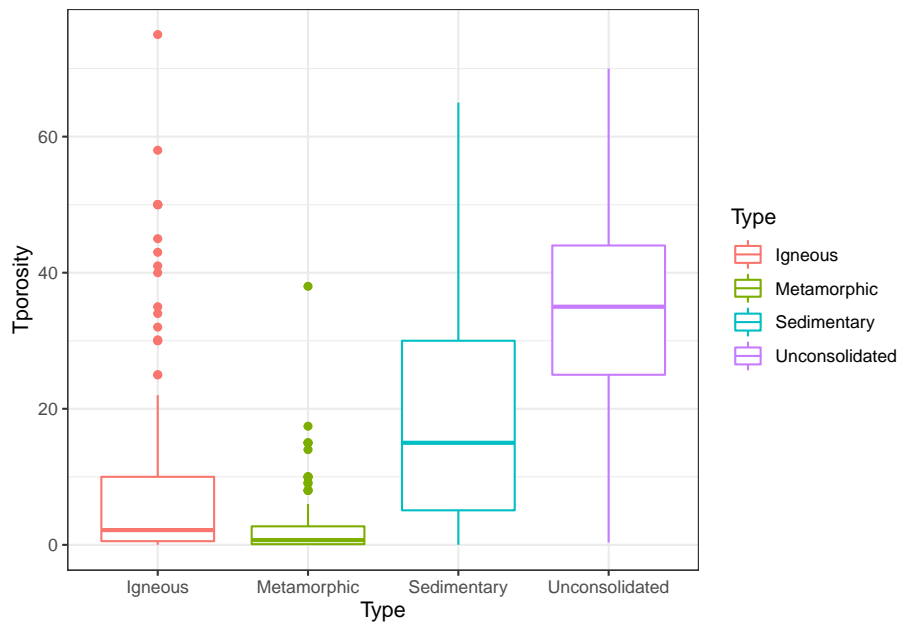


Figura 3.4: Diagramas de caja de la porosidad total (%) con ggplot2.

Almacenamiento, Rendimiento Específico y Compresibilidad. En función de su clasificación por tipo de roca y en función de su litología.

R tiene varias formas para manejar la escala logarítmica, la primera es mediante la instrucción `scale_y_log()` o `scale_x_log()` las cuales hacen una transformación de los datos y posteriormente agregan una escala; la otra forma es mediante la instrucción `coord_trans(y="log10")` o `coord_trans(x="log10")` que transforman el sistema de coordenadas sin aplicar ninguna transformación a los datos.

Para hacer diagramas de caja con escala logarítmica se utilizó una transformación en el sistema de coordenadas con la instrucción `coord_trans(y="log10")`. La razón de usar una transformación de coordenadas en lugar de una transformación de datos con la instrucción `scale_y_log10()` es porque esta última afecta la estimación de *outliers* o valores anómalos. El método se describe con el siguiente script:

```
# Boxplot con escala logarítmica en el eje y
ggplot(data=database, aes(x=Type,y=K,color=Type))+
  geom_boxplot(na.rm=TRUE)+coord_trans(y = "log10")+
  scale_y_continuous(breaks = c(1e-8,1e-4,1,1e+4),
                    minor_breaks=c(1e-8,1e-4,1,1e+4)
                    )+
```

```
annotation_logticks(scaled = FALSE)+
theme_bw()
```

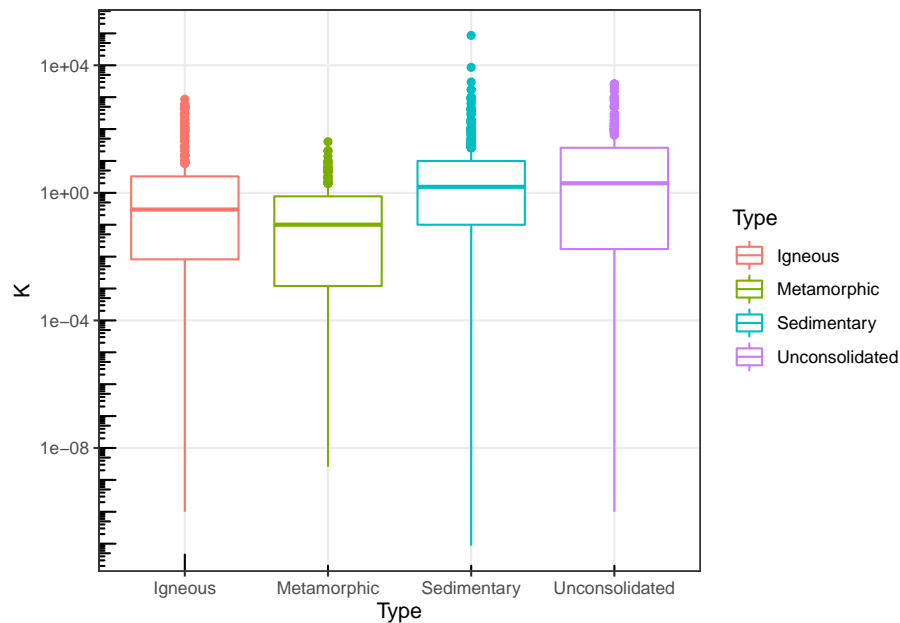


Figura 3.5: Diagramas de caja de la conductividad hidráulica (m/d) en escala logarítmica con ggplot2.

Para realizar un gráfico con una transformación en el sistema de coordenadas se especifican líneas de código para construir un boxplot y se agrega la instrucción `coord_trans()` cuyo parámetro es $y = \text{"log10"}$, que indica el tipo de transformación de las coordenadas en el eje y. La instrucción `scale_y_continuos()` se utiliza para agregar los `breaks` donde se define un vector con las medidas que indican donde se graficarán los valores de los ejes y `minor_breaks` que también es un vector donde se indican las medidas en las que se graficarán las líneas de la malla del gráfico. También puede agregarse el comando `annotation_logticks()` para agregar marcadores logarítmicos en los ejes, especificando `scale=FALSE` ya que no se ha aplicado ninguna transformación previa de los datos.

Histograma

Los histogramas para cada propiedad hidrogeológica (*Porosidad Total*, *Porosidad Efectiva*, *Conductividad Hidráulica*, *Coefficiente de Almacenamiento*, *Rendimiento Específico* y *Compresibilidad*) se construyeron para cada clasificación litológica (ígneas, sedimentarias, metamórficas, no consolidados) y cada una de las litologías más representativas.

Para construir un histograma se puede utilizar el siguiente script:

```
library(ggplot2)

database<-read.csv("hydraulic_parameters.csv",header=TRUE,sep=",")

# Histograma con ggplot2
ggplot(data=database, aes(x=S))+
  geom_histogram(na.rm=TRUE, bins=15)+ scale_x_log10()+
  theme_bw()
```

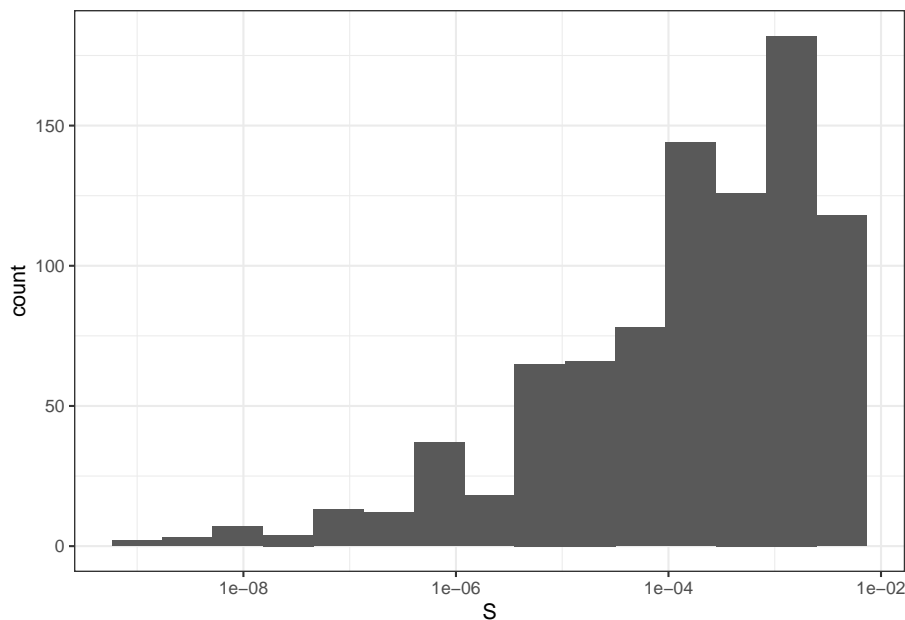


Figura 3.6: Histograma del coeficiente de almacenamiento (-) con ggplot2.

Este script permite generar un único histograma con los valores del coeficiente de almacenamiento para todas las rocas. Se construye de la misma forma que los otros gráficos con la función `ggplot()` en la que se declaran los datos y las variables que se quieren graficar, y se agrega la función `geom_histogram()` para que la gráfica resultante sea un histograma, dentro de la función se declara `na.rm=TRUE` para deshacerse de los valores faltantes `NA`, además de que se indica el número de clases con el parámetro `bins`. La instrucción `theme_bw()` cambia el estilo del gráfico.

Ggplot2 tiene muchas herramientas, una de ellas es la capacidad de generar matrices de gráficos como el siguiente:

```
library(ggplot2)

database<-read.csv("hydraulic_parameters.csv",header=TRUE,sep=",")

# Matriz de histogramas con ggplot2
ggplot(data=database, aes(x=S,fill=Type))+
  geom_histogram(na.rm=TRUE, bins=15)+
  scale_x_log10()+facet_wrap(~Type,nrow=2)
```

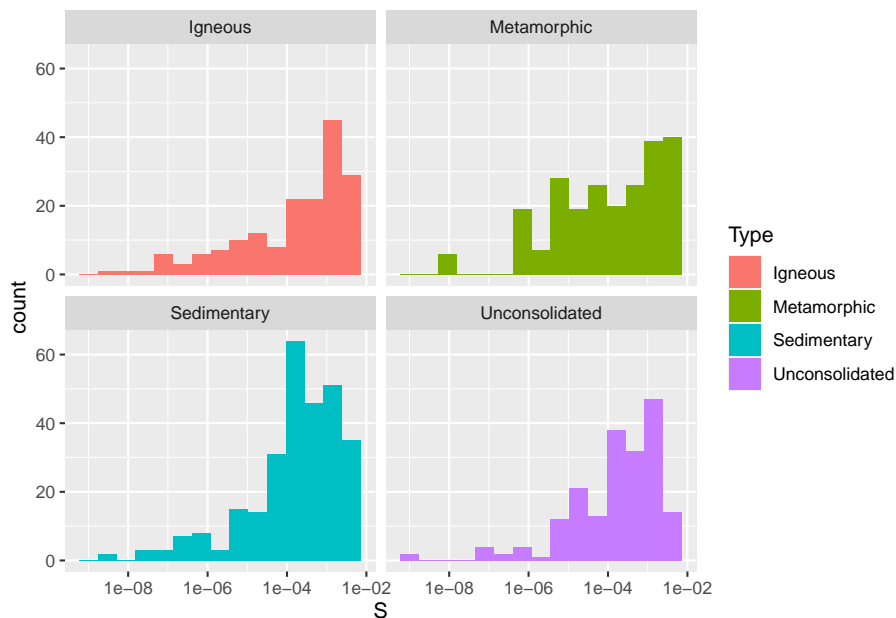


Figura 3.7: Histogramas del coeficiente de almacenamiento (-) con ggplot2.

En este nuevo ejemplo se declara dentro de *aes()* el parámetro *fill*, el cual da color a los histogramas. Para hacer la matriz, se agrega una nueva función llamada *facet_wrap()* en la cual se declaran el signo (*~*) que indica en función de "...", seguido del parámetro con el cual se hará la matriz de gráficos, en este caso el tipo de roca (*Type*). De la misma forma *nrow* indica el número de renglones de la matriz.

Diagramas de correlación

Las correlaciones se pueden graficar con ggplot como se muestra a continuación:

```
#install.packages("ggpmisc")

## Librerías utilizadas
```

```

library(ggplot2)
library(ggpmisc)

database<-read.csv("hydraulic_parameters.csv",header=TRUE,sep=",")

# Plot con ggplot2
ggplot(data=database, aes(x=K,y=Tporosity))+geom_point()+
  scale_x_log10()+
  geom_smooth(method=lm,formula=y~x,se=FALSE)+
  stat_poly_eq(formula = y~x,na.rm=TRUE,
               aes(label = paste(..eq.label.., ..rr.label..,
                                   sep = "~~~")),
               size = 4, parse = TRUE)+
  theme_bw()

```

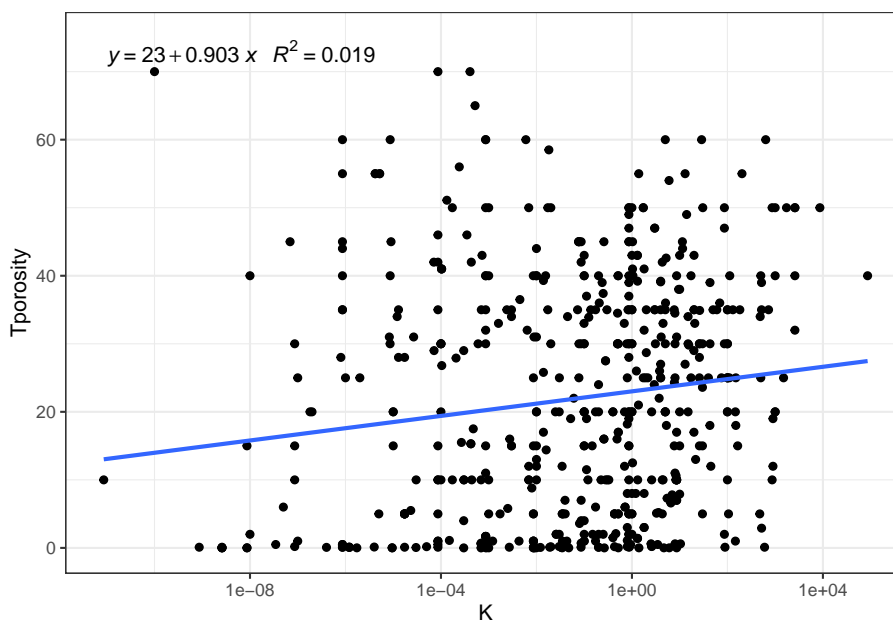


Figura 3.8: Diagrama de correlación de la conductividad hidráulica (m/d) vs Porosidad Total (%) con ggplot2.

Por la falta de valores relacionados entre sí, es decir que existiera información de más de una propiedad hidrogeológica para cada observación, en la base de datos; los diagramas de dispersión y las correlaciones se hicieron a gran escala tomando todos los tipos de rocas.

El diagrama se crea a partir de declarar los datos y las variables a graficar

dentro de la función `ggplot()` posteriormente se agrega una geometría de puntos `geom_points`, además de que se agrega una escala logarítmica el eje x `scale_x_log10()`. Se añade una nueva instrucción `geom_smooth()` cuyo primer parámetro es `method="lm"`, el cual determina las estimaciones de mínimos cuadrados de los parámetros de un modelo lineal con la `formula = y ~ x` y ajusta una línea de regresión. El comando `se` indica si se quiere indicar el intervalo de confianza. Finalmente el parámetro `theme_bw()` permite cambiar el estilo del gráfico.

Para agregar una etiqueta con la información del modelo de regresión y el coeficiente de determinación se agrega la instrucción `stat_poly_eq()` la cual es propia de la librería `ggpmisc`. Los parámetros de esta instrucción son: la fórmula de regresión `y ~ x`; `na.rm=TRUE` para eliminar valores faltantes; el tamaño de la etiqueta `size`; el parámetro `parse=TRUE` para computar la etiqueta como texto visible; así como la instrucción `aes()` con el parámetro `label` para indicar un texto con la fórmula (`..eq.label..`) y el coeficiente de determinación (`..rr.label..`).

3.3.4 Generación de gráficos interactivos con la paquetería plotly

Para darle un aspecto más dinámico y llamativo a los gráficos se utilizó la paquetería `plotly` para generar gráficos interactivos que después se pudieran acoplar a la aplicación hecha con la paquetería `shiny`.

Se debe instalar la paquetería `plotly` con la instrucción `install.packages("plotly")`, si esta no ha sido instalada previamente. En el siguiente ejemplo se muestra un script de un gráfico creado con `plotly`:

```
#install.packages("plotly")

library(plotly)

database<-read.csv("hydraulic_parameters.csv",header=TRUE,sep=",")

# Boxplot con la paquetería plotly
plot_ly(database, y = ~K, x=~Type, type = "box", boxpoints = "all",
         jitter = 0.3,color=~Type,
         pointpos = -1.8)%>%layout(yaxis=list(type="log"))
```

Este tipo de gráficos generan una serie de herramientas interactivas dentro de una pantalla emergente en `rstudio` para una mejor visualización de los datos.

Para generar un gráfico interactivo, en este caso de un boxplot; se agrega la función `plot_ly()` con una serie de parámetros. El primer parámetro son los datos con los que se va a trabajar, los siguientes parámetros son las variables a graficar, las cuales deben de estar precedidas por el símbolo (`~`) que significa en función de "...", el cuarto parámetro es el tipo de geometría, en este caso un diagrama de caja `box`. El parámetro `boxpoints="all"` permite que se grafiquen

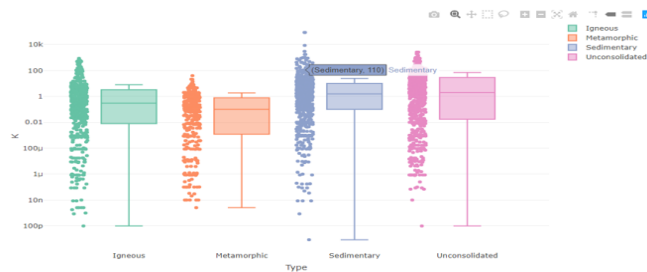


Figura 3.9: Diagramas de caja interactivos de la conductividad hidráulica.

puntos y diagramas de caja en el mismo gráfico. Los comandos *jitter* y *color* son estéticos y dan tamaño y color respectivamente, a los puntos. Para asignar un tamaño a las etiquetas desplegables se utiliza el parámetro *pointpos*. Finalmente la función *layout()* que está compuesta por el parámetro *yaxis*, asigna una escala logarítmica al gráfico.

La sintaxis de *plotly* es bastante diferente a la de *ggplot2*, no obstante, *plotly* puede incorporar gráficos de *ggplot2*, por lo que los gráficos hechos con *ggplot2* también pueden ser interactivos como a continuación se muestra:

```
library(plotly)
library(ggplot2)

database<-read.csv("hydraulic_parameters.csv",header=TRUE,sep=",")

# Histogramas con ggplot
plot<- ggplot(data=database, aes(x=Tporosity,fill=Type))+
  geom_histogram(na.rm=TRUE, bins=15)+
  facet_wrap(~Type,nrow=2)

# Grafico inteactivo del gráfico hecho con ggplot
ggplotly(plot)
```

Para hacer interactivo un gráfico hecho con *ggplot2* basta con guardar el gráfico creado en una variable, en este ejemplo la variable es *plot*. Posteriormente se hace un llamado a la función *ggplotly()* con la variable creada anteriormente como argumento.

3.3.5 Generación de tablas con medidas numéricas

La función *summary()* elabora resúmenes, por lo que al declararla se puede tener una idea general y preliminar de valores mínimos, máximos, medias, etc.

```
# Resumen de la variable "database" creada anteriormente
summary(database)

##      Type      Lithology      Tporosity      Eporosity      K
## Length:3651 Length:3651  Min. : 0.0001 Min. : 0.0  Min. : 0.00
## Class :character Class :character  1st Qu.: 2.1975  1st Qu.: 0.7  1st Qu.: 0.02
```

```
## Mode :character Mode :character Median :13.0000 Median : 6.0 Median : 0.77
## Mean :17.9953 Mean :11.4 Mean : 84.06
## 3rd Qu.:30.0000 3rd Qu.:20.0 3rd Qu.: 7.84
## Max. :75.0000 Max. :65.0 Max. :86400.00
## NA's :2631 NA's :2766 NA's :1488
##
## S Sy Compressibility Note Location
## Min. :0.0000 Min. :0.0020 Min. :0 Length:3651 Length:3651
## 1st Qu.:0.0000 1st Qu.:0.0139 1st Qu.:0 Class :character Class :character
## Median :0.0002 Median :0.0380 Median :0 Mode :character Mode :character
## Mean :0.0009 Mean :0.0724 Mean :0
## 3rd Qu.:0.0014 3rd Qu.:0.1000 3rd Qu.:0
## Max. :0.0050 Max. :0.5700 Max. :0
## NA's :2776 NA's :2479 NA's :3461
## Reference_ID
## Length:3651
## Class :character
## Mode :character
##
##
##
```

Sin embargo, esta función tiene bastantes limitantes, por lo que no es posible inspeccionar la complejidad de los datos y es necesario utilizar otras funciones.

Para poder obtener información de datos agrupados se utilizó la función `tapply()`, una función vectorizada que divide grupos y les aplica una función determinada, en este caso `summary()`. El siguiente es un ejemplo para generar resúmenes de la variable *Porosidad Total* (*Tporosity*) en función del tipo de litología (ígnea, sedimentaria, metamórfica, no consolidados):

```
# Aplicación de la función tapply para generar medidas numéricas
table1<- tapply(database$Tporosity,database$Type,summary)
```

```
table1
```

```
## $Igneous
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.   NA's
## 0.0001 0.5425 2.1750 7.3894 10.0000 75.0000   639
##
## $Metamorphic
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.   NA's
## 0.0001 0.1025 0.7000 2.8103 2.7250 38.0000   544
##
## $Sedimentary
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.   NA's
## 0.010 5.075 15.000 17.947 30.000 65.000   769
##
## $Unconsolidated
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.   NA's
## 0.35 25.00 35.00 35.25 44.00 70.00   679
```

Esta función está formada por tres parámetros. El primero es la columna de la variable de la cual se quieren obtener medidas numéricas, para este caso la columna de la *Porosidad Total*; el segundo parámetro es la columna a partir de la cual se quieren generar subgrupos, en este caso los subgrupos de la columna *Type*. El último parámetro es la función que se quiere realizar, para este caso `summary()`.

Una manera mucho más flexible para crear tablas con medidas numéricas es mediante la paquetería *dplyr* la cual contiene el comando *summarise()* que crea una o más variables escalares, que resumen las variables de una tabla existente. Un ejemplo se muestra a continuación, donde se realiza una tabla a manera de resumen de la media aritmética, mediana y media geométrica (para usar la función de media geométrica se necesita habilitar la paquetería *EnvStats*). Si no están instaladas las paqueterías necesarias se debe proceder a instalarlas:

```
##Instalar paqueterías
#install.packages(c(dplyr,EnvStats))

library(dplyr)
library(EnvStats)

# Tabla de medidas de tendencia central
table2<-database %>%
  group_by(Type) %>%
  summarise(mean=round(mean(K,na.rm=TRUE),3),
            median=round(median(K,na.rm=TRUE),3),
            geometric_mean=round(geoMean(K,na.rm=TRUE),3)
            )

table2

## # A tibble: 4 x 4
##   Type          mean median geometric_mean
##   <chr>        <dbl> <dbl>         <dbl>
## 1 Igneous      25.8   0.3           0.078
## 2 Metamorphic  1.00   0.1           0.017
## 3 Sedimentary 174.    1.54          0.607
## 4 Unconsolidated 76.0   2             0.446
```

Para realizar el código se creó una nueva variable *table2* para guardar la información de la tabla. El operador pipe (*%>%*) permite concatenar múltiples funciones para implementar procesos. El comando *group_by(Type)* literalmente describe el proceso de agrupar de acuerdo a las categorías de la columna *Type*. Posteriormente se concatena con el comando *summarise()*. Los argumentos del comando *summarise()* corresponden con la funciones: media *mean()*, media geométrica *geoMean()* y la mediana *median()*; cuyos argumentos son la variables de las cuales se quiere hacer el resumen (*Columna “K” de la conductividad hidráulica*); y el parámetro para eliminar valores faltantes (*na.rm=TRUE*). Además se le asigna la función *round()*, indicando el número de decimales para redondear los valores de salida. Finalmente, para visualizar la nueva tabla se utiliza el comando *View()*.

```
# Visualizar la nueva tabla
View(table2)
```

Este proceso se realizó para cada uno de los valores de interés como: valores mínimos (*min()*), valores máximos (*max()*), cuartiles (*quantile()*), desviación estándar (*sd()*). Para saber la función de cada comando se utilizó la intrucción *help()*, por ejemplo: *help("quantile")*.

3.3.6 Generación de tabla interactiva con la paquetería DT

La paquetería *DT* permite generar una interfaz para la manipulación de tablas con herramientas como filtros y búsquedas, además de que puede acoplarse con la librería *shiny* para la creación de aplicaciones.

```
#install.packages("DT") #Instalación de paquetería
```

```
library(DT)
```

```
datatable(database, filter="top")
```

La función *datatable()* tiene como argumentos, el objeto *data.frame* (*database*) y el parámetro *filter="top"* que permite agregar casillas para filtrar datos en el encabezado de la tabla.

	Type	Litology	Tporosity	Eporosity	K	S	Sy	Compressibility	Note	Location	Source
1	Igneous	Andesite	0.5						NS	Doenstedt, Germany	77
2	Igneous	Andesite			3.6	0.00923			Fractured	Taipei basin, Taiwan	78
3	Igneous	Andesite			8.38	0.00125			Fractured	Taipei basin, Taiwan	78
4	Igneous	Andesite			6.35	0.00529			Fractured	Taipei basin, Taiwan	78
5	Igneous	Andesite			1.91151	0.00266			Fractured	Taipei basin, Taiwan	78
6	Igneous	Andesite			3.29	0.00078			Fractured	Taipei basin, Taiwan	78
7	Igneous	Andesite				0.00301			NS	Taipei basin, Taiwan	78
8	Igneous	Andesite		0.000864					Dense	Zacapu basin, MichoacÁn	111
9	Igneous	Andesite			8.64				Fractured	Zacapu basin, MichoacÁn	111
10	Igneous	Andesite		0.27					Fractured	L.L.umbredebris-fleotrackerVoleÁndeColima, MÃ©xico	113

Figura 3.10: Interfaz interactiva de la tabla "database" con la paquetería DT.

3.3.7 Creación de una aplicación con la paquetería shiny

Para crear una nueva aplicación desde *Rstudio* primero se instaló la paquetería *shiny* con la instrucción *install.packages("shiny")*. Una vez instalada se creó un nuevo directorio de trabajo con el nombre "app" en Windows en el directorio *c:/*, en el cual se guardó el archivo de la base de datos (.csv) además de dos scripts con los nombres *ui.R* y *server.R*. Estos dos scripts corresponden con la

interfaz (*ui.R*) y con el servidor (*server.R*). Para conocer las funcionalidades de la paquetería *shiny* se consultó la hoja de referencia <https://rstudio.com/wp-content/uploads/2015/03/shiny-spanish.pdf>.

```
# Nuevo directorio de trabajo
setwd("C:/app")
```

Los dos scripts deben contener las siguientes instrucciones:

ui.r Script

```
library(shiny)

# Define la interfaz que será visible al usuario,
##así como los datos de entrada

shinyUI(
  navbarPage()
)
```

server.r Script

```
library(shiny)

# Define la función donde se llevarán a cabo
##los procesos matemáticos y lógicos

shinyServer(function(input, output) {
})
```

La función *shinyUI()* dentro del script *ui.R* permite crear la interfaz que será visible al usuario, así como los valores entrada (inputs) que procesará el servidor y donde serán visibles los valores de salida (outputs). El parámetro *navbarPage()* permite crear múltiples páginas y una barra de navegación.

Del lado del servidor en el script *server.R* se declara la función *shinyServer(function(input, output) {})* la cual recibirá las entradas (inputs) desde el archivo *ui.R*, las procesará y creará valores de salida (outputs) que se desplegarán en la interfaz.

Dentro de la función *navbarPage()* se declaran separados por comas el título de la aplicación, la clave *id* para poder identificarlo fácilmente y las funciones *tabPanel()*. Las funciones *tabPanel()* crean cada una de las páginas, dentro de estas se declararán diferentes parámetros como el título de cada página. El llamado a estas nuevas funciones se describe a continuación:

ui.r Script

```

# Librerías que se utilizarán para el procesado de datos
library(shiny)
library(DT)
library(ggplot2)
library(plotly)

# Define la interfaz que y los datos de entrada
shinyUI(
  navbarPage(
    #Crea una barra de navegación
    "AquiParameter", #Nombre que aparecerá en la barra
    id="barra",      #id para identificar la barra

    tabPanel("Data"),      #Nueva página con nombre "data"
    tabPanel("Histograms") #Nueva página con nombre "histograms"

  )
)

```

server.r Script

```

# Librerías que se utilizarán para el procesado de datos
library(shiny)
library(DT)
library(ggplot2)
library(plotly)

# Función donde se llevarán a cabo procesos matemáticos y lógicos
shinyServer(function(input, output) {

})

```

Con esta modificación la app ahora tiene dos páginas “Data” e “Histogram”, en las cuales se puede navegar a partir de una barra de navegación en la parte superior de cada página.

Para crear una tabla de consulta con la paquetería *DT* e incorporarla a la aplicación se realizó el siguiente proceso: Primero se cargó la base de datos con el comando `read.csv()` con sus respectivos parámetros. Después desde el servidor *server.R* dentro de la función `shinyServer(){}` se crea una variable de salida `data` mediante el comando `output$data` en la cual guarda un objeto del tipo `renderDataTable()` cuyos argumentos son los mismos que se utilizan para crear una tabla con la paquetería *DT*.

Por el lado de la interfaz *ui.R* dentro de la función `tabPanel()` se declara el

nombre de la página con el comando `titlePanel()` y separado por una coma se declara `fluidRow()` que permite un diseño de página que consta de filas que a su vez incluyen columnas. Las columnas se declaran con la función `column()` y el parámetro `width=12` permite que la tabla ocupe toda la pantalla. Finalmente dentro de la función `dataTableOutput()` se declara el nombre de la variable de salida `data` creada en el servidor.

ui.r Script

```
# Librerías que se utilizarán para el procesamiento de datos
library(shiny)
library(DT)
library(ggplot2)
library(plotly)

#Carga la base de datos .csv
database<-read.csv("hydraulic_parameters.csv",header=TRUE,sep="," ,
                  fileEncoding = "UTF-8")

# Interfaz y los datos de entrada
shinyUI(
  navbarPage(
    "AquiParameter",
    id="barra",

    tabPanel("Data"
             titlePanel("Hydrogological Parameters Dataset"), #Titulo
             fluidRow(column(dataTableOutput("data"), #Variable de salida
                             width=12))
             ),
    tabPanel("Histograms")
  )
)
```

server.r Script

```
# Librerías que se utilizarán para el procesamiento de datos
library(shiny)
library(DT)
library(ggplot2)
library(plotly)

#Carga la base de datos .csv
database<-read.csv("hydraulic_parameters.csv",header=TRUE,sep="," ,
```

```

        fileEncoding = "UTF-8")

# Función donde se llevarán a cabo procesos matemáticos y lógicos
shinyServer(function(input, output) {

  # Crea la variable de salida "data" que se declara en el ui.R

  output$data<-renderDataTable(
    datatable(
      database,
      filter="top",
      selection="multiple",
      style="bootstrap"
    )
  )
})

```

Para crear páginas con variables de entrada para la creación de histogramas se siguió el siguiente proceso. Primero se creó dentro de la interfaz *ui.R* una función *sidebarLayout()* dentro del *tabPanel()* de la página *histograms*. La función *sidebarLayout()* permite crear una interfaz de página con un panel lateral *sidebarPanel()* y una página principal *mainPanel()* las cuales se declaran dentro de la función separados por comas. Dentro del *sidebarLayout()* se incluyen los siguientes botones:

- *radioButtons()*: El cual crea opciones de selección en la interfaz y regresa un único valor de tipo “character” seleccionado por el usuario, en este caso un valor con la propiedad que se quiere medir. Los argumentos de la función son los siguientes: un *id* para identificarlo y declararlo en el servidor; un nombre que será visible al usuario; un argumento *choices* que incluye un vector con los nombres de las columnas (propiedades) de la base de datos; y un argumento *selected*=“*K*” que indica el valor con el que se iniciará el proceso
- *selectInput()*: El cual crea opciones de selección en la interfaz y regresa como valores de entrada un vector de caracteres con las elecciones seleccionados por el usuario, en este caso litologías. Los argumentos separados por comas son los siguientes: un *id* para poder declararlo desde el servidor; un nombre que será visible al usuario; una lista *choices()* que incluye los valores únicos *unique* de la columna *Lithology* que el usuario podrá escoger, un argumento *multiple*=*TRUE* para que se pueden seleccionar varias elecciones; y un argumento *select*= “*Basalt*” para que la aplicación inicie con esa elección
- *checkboxInput()*: El cual crea una única opción en la interfaz y regresa como resultado un valor “booleano”, *TRUE* si está seleccionado y *FALSE*

si no está seleccionado. Los argumentos separados por comas son los siguientes; un *id* para que se pueda declarar desde el servidor, un nombre que será visible al usuario y un argumento *value=TRUE* para que al iniciar sesión la opción esté seleccionada

- *sliderInput()*: El cual crea un deslizador que regresa como valor de entrada un valor del tipo “numeric”. Los argumentos son los siguientes: un *id* para poder declararlo en el servidor, un nombre que será visible al usuario, valores mínimos y máximos *min* y *max* ; un argumento *value=20* que indica el valor en el cual se iniciará la aplicación; y un argumento *step=5* para indicar los saltos al momento de mover el deslizador

Para conocer todos los tipos de botones así como sus funciones y estructura se consultó el ejemplo de *Shiny Widgets Gallery* en la dirección (<https://shiny.rstudio.com/gallery/widget-gallery.html>)[<https://shiny.rstudio.com/gallery/widget-gallery.html>].

Del lado del servidor se creó una nueva variable llamada *properties* , la cual guarda un subgrupo creado a partir de los valores de entrada *input* seleccionados por el usuario mediante los botones. La variable *properties* se desarrolla a continuación:

```
# No compilar, ejemplo de extracción de datos
##a partir de las elecciones del ui, que se usará después

#Nuevo subgrupo creado con los "inputs" de los botones dentro de una
#función "reactive"
properties<-reactive({
  database[database$Lithology%in%input$lithology,
           c("Type","Lithology",input$property)]
})
```

En esta nueva variable se indica una función *reactive* la cual propicia que cada vez que el usuario seleccione una opción de entrada *input* , la función *properties()* volverá leer la información. La extracción de datos es mediante renglones y columnas, separados por una coma. Los renglones se extraen de la columna *Lithology* , la instrucción *%in%* indica si los valores coinciden con un valor, en este caso el vector creado con los valores de entrada del botón con id *lithology* (*input\$lithology*). Las columnas se declaran con un vector, las columnas de la base de datos a analizar son *Type* , *Lithology* y un valor de entrada del botón con id *property*.

Se declaró una variable de salida llamada *histogram* con la instrucción *output\$histogram* , esta variable guarda una función *renderPlotly()* , ya que el producto de salida será una gráfica interactiva con la paquetería *plotly*. Dentro de la función *renderPlotly()* es necesario crear una función lógica para crear diagramas con escala logarítmica o escala normal de acuerdo a la elección del botón con id *logx* (*input\$logx*) por parte del usuario. Los histogramas con *ggplot* y *plotly* se construyeron siguiendo el mismo principio del ejemplo de la

figura 3.9.

En la función `ggplot()` se declara como primer argumento el subgrupo creado y guardado en la variable `properties`, declarándose como una función `properties()`. Los valores del `aes()` incluyen la variable a graficar en `x`, para este caso la columna 3 del subgrupo creado `properties()`, y el comando `fill` que indica el color de relleno. También se añade el valor de entrada del botón con id `bins` (`input$bins` dentro de la función `geom_histogram()`).

```
#No compilar, ejemplo de la variable de salida histogram
# que se usara después
output$histogram<-renderPlotly({

  if(input$logx==TRUE){

    plot1<-ggplot(properties(),
                  aes(x=properties()[,3], fill=Lithology))+
    geom_histogram(bins=input$bins,na.rm=TRUE)+
    facet_wrap(~Lithology, ncol=2)+
    scale_x_log10()

    ggplotly(plot1,tooltip=c("y","Type"))

  }else{

    plot2<-ggplot(properties(),
                  aes(x=properties()[,3],fill=Lithology))+
    geom_histogram(bins=input$bins,na.rm=TRUE)+
    facet_wrap(~Lithology, ncol= 2)

    ggplotly(plot2,tooltip=c("y","Type"))

  }

})
```

Finalmente para desplegar el gráfico se declaró en el script `ui.R` dentro de la función `mainPanel()` la siguiente instrucción `plotlyOutput()` cuyo argumento es el nombre de la variable de salida declarada en el servidor. El ejemplo reproducible es el siguiente:

ui.r Script A

```

# Librerías que se utilizarán para el procesado de datos
library(shiny)
library(DT)
library(ggplot2)
library(plotly)

#Carga la base de datos .csv
database<-read.csv("hydraulic_parameters.csv",header=TRUE,sep=",",
                  fileEncoding = "UTF-8")

# Define la interfaz, así como los datos de entrada
shinyUI(
  navbarPage(
    "AquiParameter",
    id="barra",
    tabPanel("Data",
             titlePanel("Hydrogeological Parameters Dataset"),
             fluidRow(column(DT::dataTableOutput("data"),
                             width=12))

    ),

  tabPanel("Histograms",
           sidebarLayout(
             sidebarPanel(

               radioButtons("propertie",
                            "Select Hydraulic Propertie",
                            choices =c("Tporosity","Eporosity",
                                       "K","S",
                                       "Sy","Compressibility"),
                            selected="K"

               ),

               selectInput("lithology",
                           "Select Lithology",
                           choices=unique(database$Lithology),
                           multiple=TRUE,
                           selected="Basalt"

               ),

             )
  )
)

```



```

    database,
    filter="top",
    selection="multiple")
)

# Crea la variable de salida histogram
output$histogram<-renderPlotly({
  if(input$logx==TRUE){

    plot1<-ggplot(properties(),
                  aes(x=properties()[,3],fill=Lithology))+
    geom_histogram(bins=input$bins,na.rm=TRUE)+
    facet_wrap(~Lithology, ncol =2)+
    scale_x_log10()

    ggplotly(plot1,tooltip=c("y","Type"))

  }else{

    plot2<-ggplot(properties(),
                  aes(x=properties()[,3],fill=Lithology))+
    geom_histogram(bins=input$bins,na.rm=TRUE)+
    facet_wrap(~Lithology, ncol =2)

    ggplotly(plot2)

  }

})

})

```

Para crear los diagramas de caja se utilizó el mismo algoritmo que se siguió para crear los histogramas en el script antes mencionado. Para las funciones más avanzadas y aspectos estéticos se consultó la hoja de referencia de *ggplot* <https://rstudio.com/wp-content/uploads/2015/04/ggplot2-spanish.pdf> y las páginas oficiales de *plotly* <https://plotly.com/r/> y la página de *Shiny from Rstudio* <https://shiny.rstudio.com/tutorial/>.

Capítulo 4

Discusión y resultados

4.1 Base de datos: detalles y descripción

Se generó una base de datos con más de 6000 registros de las propiedades hidrogeológicas consideradas fundamentales para la caracterización hidrogeológica, de acuerdo con Jones et al. (2000, para diferentes tipos de litologías; para cada observación se presentan 11 columnas con la información de los parámetros hidrogeológicos, notas, locación y referencias. La base de datos se encuentra disponible en formato *.csv* delimitado por comas y puede consultarse y descargarse en la siguiente dirección: aquiparameters.rua.unam.mx, en el apartado *Downloads*.

La base de datos está diseñada para identificar rápidamente información de interés geológico asociada a cada una de las propiedades hidrogeológicas, por lo que existen etiquetas de litologías, además de información importante anexada en la base de datos, como comentarios de cada litología, ubicaciones y referencias.

Contiene valores de los siguientes parámetros hidrogeológicos: *Porosidad Total(%)*, *Porosidad Efectiva(%)*, *Conductividad Hidráulica (m/d)*, *Coefficiente de Almacenamiento (-)*, *Rendimiento Específico (-)* y *compresibilidad (m^2/N)*; Para más de 40 Tipos de Litologías y sedimentos no consolidados. Las litologías presentes en la base de datos son las siguientes:

## [1] "Andesite"	"Basalt"	"Dacite"	"Diorite"
## [5] "Dolerite"	"Gabbro"	"Granite"	"Granodiorite"
## [9] "Ignimbrite"	"Latite"	"Tuff"	"Scoria"
## [13] "Peridotite"	"Pumice"	"Monzonite"	"Rhyolite"
## [17] "Syenite"	"Amphibolite"	"Gneiss"	"Granulite"
## [21] "Marble"	"Migmatite"	"Phyllite"	"Quartzite"
## [25] "Schist"	"Serpentine"	"Scarn"	"Slate"

Tabla 4.1: Número de observaciones por propiedad hidrogeológica para cada tipo de litología

Type	total_porosity	effective_porosity	hydraulic_conductivity	storage_coefficient	specific_yield	compressibility
Igneous	240	233	491	173	254	20
Metamorphic	106	79	346	230	278	12
Sedimentary	432	338	692	282	321	49
Unconsolidated	242	235	634	190	319	109

```
## [29] "Anhydrite"           "Breccia"             "Chalk"               "Conglomerate"
## [33] "Dolomite"           "Gypsum"              "Halite"              "Limestone"
## [37] "Sandstone"          "Shale"               "Siltstone"           "Sand and Gravel"
## [41] "Gravel"             "Sand"                "Silt"                "Silt and Clay"
## [45] "Glacial till"       "Clay"                "Silty Sand"          "Clay Sand"
## [49] "Sand silt and gravel" "Sand clay and gravel" "Sand silt and clay"
```

Las definiciones de cada litología se consultaron en: <https://mrdata.usgs.gov/catalog/term-simple.php?thcode=4&code=root>.

La base de datos tiene una gran cantidad de datos. El número de datos de cada propiedad hidrogeológica para cada tipo de litología se presenta en la tabla 4.1. En esta tabla es posible apreciar que la mayoría de los datos recolectados fueron de mediciones de conductividad hidráulica, mientras que la menor cantidad de registrados corresponden con los valores de compresibilidad.

La tabla 4.2 muestra el número de datos de cada propiedad para cada litología en específico. En donde se puede apreciar que son las areniscas y las calizas las rocas de las que se tienen más registros de propiedades hidrogeológicas. Mientras que algunas rocas metamórficas y algunas rocas ígneas presentan la menor cantidad de datos.

4.2 Aplicación de consulta

Con el fin de publicar los resultados generados en este trabajo se creó una plataforma web interactiva, completamente programada con R. Esta plataforma tiene como objetivo principal ser una forma de consulta, interactiva, práctica e intuitiva, de la base de datos, así como proveer un rápido y fácil acceso a los datos crudos y a las referencias citadas. La página de web y puede consultarse en el siguiente enlace: aquiparameters.rua.unam.mx; el código fuente está disponible en la plataforma de github: https://github.com/Darcy1856/hydrogeologicalparameters_db

La aplicación tiene 6 apartados: *Data*, *Boxplot*, *Histograms*, *Correlation*, *Download* y *About*. A los cuales se puede acceder mediante la barra de navegación.

La página principal corresponde con el apartado *Data* en donde se puede interactuar con una tabla que incluye todos los valores de la base de datos y en la cual se pueden hacer búsquedas, aplicar filtros, así como ordenar los datos de menor a mayor o viceversa. Para hacer búsquedas se necesita colocar una

Tabla 4.2: Número de observaciones por propiedad hidrogeológica para cada litología

Lithology	total_porosity	effective_porosity	hydraulic_conductivity	storage_coefficient	specific_yield	compressibility
Amphibolite	1	8	19	19	11	2
Andesite	14	8	17	10	7	1
Anhydrite	6	6	7	2	0	4
Basalt	59	37	162	31	81	7
Breccia	0	5	4	2	0	0
Chalk	45	38	117	61	61	4
Clay	37	21	91	28	37	35
Clay Sand	3	4	25	19	8	3
Conglomerate	9	4	23	9	11	0
Dacite	7	5	4	0	2	1
Diorite	7	3	7	4	7	0
Dolerite	6	4	23	3	8	2
Dolomite	33	71	102	19	44	3
Gabbro	11	4	35	12	8	2
Glacial till	23	10	64	7	13	2
Gneiss	35	23	116	87	99	3
Granite	73	86	161	79	92	6
Granodiorite	18	20	7	1	1	0
Granulite	3	5	10	10	1	3
Gravel	30	18	44	8	25	8
Gypsum	20	21	14	1	5	4
Halite	4	3	4	3	1	2
Ignimbrite	4	28	15	8	11	0
Latite	2	0	0	0	0	0
Limestone	112	101	160	70	107	8
Marble	12	0	21	8	13	0
Migmatite	0	0	16	10	11	0
Monzonite	2	0	0	0	0	0
Peridotite	0	0	7	1	2	0
Phyllite	3	4	19	13	29	2
Pumice	2	0	0	0	0	0
Quartzite	17	11	40	24	53	0
Rhyolite	10	6	17	7	10	1
Sand	63	85	195	59	116	33
Sand and Gravel	26	41	98	33	56	9
Sand clay and gravel	1	0	0	0	1	1
Sand silt and clay	15	18	28	2	21	0
Sand silt and gravel	1	6	8	4	2	1
Sandstone	138	77	188	100	72	17
Scarn	2	0	0	0	0	0
Schist	22	22	86	52	49	2
Scoria	2	2	0	0	2	0
Serpentine	2	2	0	2	0	0
Shale	40	8	44	9	12	3
Silt	27	22	47	14	21	6
Silt and Clay	13	6	23	9	13	10
Siltstone	25	4	29	6	8	4
Silty Sand	3	4	11	7	6	1
Slate	9	4	19	5	12	0
Syenite	1	1	0	0	0	0
Tuff	22	29	36	17	23	0

Hydrogeological Parameters Dataset

Show 10 entries

Search:

Type	Litology	Φ T (%)	ϕ e (%)	K (m/d)	S (-)	Sy (-)	α (Pa-1)	Note	Location	Source
1	Igneous Andesite	0.5						NS	Doenstedt, Germany	77
2	Igneous Andesite			3.6	0.000923			Fractured	Taipei basin, Taiwan	78
3	Igneous Andesite			8.38	0.00125			Fractured	Taipei basin, Taiwan	78
4	Igneous Andesite			6.35	0.000529			Fractured	Taipei basin, Taiwan	78
5	Igneous Andesite			1.91151	0.00266			Fractured	Taipei basin, Taiwan	78
6	Igneous Andesite			3.29	0.000778			Fractured	Taipei basin, Taiwan	78
7	Igneous Andesite				0.000301			NS	Taipei basin, Taiwan	78
8	Igneous Andesite			0.000864				Dense	Zacapu basin, Michoacán	111
9	Igneous Andesite			8.64				Fractured	Zacapu basin, Michoacán	111
10	Igneous Andesite		0.27					Fractured	LaLumbredebris-	113

Figura 4.1: Página principal de la plataforma web (Apartado "Data")

palabra dentro de la barra *search* ubicada en la parte superior izquierda de la base de datos.

Para filtrar los datos se necesita colocar un valor o un rango de valores en las casillas colocadas en la parte superior de cada una de las columnas de la tabla. De la misma manera para ordenar los valores de menor a mayor es necesario hacer *click* en las flechas colocadas al lado de los nombres de las columnas en la tabla.

El apartado *Boxplot* permite visualizar los datos mediante diagramas de caja con el fin de apreciar valores mínimos, máximos, medianas y cuartiles de los datos. Se pueden visualizar seleccionando diversas opciones en el panel lateral izquierdo. Las opciones incluyen: el parámetro a visualizar, la escala (logarítmica o continua), el tipo de litología; o bien hacer una búsqueda de una litología específica con el buscador ubicado en la parte inferior izquierda. En esta página también se presenta una tabla con los valores puntuales mínimos, máximos y los cuartiles, que muestra los datos de la litología o las litologías seleccionadas.

La página *Histogram* permite el acceso a los histogramas creados a partir de la base de datos. Los histogramas se muestran de acuerdo con las selecciones del panel lateral izquierdo por parte del usuario, estas selecciones incluyen: el parámetro a visualizar, la escala (logarítmica o continua), el tipo de litología; o bien hacer una búsqueda de una litología específica con el buscador ubicado en la parte inferior izquierda. En este apartado también es posible visualizar una tabla con medidas de tendencia central y medidas de dispersión.

El apartado *Correlation* da acceso a los diagramas de dispersión de los datos. Estos diagramas se crean a partir de las selecciones del usuario en el panel lateral izquierdo. Las opciones son las siguientes: los parámetros a visualizar y la escala (logarítmica o continua) en el eje x y en el eje y.

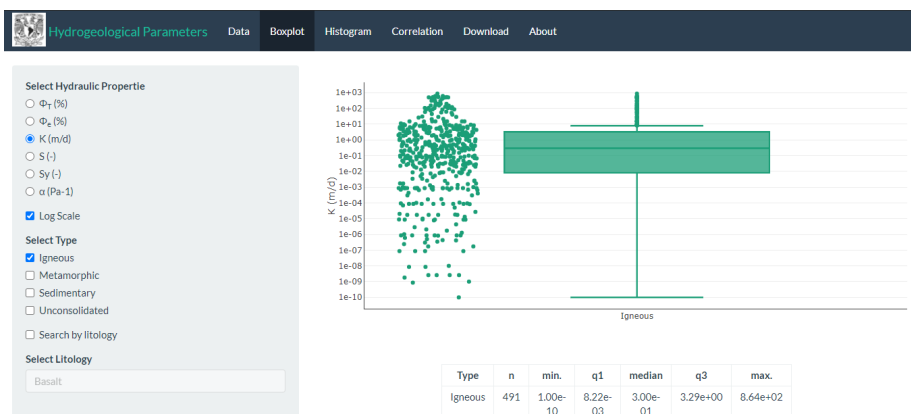


Figura 4.2: Apartado "Boxplot" de la plataforma

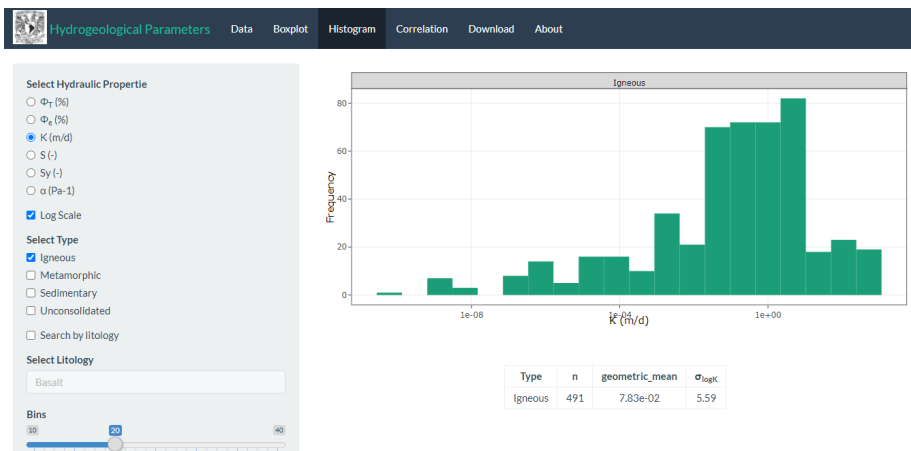


Figura 4.3: Apartado "Histogram" de la plataforma

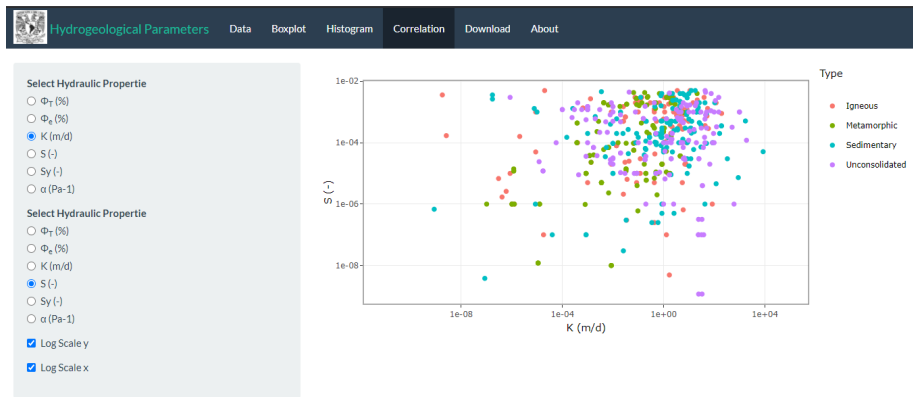


Figura 4.4: Apartado "Correlation" de la plataforma

La página *Download* da acceso a los enlaces de descarga de la base de datos en formato (.csv) delimitado por comas y a las referencias citadas en formato (.pdf).

	Type	Litology	ΦT (%)	Φe (%)	K (m/d)	S(-)	Sy(-)	α (Pa-1)	Note	Location	Source
1	Igneous	Andesite	0.5						NS	Doenstedt_Germany.	77
2	Igneous	Andesite			3.6	0.000923			Fractured	Taipei basin_Taiwan	78
3	Igneous	Andesite			8.38	0.00125			Fractured	Taipei basin_Taiwan	78
4	Igneous	Andesite			6.35	0.000529			Fractured	Taipei basin_Taiwan	78
5	Igneous	Andesite			1.91151	0.00266			Fractured	Taipei basin_Taiwan	78
6	Igneous	Andesite			3.29	0.000778			Fractured	Taipei basin_Taiwan	78
7	Igneous	Andesite				0.000301			NS	Taipei basin_Taiwan	78
8	Igneous	Andesite			0.000864				Dense	Zacapu basin_Michoacán	111
9	Igneous	Andesite			8.64				Fractured	Zacapu basin_Michoacán	111
10	Igneous	Andesite	0.27						Fractured	LaLumbredebris-	113

Figura 4.5: Apartado "Download" de la plataforma

El apartado *About* tiene la información referente al trabajo: autor, desarrollador, director, y cita.

4.3 Análisis estadístico

A continuación, se presenta la estadística descriptiva, a partir de la base de datos, mediante gráficos y medidas numéricas con el fin de analizar la distribución,

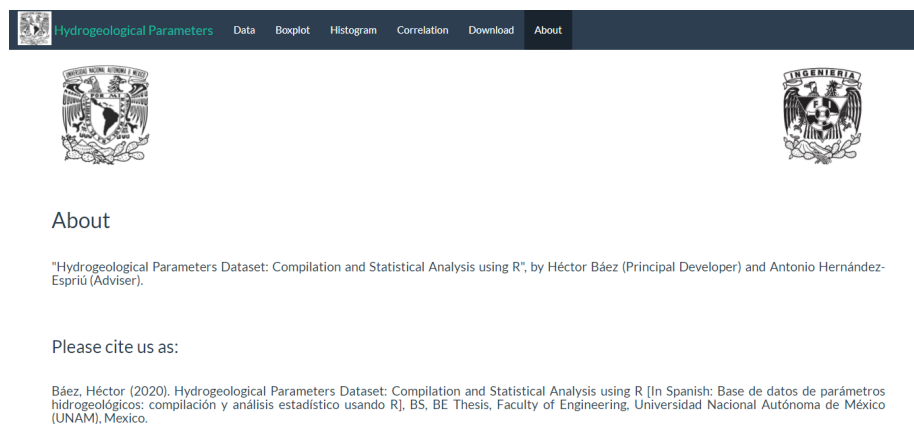


Figura 4.6: Apartado "About" de la plataforma

datos anómalos, medidas de tendencia central y medidas de dispersión. Para facilitar la visualización, se crearon subgrupos para el análisis.

4.3.1 Porosidad Total

El rango de porosidad total para las diferentes clasificaciones litológicas va desde valores cercanos a 0 hasta alcanzar estimaciones muy elevadas, más del 30% para todos los tipos de litologías fig. 4.7. La existencia de valores tan elevados de porosidad se debe a la presencia de valores anómalos dentro de la muestra.

Los valores de porosidad total en rocas ígneas resultan ser bastante elevados, superando en un caso extremo más del 70% de porosidad total. Sin embargo, desde los trabajos de Davis and DeWiest (1966) se ha registra que el intemperismo en depósitos volcánicos puede aumentar la porosidad en valores extremos de hasta un 60%.; existen litologías como las tobas o el basalto, cuya porosidad es primaria y se genera por una alta cantidad de gases en el magma del cual se generan estas rocas. En un caso extremo la toba alcanza hasta un 87%.

Analizando más a fondo el comportamiento del rango de la porosidad total, la figura 4.8 muestra los rangos para las litologías más representativas, siendo para las rocas ígneas, la variedad de tobas, un ejemplo de rocas con una porosidad anómala en términos estadísticos. Otro ejemplo es la roca "pumita", la cual no se muestra en los gráficos debido a la escasez de datos, pero que puede verificarse en la base de datos; teniendo un registro de hasta el 75% de porosidad.

Dentro de las rocas metamórficas es el esquisto intemperizado el que presenta valores atípicos de porosidad total, alcanzando hasta un 38% de porosidad. Mientras que el resto de las observaciones muestran que este tipo de rocas no supera el 5% de porosidad.

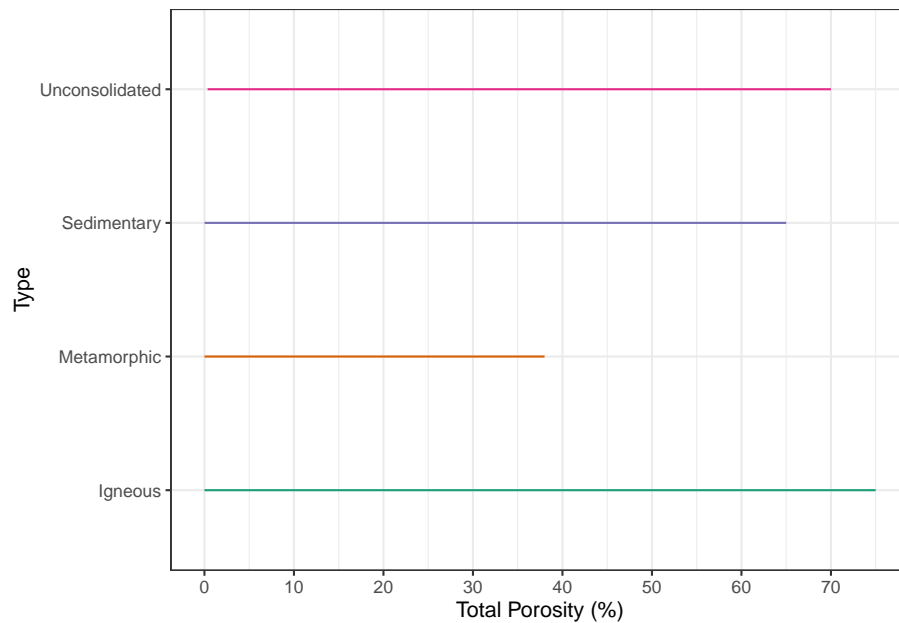


Figura 4.7: Diagrama de rangos de la porosidad total (%) para los diferentes tipos de litología.

A su vez las rocas sedimentarias de mayor rango corresponden con los carbonatos con más del 60% de porosidad total. Los sedimentos no consolidados también muestran rangos que alcanzan hasta un 70% de porosidad por la presencia de valores extremos principalmente en limos y arcillas.

En el diagrama 4.9 se muestran los diagramas de caja donde puede identificar valores anómalos para las rocas ígneas y metamórficas, fracturadas. Por lo cual se puede definir un nuevo rango de valores para este tipo de rocas, las cuales no exceden el 25% de porosidad para el caso de rocas ígneas y 10% para rocas metamórficas sin la presencia de valores anómalos.

Las rocas sedimentarias y los sedimentos no consolidados no presentan valores anómalos por lo que sus valores de porosidad varían desde valores cercanos a cero hasta casi el 70% de porosidad en sedimentos no consolidados. En la tabla 4.3 se muestran los diferentes valores mínimos y máximos, así como los cuartiles: Q1, Q2 (Mediana) y Q3, para las diferentes clasificaciones litológicas.

En la tabla 4.4 se observan las medidas de tendencia central para la porosidad total (%). Se muestra el siguiente comportamiento: en rocas ígneas y metamórficas la mediana y la media geométrica varían en menor proporción una de otra a diferencia de la media aritmética que difiere en mayor cantidad de las otras dos medidas de tendencia central; en sedimentos no consolidados

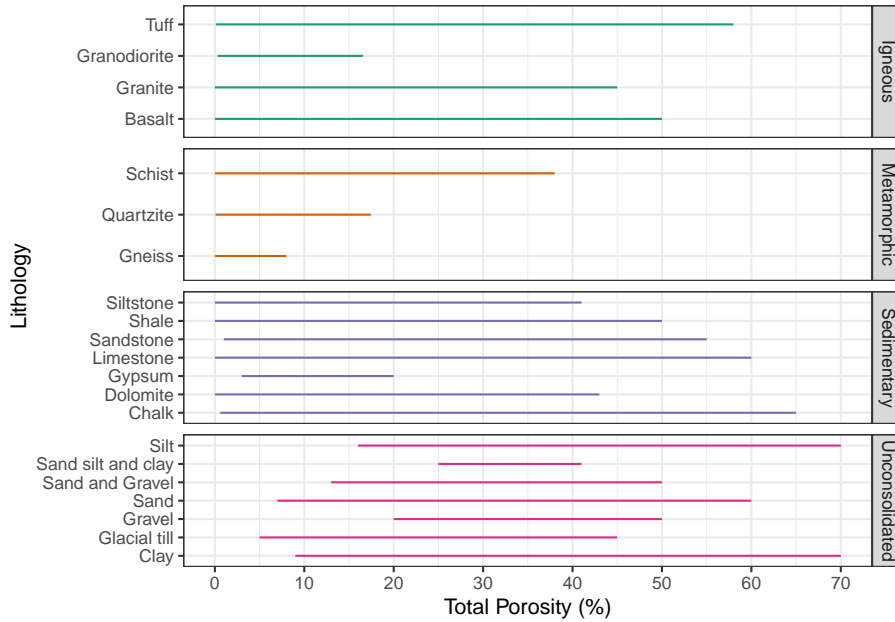


Figura 4.8: Diagrama de rangos de la porosidad total (%) para diferentes litologías.

Tabla 4.3: Valores mínimos, máximos y cuartiles de la porosidad total (%)

Tipo	No.Datos	min.	Q1	Mediana	Q3	max.
Igneous	240	0.0001	0.542	2.175	10.000	75
Metamorphic	106	0.0001	0.103	0.700	2.725	38
Sedimentary	432	0.0100	5.075	15.000	30.000	65
Unconsolidated	242	0.3500	25.000	35.000	44.000	70

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la porosidad total (%) para las diferentes clasificaciones litológicas. La columna **No. datos** corresponde con el número de observaciones

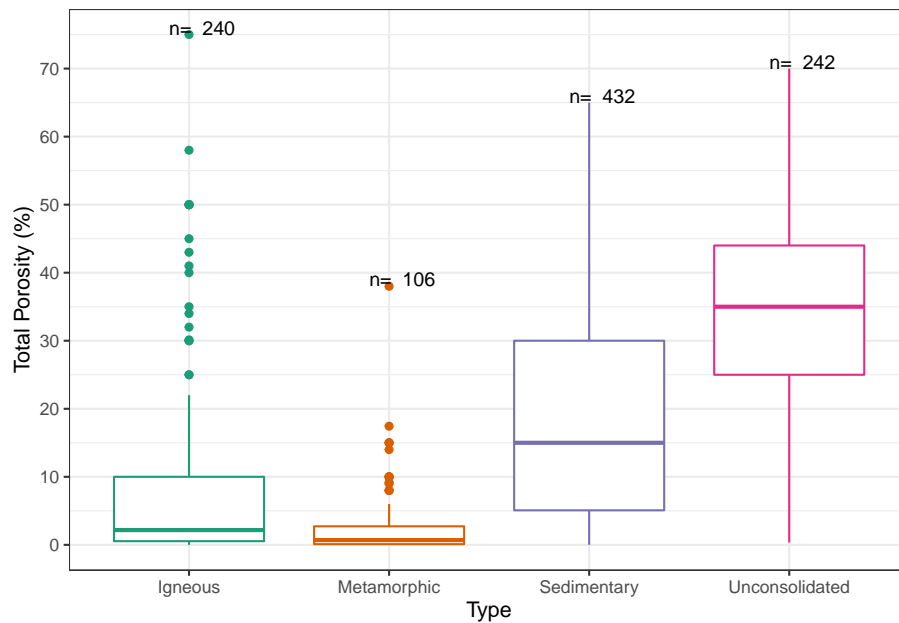


Figura 4.9: Diagramas de caja de la porosidad total (%) para las diferentes clasificaciones litológicas.

Tabla 4.4: Comparación de las medidas de tendencia central de la porosidad total (%)

Tipo	Media	Media geométrica	Mediana
Igneous	7.389	1.796	2.175
Metamorphic	2.810	0.549	0.700
Sedimentary	17.947	9.131	15.000
Unconsolidated	35.251	31.745	35.000

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la porosidad total (%) de los diferentes tipos de litologías.

la media aritmética, la mediana y la media geométrica se diferencian en menor grado una de otra; mientras que en rocas sedimentarias la mediana difiere en menor cantidad de la media aritmética que de la media geométrica.

La tabla 4.5 muestra las medidas de dispersión de la porosidad total (%) para las diferentes clasificaciones litológicas. Analizando la desviación estándar se muestra que las rocas sedimentarias y sedimentos no consolidados presentan una mayor dispersión entre sus datos, mientras que las rocas metamórficas son

Tabla 4.5: Medidas de dispersión de la porosidad total (%)

Tipo	Rango	IQR	σ	DM	DMd	DMg
Igneous	75.00	9.457	11.755	7.639	6.700	6.724
Metamorphic	38.00	2.622	5.207	3.239	2.600	2.606
Sedimentary	64.99	24.925	14.173	11.968	11.701	12.557
Unconsolidated	69.65	19.000	13.139	10.323	10.294	10.859

Nota:

Tabla de comparación de las medidas de dispersión de la porosidad total (%) en los diferentes tipos de litologías. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

las de menor dispersión.

Las rocas ígneas, sedimentarias y metamórficas tienen una menor desviación respecto a su mediana entre sus datos, en comparación con sus respectivas desviaciones respecto a la media aritmética y la media geométrica. En sedimentos no consolidados las tres desviaciones medias absolutas tienden a ser parecidas, no obstante, se puede concluir que sus datos tienen una menor desviación respecto a la mediana.

La gran dispersión que existe entre los datos se debe a que diferentes rocas con diferentes propiedades se clasifican dentro de un mismo tipo de litología.

El análisis de la fig.4.10 muestra que todos los histogramas son unimodales para todos los casos. En rocas ígneas y metamórficas el sesgo es positivo con una forma leptocúrtica; sesgado hacia valores cercanos a cero, en donde las mayores frecuencias se presentan en rangos menores al 2% de porosidad total en rocas ígneas y metamórficas.

Las rocas sedimentarias también tienen una tendencia hacia un sesgo positivo y forma leptocúrtica. Además, tienen una mayor frecuencia en valores cercanos al 5% de porosidad total, sin embargo, estimaciones cercanas al 28% también tienen un alta frecuencias por lo que podría considerarse como otra moda.

El histograma de los sedimentos no consolidados se diferencia mucho del resto de histogramas y muestra un histograma simétrico y mesocúrtico con altas frecuencias para un valor de 38% de porosidad total.

Correlaciones Porosidad Total

En la figura 4.11 se presentan los diagramas de correlación de la porosidad total en función de cada una de las propiedades hidrogeológicas (porosidad efectiva,

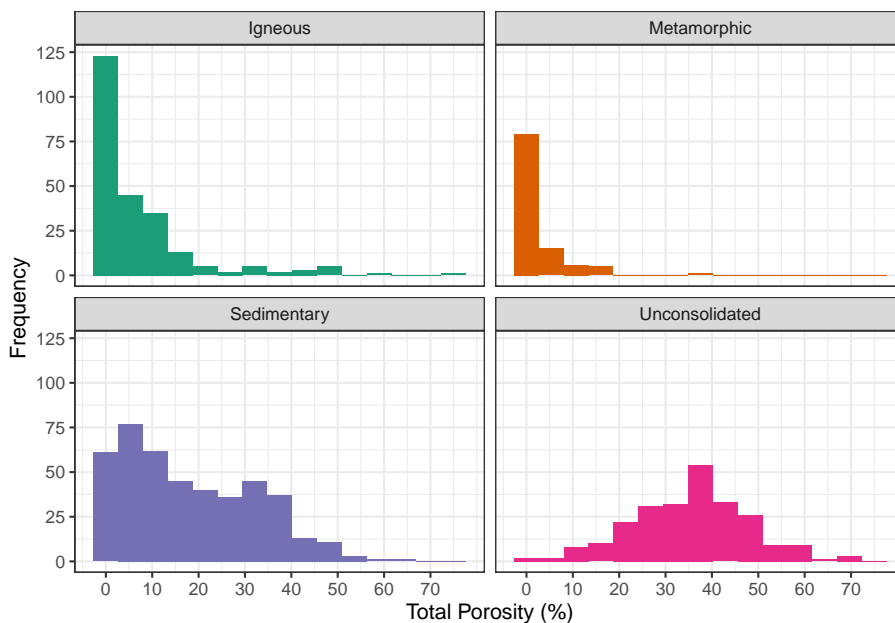


Figura 4.10: Histogramas de la porosidad total (%) para las diferentes clasificaciones litológicas. Histogramas construidos a partir de 15 clases

conductividad hidráulica, coeficiente de almacenamiento, rendimiento específico y compresibilidad).

Analizando los diagramas y el coeficiente de determinación se puede concluir que los casos de los gráficos B, C, D, E no se ajustan a un modelo lineal, aun aplicando métodos linealizables mediante transformaciones logarítmicas de los datos. Por lo que se puede concluir que no hay un modelo de regresión lineal que explique un porcentaje importante de la proporción de la varianza de la variable dependiente en estos casos.

En el caso del diagrama A (porosidad total vs porosidad efectiva) se tiene un coeficiente de determinación bastante alto en comparación con los otros modelos expuestos en la figura de correlaciones, lo que indica una mayor proporción de varianza de la variable dependiente explicada por el modelo.

4.3.1.1 Rocas ígneas

La fig. 4.12 muestra los diagramas de caja de la porosidad total (%) de algunas rocas ígneas, siendo la toba la que difiere por mucho de las demás y por lo tanto la roca que genera los valores anómalos cuando se hace un análisis de todas las rocas ígneas, con estimaciones que alcanzan hasta el 80 % de porosidad total.

Como ya se mencionó con anterioridad los trabajos de Davis and DeWiest (1966)

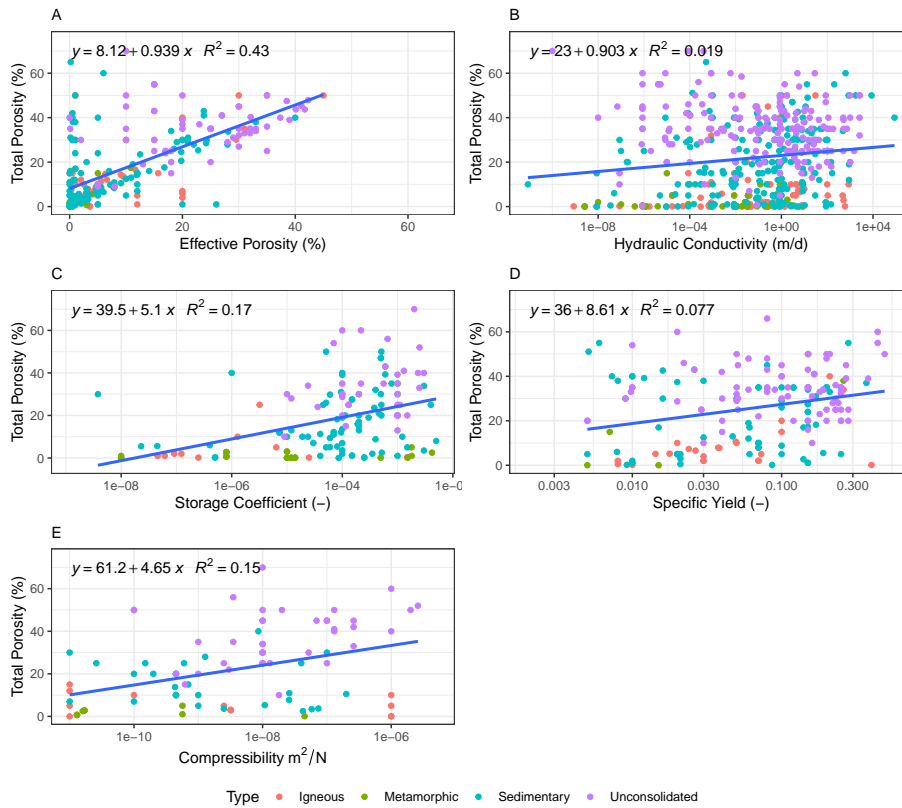


Figura 4.11: Diagramas de correlación de la porosidad total en función de las diferentes propiedades hidrogeológicas. A: Porosidad total vs Porosidad Efectiva. B: Porosidad Total vs Conductividad Hidráulica (transformación logarítmica en base 10 del eje x). C: Porosidad Total vs Coeficiente de Almacenamiento (transformación logarítmica en base 10 del eje x). D: Porosidad Total vs Rendimiento Específico (transformación logarítmica en base 10 del eje x). E: Porosidad Total vs Compresibilidad (transformación logarítmica en base 10 del eje x)

muestran la existencia de registros en litologías como toba y la pumita con valores de porosidad que superan el 60 % de porosidad total; el término de toba es muy general y puede englobar una variedad de litologías, lo cual puede ser una explicación para el amplio rango de valores. Para el basalto, aquellas estimaciones mayores al 20% son anómalas, mientras que para el granito son aquellas mayores al 8%; otra roca registrada es la granodiorita, la cual no presenta valores anómalos y llega a alcanzar hasta el 15% de porosidad.

El elevado valor de porosidad total (%) en rocas ígneas se debe a la presencia de rocas muy fracturas e intemperizadas. Por ejemplo, los casos del granito y granodiorita intemperizados expuestos por Brassington (2007) con porosidades

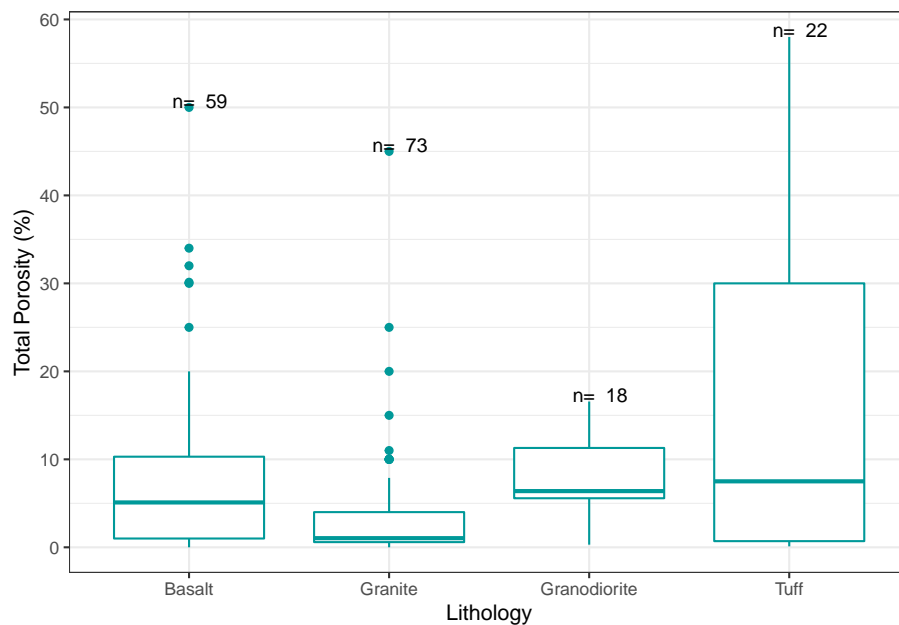


Figura 4.12: Diagramas de caja de la porosidad total (%) en diferentes rocas ígneas.

Analizando la tabla 4.6 se puede inferir que el basalto es la roca que registra un mayor valor de la mediana con 7.5 %, mientras que el granito tiene el menor valor de la mediana con 1.04 %. La toba registra los mayores valores máximos con 58%, seguido por el basalto con 50 %

Las medidas de tendencia central se registran en la tabla 4.7. En rocas como el basalto, granito y granodiorita, la mediana y la media geométrica se diferencian en menor cantidad, mientras que la media aritmética difiere en mayor grado de las otras dos medidas de tendencia central. En la toba las tres medidas de tendencia central tienden a ser muy diferentes unas de otras.

La tabla 4.8 muestra las medidas de dispersión de la porosidad total en relación

Tabla 4.6: Valores mínimos, máximos y cuartiles de la porosidad total (%)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Basalt	59	0.0010	1.000	5.10	10.300	50.00
Granite	73	0.0001	0.590	1.04	4.000	45.00
Granodiorite	18	0.3000	5.578	6.39	11.293	16.57
Tuff	22	0.1100	0.703	7.50	30.000	58.00

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la Porosidad Total (%) en diferentes rocas ígneas. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.7: Comparación de las medidas de tendencia central de la porosidad total (%)

Litología	Media	Media_geométrica	Mediana
Basalt	8.669	2.433	5.10
Granite	3.643	1.013	1.04
Granodiorite	7.710	5.215	6.39
Tuff	15.738	3.911	7.50

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la porosidad total (%) en diferentes rocas ígneas.

con las diferentes rocas ígneas. Analizando la desviación estándar se puede concluir que los datos de porosidad total (%) en las tobas están muy dispersos, mientras que en la granodiorita esta dispersión es menor. En todos los casos la desviación de los datos de porosidad total respecto a la mediana son menores que si se comparan con las desviaciones respecto a la media geométrica y la media aritmética.

La gran dispersión que existe entre los datos de porosidad total de las diferentes rocas ígneas se debe a la existencia de rocas fracturadas y rocas intemperizadas, así como rocas inalteradas dentro de un mismo grupo.

Los histogramas de la fig. 4.13 muestran el comportamiento de la distribución de los datos para rocas ígneas. Puede interpretarse que para todos los casos se tiene un histograma con sesgo positivo y forma leptocúrtica, también puede concluirse unimodalidad para los cuatro casos.

Para el basalto, el granito y la toba las mayores frecuencias de datos de porosidad total, se concentran en estimaciones menores al 1%, por su parte la granodiorita, tiene una mayor frecuencia en datos cercanos al 5% de porosidad total.

Tabla 4.8: Medidas de dispersión de la porosidad total (%)

Litología	Rango	IQR	σ	DM	DMd	DMg
Basalt	49.999	9.300	10.182	7.284	6.818	7.492
Granite	45.000	3.410	6.758	4.013	3.116	3.117
Granodiorite	16.270	5.715	4.825	3.857	3.681	4.072
Tuff	57.890	29.298	19.765	16.325	14.625	14.724

Nota:

Tabla de comparación de las medidas de dispersión de la porosidad total (%) en diferentes rocas ígneas. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (Dmg).

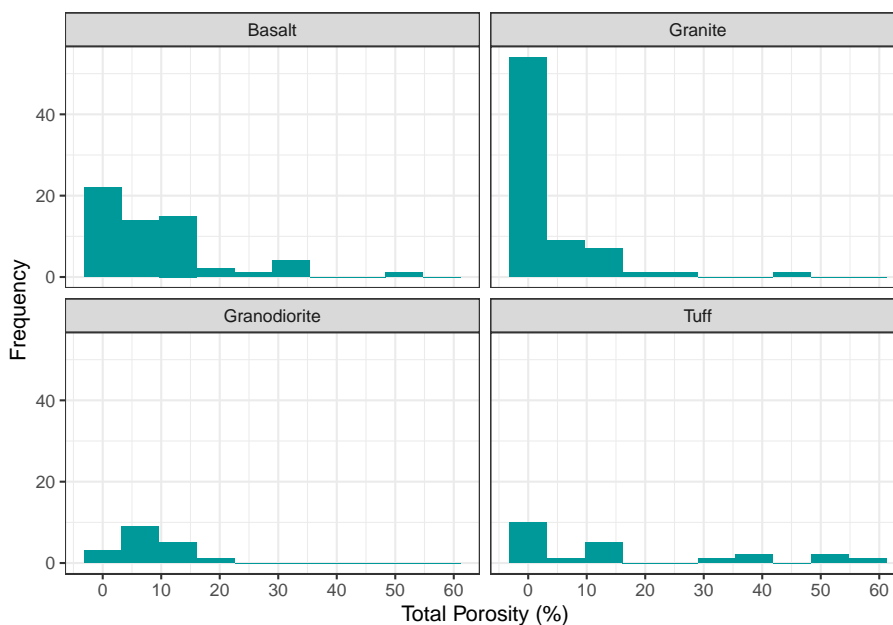


Figura 4.13: Histogramas de la porosidad total (%) en diferentes rocas ígneas. Histogramas construidos a partir de 10 clases

4.3.1.2 Rocas Metamórficas

Las rocas metamórficas presentan las menores porosidades totales, incluso cuando una gran cantidad de los datos están representados por rocas fracturadas y/o intemperizadas. En la figura 4.14 se muestran los diagramas de caja de algunas rocas metamórficas, siendo el esquisto el que presenta valores atípicos máximos, alcanzando hasta un 38% de porosidad en un caso anómalo. El gneis y la cuarcita muestran un comportamiento en el cual las porosidades no superan el 18% ni siquiera en casos extremos.

La presencia de valores anómalos en el gráfico se debe a la existencia de rocas extremadamente fracturadas e intemperizadas que forman parte de la muestra. Por ejemplo, Brassington (2007) registra una porosidad total del 38% en esquistos intemperizados, otro ejemplo son las cuarcitas intemperizadas de Delhi, India expuestas por Gupta and Rao (2001) con porosidades superiores al 15%.

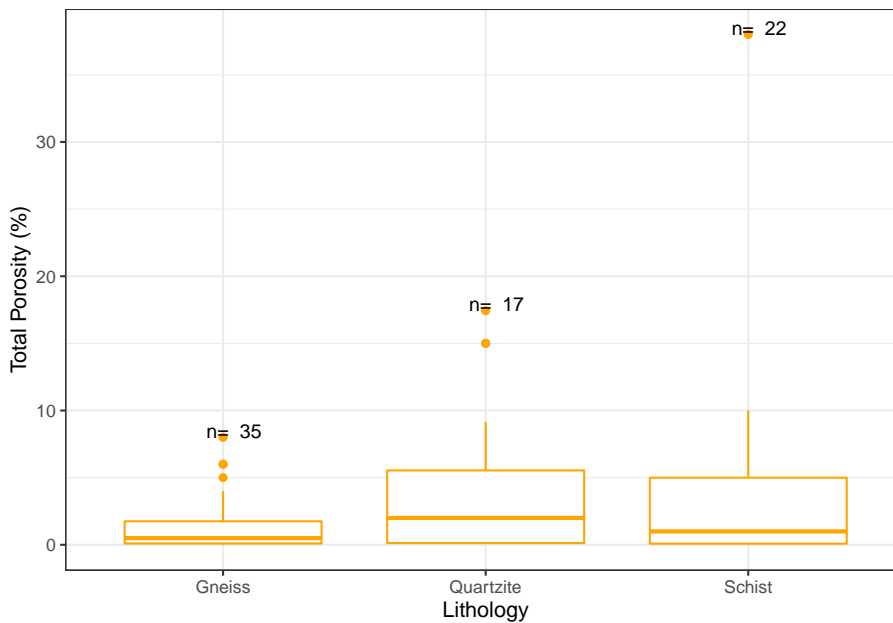


Figura 4.14: Diagramas de caja de la porosidad total (%) en diferentes rocas metamórficas.

En la tabla 4.9 se muestran los cuartiles y los valores máximos de algunas rocas metamórficas. En la tabla se puede apreciar que el gneis tiene la mediana más baja, seguido por el esquisto y finalmente la cuarcita tiene la mediana más alta.

La tabla 4.10 muestra la comparación de las diferentes medidas de tendencia central que se obtuvieron para cada roca metamórfica. La media aritmética resulta ser para los tres casos la que más difiere de las tres medias ya que varía

Tabla 4.9: Valores mínimos, máximos y cuartiles de la porosidad total (%)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Gneiss	35	1e-04	0.100	0.5	1.75	8.00
Quartzite	17	8e-02	0.130	2.0	5.54	17.44
Schist	22	1e-03	0.085	1.0	5.00	38.00

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la Porosidad Total (%) en diferentes rocas metamórficas. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.10: Comparación de las medidas de tendencia central de la porosidad total (%)

Litología	Media	Media_geométrica	Mediana
Gneiss	1.372	0.341	0.5
Quartzite	4.443	1.230	2.0
Schist	4.219	0.477	1.0

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la porosidad total (%) en diferentes rocas metamórficas.

en más de una unidad en los tres tipos de rocas. En la cuarcita y el esquisto el valor de la media geométrica resulta ser parecido a las otras dos medidas de tendencia central.

La tabla 4.11 muestra la comparación entre diferentes medidas de dispersión de la porosidad total (%) para diferentes rocas metamórficas. En esta tabla se muestra un valor alto de la desviación estándar para el esquisto, en comparación con las otras tres rocas lo que muestra una gran dispersión entre los datos, mientras que el gneis presenta una menor dispersión. Las distintas desviaciones medias absolutas de un punto central resultan ser muy similares en el caso del gneis. En la cuarcita la desviación respecto a la media aritmética y la desviación respecto a la media geométrica son más parecidas y son mayores que la desviación respecto a la mediana. Finalmente, para el esquisto la desviación respecto a la media aritmética es mucho mayor, mientras que se presenta una menor desviación de los datos respecto a la mediana.

La existencia de grandes valores de desviación se debe a la presencia de diferentes grados de fracturamiento e intemperismo en los diferentes valores registrados.

En la figura 4.15 se muestran los diferentes histogramas de la porosidad total

Tabla 4.11: Medidas de dispersión de la porosidad total (%)

Litología	Rango	IQR	σ	DM	DMd	DMg
Gneiss	8.000	1.650	2.050	1.524	1.219	1.226
Quartzite	17.360	5.410	5.932	4.697	3.926	4.059
Schist	37.999	4.915	8.374	5.133	3.963	4.056

Nota:

Tabla de comparación de las medidas de dispersión de la porosidad total (%) en diferentes rocas metamórficas. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (Dmg).

(%) de algunas rocas metamórficas. En esta figura se aprecia una gráfica con sesgo positivo en las distribuciones, así como una forma platicúrtica en los 3 casos; otro rasgo característico es que se tiene unimodalidad. En todos casos de rocas metamórficas expuestos se observa que las mayores frecuencias de datos de porosidad total se concentran en valores entre el 0 y el 1%.

4.3.1.3 Rocas Sedimentarias

Las rocas sedimentarias varían desde valores cercanos a cero hasta alcanzar porosidades importantes, en la mayoría de los casos superando el 40% de porosidad total (figura 4.16). La lutita y la caliza son las únicas rocas que presentan valores anómalos de porosidad total (%), siendo en lutitas los valores mayores al 40% considerados anómalos y en calizas los valores mayores al 60%.

Las rocas carbonatadas y las areniscas son las que presentan una mayor porosidad total, llegando a alcanzar hasta un 60% en el caso de los carbonatos y hasta un 50% para areniscas. Valores muy altos de porosidad total se registran, por ejemplo, en los carbonatos del norte de Europa donde Fretwell et al. (2005) registra hasta un 65% de porosidad total. Las rocas de grano fino como las lutitas y las limolitas registran menores porosidades totales; cercanas al 40% y en casos anómalos hasta el 50%.

En la tabla 4.12 se muestra el resumen de los diagramas de caja expuestos en la figura 4.16. La tabla nos permite comparar los valores puntuales de las medianas indicando que la roca que mayor mediana presenta es la creta, mientras que el yeso registra un menor valor de la mediana en comparación con las otras rocas.

En la tabla 4.13 se hace una comparación de diferentes medidas de tendencia central de la porosidad total en las diferentes rocas sedimentarias. Se puede concluir que, en rocas como la creta, caliza, dolomita y la limolita, las tres medidas de tendencia central resultan ser bastante diferentes entre sí, siendo la

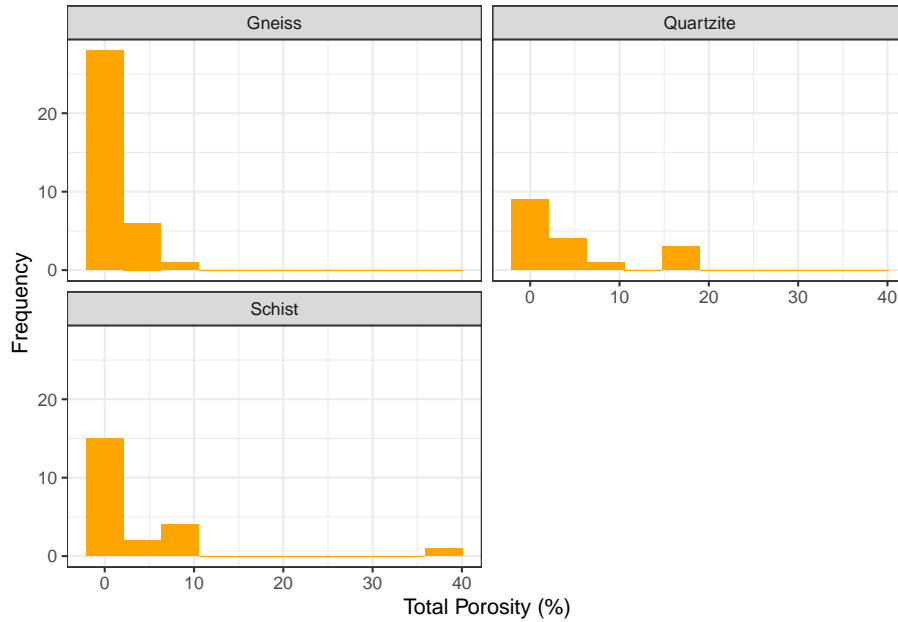


Figura 4.15: Histogramas de la porosidad total (%) en diferentes rocas metamórficas. Histogramas construidos a partir de 10 clases

Tabla 4.12: Valores mínimos, máximos y cuartiles de la porosidad total (%)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Chalk	45	0.60	15.000	31.00	40.000	65
Dolomite	33	0.01	1.000	10.00	25.000	43
Gypsum	20	3.00	5.600	8.20	13.300	20
Limestone	112	0.01	3.900	12.35	25.625	60
Sandstone	138	1.00	11.000	20.00	30.000	55
Shale	40	0.01	0.645	10.00	16.475	50
Siltstone	25	0.01	7.000	11.50	25.000	41

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la Porosidad Total (%) en diferentes rocas sedimentarias. La columna **No.datos** corresponde con el número de observaciones.

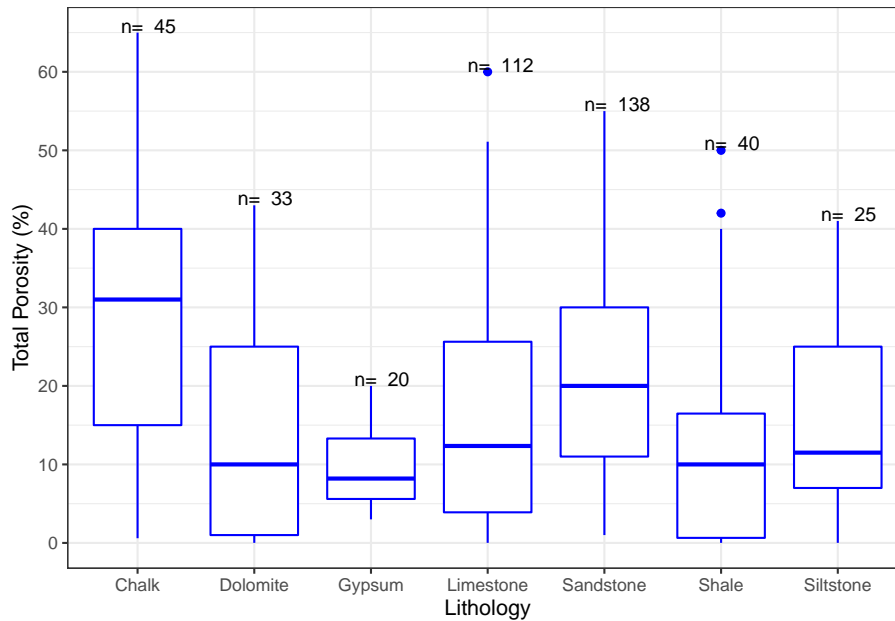


Figura 4.16: Diagramas de caja de la porosidad total (%) en diferentes rocas sedimentarias.

media geométrica la que se diferencia en mayor grado de las medias aritmética y mediana. En el caso del yeso, la media geométrica resulta ser más parecida con la mediana. En areniscas y lutitas la media aritmética y la mediana son más parecidas entre sí, que con la media geométrica.

En la tabla 4.14 se muestran las comparaciones de las distintas medidas de dispersión de los datos de porosidad total (%) para rocas sedimentarias. Se puede concluir a partir del análisis de la desviación estándar que el yeso es la roca que presenta una menor dispersión entre sus datos, mientras que por el contrario los carbonatos con la creta en primer lugar son las rocas que tienen una mayor dispersión entre sus datos de porosidad total.

En rocas como la dolomita, creta, yeso y arenisca las tres desviaciones medias absolutas respecto a un punto central resultan ser tan parecidas que no varían ni en una unidad, pero puede inferirse que existe una menor desviación de sus datos respecto a la mediana. En lutitas y limolitas es más claro que la desviación de los datos respecto a la mediana es menor que las desviaciones respecto a la media aritmética y la media geométrica. Finalmente, la caliza tiene una menor desviación de sus datos respecto a la media geométrica en comparación con las desviaciones respecto a la media aritmética y la mediana.

Los valores altos de dispersión entre los datos de la porosidad total se deben a la existencia de diferentes texturas y estructuras que predominan en cada tipo de

Tabla 4.13: Comparación de las medidas de tendencia central de la porosidad total (%)

Litología	Media	Media_geométrica	Mediana
Chalk	29.300	20.079	31.00
Dolomite	14.380	4.250	10.00
Gypsum	9.740	8.460	8.20
Limestone	16.831	6.670	12.35
Sandstone	20.462	16.550	20.00
Shale	12.279	2.949	10.00
Siltstone	15.012	8.743	11.50

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la porosidad total (%) en diferentes rocas sedimentarias.

Tabla 4.14: Medidas de dispersión de la porosidad total (%)

Litología	Rango	IQR	σ	DM	DMd	DMg
Chalk	64.40	25.000	16.681	13.982	13.758	16.865
Dolomite	42.99	24.000	14.930	12.852	12.287	12.658
Gypsum	17.00	7.700	5.197	4.304	4.170	4.176
Limestone	59.99	21.725	15.422	12.804	12.471	13.293
Sandstone	54.00	19.000	11.350	9.451	9.411	9.781
Shale	49.99	15.830	13.251	10.144	9.346	11.014
Siltstone	40.99	18.000	11.415	9.120	8.496	9.053

Nota:

Tabla de comparación de las medidas de dispersión de la porosidad total (%) en diferentes rocas sedimentarias. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

roca y que modifican sus propiedades originales, aumentando o disminuyendo en gran medida la porosidad total para un mismo tipo de roca.

Analizando los histogramas de la figura 4.17 se muestran diferentes comportamientos dependiendo del tipo de roca. La dolomita, la caliza y la lutita, tienen histogramas unimodales con un sesgo positivo, por lo que sus mayores frecuencias de datos se encuentran en valores entre el 1 y 4% de porosidad total.

Una roca que difiere por mucho de las demás es la creta, cuya distribución tiene un sesgo negativo y cuyas mayores frecuencias se encuentran en valores entre el 40 y 42% de porosidad total.

Continuando con el análisis de la figura 16; el yeso, la arenisca y la limolita, también tienen cierto sesgo positivo, siendo la arenisca la menos sesgada y la que tiende a la simetría. El yeso tiene sus mayores frecuencias en estimaciones del 5% de porosidad, mientras que la limolita concentra la mayoría de sus datos en valores entre el 8 y el 15%; por su parte el histograma de la arenisca es el único bimodal, con las mayores frecuencias en valores cercanos al 12 y 30%.

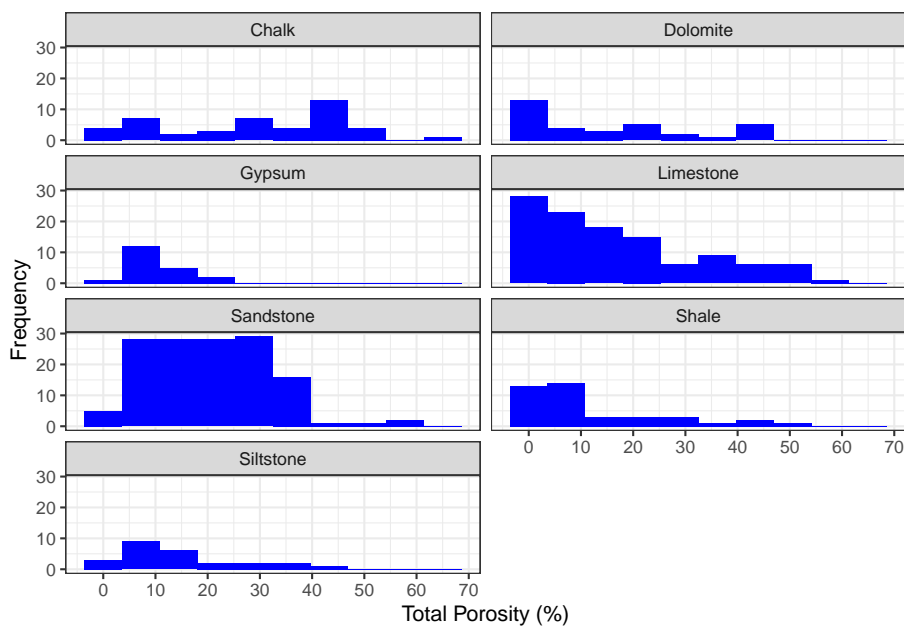


Figura 4.17: Histogramas de la porosidad total (%) en diferentes rocas sedimentarias. Histogramas construidos a partir de 10 clases

4.3.1.4 Sedimentos no consolidados

Los sedimentos no consolidados alcanzan las estimaciones más elevadas de porosidad total en comparación con el resto de las litologías. En la figura 4.18 se muestran diferentes tipos de sedimentos, cuya porosidad puede llegar a alcanzar hasta el 50% en la mayoría de los casos. Una característica para resaltar es el hecho de que las litologías más finas como es el caso de las arcillas y limos tienden a ser más porosas que las litologías más gruesas como las gravas.

La combinación de materiales tiende a disminuir las porosidades; un ejemplo es el caso de una combinación de arena y grava, mostrando una porosidad menor que la grava o arena limpia.

En ningún caso a excepción de la combinación de arena, limo y arcilla se registran valores anómalos.

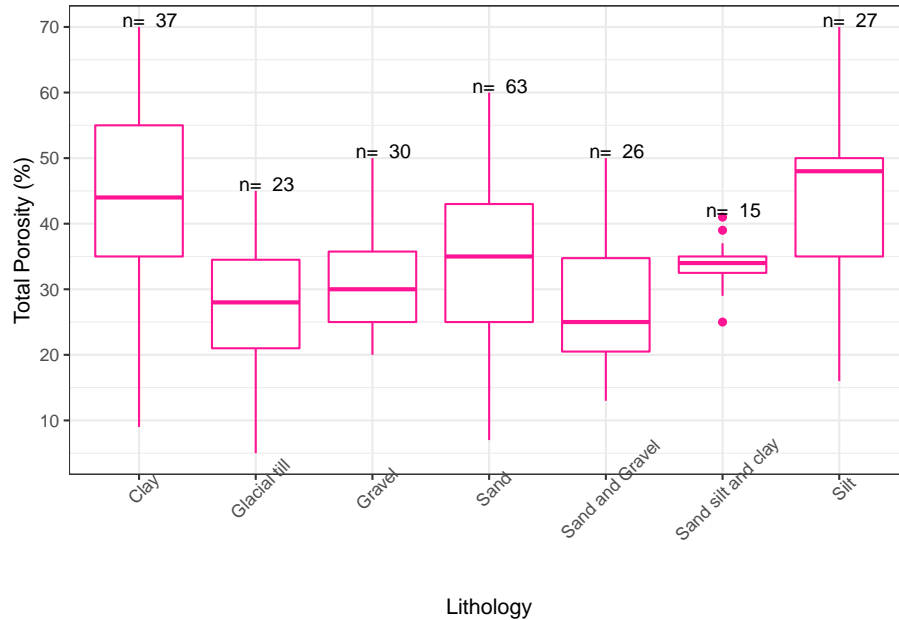


Figura 4.18: Diagramas de caja de la porosidad total (%) en diferentes sedimentos no consolidados.

Con la tabla 4.15 se puede resumir la información de los diagramas de caja de la porosidad total (%) en sedimentos no consolidados. En esta tabla es más fácil observar los valores puntuales de los valores máximos y mínimos, así como los cuartiles Q1 y Q2 de cada litología.

El valor mínimo se registra en las combinaciones de sedimentos de origen glacial (glacial till) con un 5%; mientras que el valor máximo está presente en las arcillas con un 70%. Los limos son los que tienen una mayor medida de la mediana con un 35%, mientras que la combinación de arenas y gravas registran el menor valor de la mediana con un 15%

La comparación de las medidas de tendencia central de la porosidad total (%) en sedimentos no consolidados se presentan en la tabla 4.16. Se observa que, en las arcillas, sedimentos de origen glacial y arenas, la mediana resulta ser más parecida a la media aritmética que a la media geométrica.

En el caso de las gravas la mediana y la media geométrica se diferencian en menor grado de la media aritmética. En la combinación de arenas, limos y arcillas las tres medidas de tendencia central son muy parecidas entre sí variando solo en decimales. Para los limos y la combinación de arenas y gravas las tres medidas

Tabla 4.15: Valores mínimos, máximos y cuartiles de la porosidad total (%)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Clay	37	9	35.0	44	55.00	70
Glacial till	23	5	21.0	28	34.50	45
Gravel	30	20	25.0	30	35.75	50
Sand	63	7	25.0	35	43.00	60
Sand and Gravel	26	13	20.5	25	34.75	50
Sand silt and clay	15	25	32.5	34	35.00	41
Silt	27	16	35.0	48	50.00	70

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la Porosidad Total (%) en diferentes sedimentos no consolidados. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.16: Comparación de las medidas de tendencia central de la porosidad total (%)

Litología	Media	Media_geométrica	Mediana
Clay	43.203	38.599	44
Glacial till	26.783	24.049	28
Gravel	31.367	30.245	30
Sand	34.659	32.507	35
Sand and Gravel	27.692	26.267	25
Sand silt and clay	33.867	33.652	34
Silt	44.296	42.455	48

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la porosidad total (%) en diferentes sedimentos no consolidados.

de tendencia central son bastante diferentes entre sí, variando en más de una unidad.

Analizando la desviación estándar de la tabla 4.17 se puede inferir que las arcillas son los sedimentos que presentan una mayor dispersión entre sus datos de porosidad total, mientras que la combinación de arenas, limos y arcillas tienen una menor dispersión en comparación con los otros tipos de sedimento. Observando las desviaciones medias absolutas respecto a un punto central se puede observar que en las gravas, arenas, limos, arcillas y sedimentos de origen glacial la desviación respecto a la mediana es menor que la desviación respecto a la media aritmética y respecto a la media geométrica.

Para la combinación de arenas y gravas y la combinación de arenas, limos y

Tabla 4.17: Medidas de dispersión de la porosidad total (%)

Litología	Rango	IQR	σ	DM	DMd	DMg
Clay	61	20.00	16.541	12.684	12.662	13.630
Glacial till	40	13.50	10.535	8.265	8.000	8.942
Gravel	30	10.75	8.660	7.058	6.967	6.983
Sand	53	18.00	11.162	8.956	8.897	9.363
Sand and Gravel	37	14.25	9.298	7.521	6.846	7.139
Sand silt and clay	16	2.50	3.852	2.693	2.667	2.736
Silt	54	15.00	11.740	9.108	8.741	9.586

Nota:

Tabla de comparación de las medidas de dispersión de la porosidad total (%) en diferentes sedimentos consolidados. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

arcillas también se puede concluir que la desviación de sus datos respecto a la mediana es menor que la desviación respecto a la media aritmética y la desviación respecto a la media geométrica.

La gran dispersión que existe entre los datos de porosidad efectiva de un mismo tipo de sedimento se debe a la existencia de diferentes tamaños y formas de grano dentro de sedimentos del mismo tipo, así como la combinación con otros materiales que afectan los valores de porosidad total.

En la figura 4.19 se presentan los histogramas de la porosidad total (%) para cada uno de los tipos de sedimento. Estos histogramas varían en gran medida para cada tipo. En el caso de las arcillas, arenas, limos y la combinación de arenas limos y arcillas, se presentan distribuciones que tienen cierta simetría y unimodalidad. Las arcillas, arenas y limos tienen las mayores frecuencias en estimaciones cercanas a 40%, 40% y 45% de porosidad total respectivamente. La combinación de arenas limos y arcillas tiene una mayor frecuencia en 35% de porosidad total.

Para el caso de los sedimentos de origen glacial se tiene un histograma con sesgo negativo y la mayor frecuencia en estimaciones cercanas al 35% de porosidad. En el caso de gravas y combinación de arena y grava las distribuciones tienden a tener un sesgo positivo; para la combinación de arena y grava las mayores frecuencias se presentan en valores cercanos al 25% de porosidad; y finalmente para las gravas se tiene una bimodalidad con altas frecuencias en valores cercanos a 18% y al 32% de porosidad total.

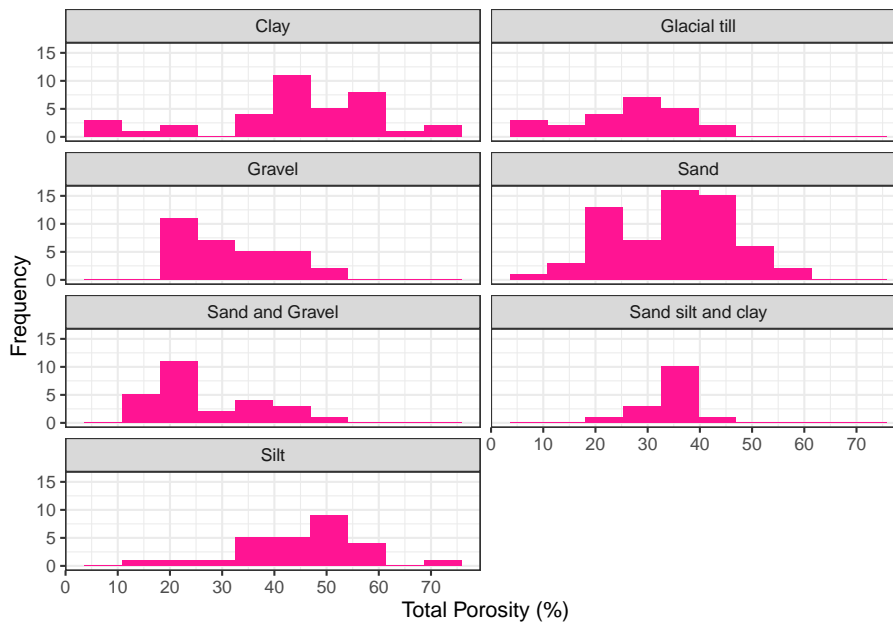


Figura 4.19: Histogramas de la porosidad total (%) en diferentes sedimentos no consolidados. Histogramas construidos a partir de 10 clases

4.3.2 Porosidad Efectiva

La porosidad efectiva tiene un amplio rango de valores para todos los tipos de litologías. Estos rangos van desde estimaciones cercanas a 0 hasta alcanzar valores del 60% (figura 4.20). La presencia de estos valores en extremo elevados se debe a estimaciones atípicas de litologías específicas en los registros.

La figura 4.21 muestra los rangos para algunas litologías específicas tomando en cuenta los valores anómalos. Se puede identificar que al igual que para la porosidad total, es la toba la que presenta los mayores valores máximos y anómalos para rocas ígneas en general. El basalto puede llegar a alcanzar un 30% y la ignimbrita más del 30%.

El gneis y el esquisto son las rocas más representativas de las rocas metamórficas en la base de datos ya que son las únicas litologías con los suficientes datos para un análisis de datos de porosidad efectiva. Rocas sedimentarias y sedimentos consolidados presentan altos valores de porosidad efectiva por la presencia de ciertas litologías que se describirán más adelante.

En la figura 4.22 se presentan los diagramas de caja de porosidad efectiva para las diferentes clasificaciones litológicas, en la cual se pueden apreciar mejor los rangos sin la presencia de valores anómalos. Los sedimentos no consolidados son los que alcanzan un mayor registro de porosidad efectiva de hasta el 50% sin

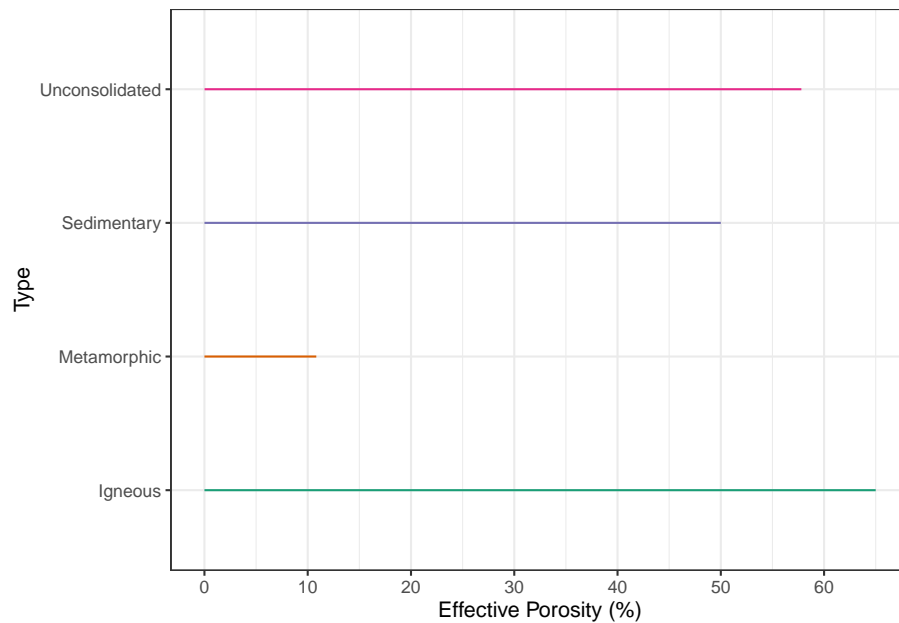


Figura 4.20: Diagrama de rangos de la porosidad efectiva (%) para los diferentes tipos de litología.

contar valores anómalos.

Para el caso de rocas ígneas fracturadas, aquellos valores que son superiores al 35% se consideran anómalos, mientras que las rocas metamórficas tienen valores anómalos a partir del 2% de porosidad efectiva. Finalmente, las porosidades efectivas superiores al 38% en rocas sedimentarias se consideran atípicas.

El resumen de los diagramas de caja se presenta en la tabla 4.18 en donde se muestran los valores máximos y mínimos, así como los cuartiles. Podemos observar que las rocas ígneas, sedimentarias y metamórficas tienen valores mínimos de 0.001% y para el caso de los sedimentos no consolidados, el valor mínimo es 0.01%. Las litologías que tienen los mayores valores máximos son las rocas sedimentarias y sedimentos no consolidados, con 50 y 57.8% de porosidad efectiva.

Analizando el valor de las medianas, se puede inferir que las rocas sedimentarias y los sedimentos no consolidados, tienen un mayor valor de las medianas con un 3.45 y 23% de porosidad efectiva respectivamente, mientras que las rocas ígneas y metamórficas tienen los menores valores para la mediana con 2 y 0.5% de porosidad efectiva respectivamente.

Comparando las distintas mediciones de tendencia central (4.19) podemos observar que, en rocas ígneas, sedimentarias y metamórficas, la mediana y la

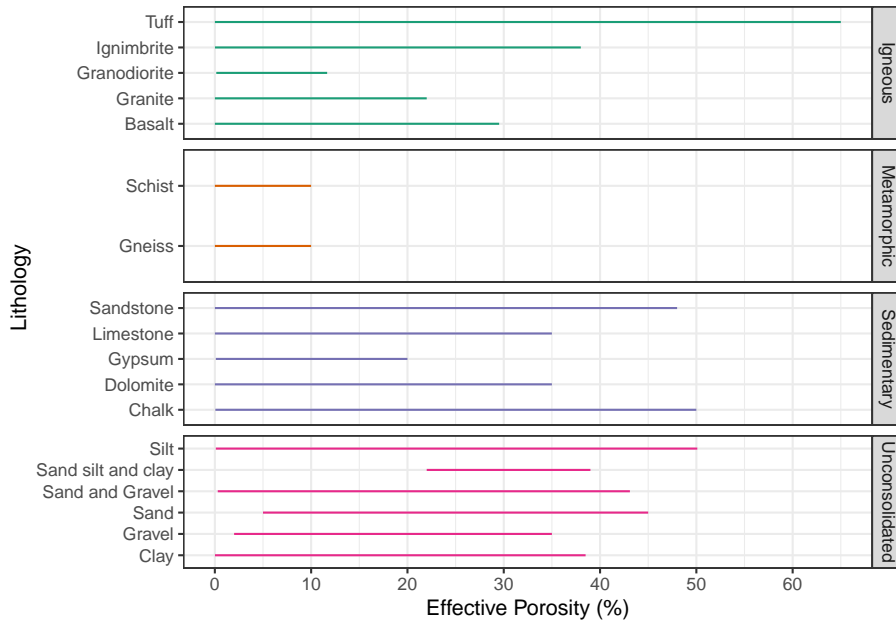


Figura 4.21: Diagrama de rangos de la porosidad efectiva (%) para diferentes litologías.

Tabla 4.18: Valores mínimos, máximos y cuartiles de la porosidad efectiva (%)

Tipo	No.Datos	min.	Q1	Mediana	Q3	max.
Igneous	233	1e-04	0.26	2.00	12.0	65.00
Metamorphic	79	1e-05	0.01	0.50	1.0	10.83
Sedimentary	338	1e-03	0.60	3.45	15.0	50.00
Unconsolidated	235	1e-02	15.00	23.00	30.5	57.80

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la porosidad efectiva (%) para las diferentes clasificaciones litológicas. La columna **No. datos** corresponde con el número de observaciones.

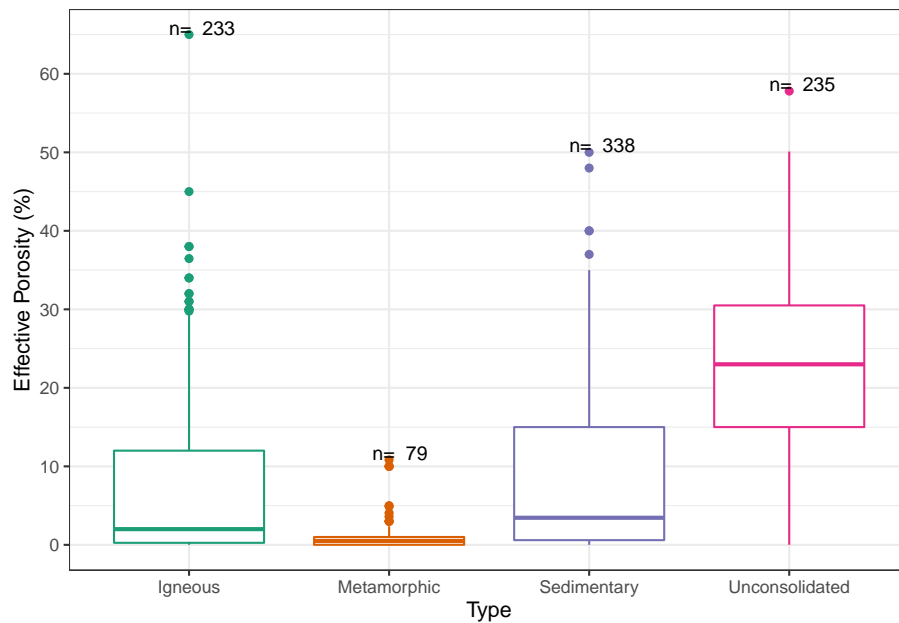


Figura 4.22: Diagramas de caja de la porosidad efectiva (%) para las diferentes clasificaciones litológicas.

Tabla 4.19: Comparación de las medidas de tendencia central de la porosidad efectiva (%)

Tipo	Media	Media_geométrica	Mediana
Igneous	7.611	1.555	2.00
Metamorphic	1.249	0.146	0.50
Sedimentary	8.668	2.211	3.45
Unconsolidated	22.505	17.403	23.00

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la porosidad efectiva (%) de los diferentes tipos de litologías.

media geométrica son más parecidas entre sí, que en comparación con la media aritmética, mientras que en sedimentos no consolidados la media aritmética y la mediana son más parecidas en comparación con la media geométrica.

Analizando las medidas de dispersión expuestas en la tabla 4.20, específicamente la desviación estándar, se puede concluir que los sedimentos no consolidados presentan una mayor dispersión en sus datos de porosidad efectiva (%) en comparación con las otras litologías, mientras que las rocas metamórficas son las

Tabla 4.20: Medidas de dispersión de la porosidad efectiva (%)

Tipo	Rango	IQR	σ	DM	DMd	DMg
Igneous	65.000	11.74	10.804	8.441	7.096	7.109
Metamorphic	10.830	0.99	2.349	1.409	1.127	1.187
Sedimentary	49.999	14.40	10.428	8.566	7.784	7.849
Unconsolidated	57.790	15.50	11.201	9.257	9.255	10.010

Nota:

Tabla de comparación de las medidas de dispersión de la porosidad efectiva (%) en los diferentes tipos de litologías. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

que tienen una menor dispersión. Analizando las desviaciones medias absolutas de un punto central se puede concluir que para todos los tipos de litologías la desviación de los datos respecto a la mediana es menor en comparación con las desviaciones respecto a la media aritmética y la media geométrica.

En la figura 4.23 se representan las distribuciones de la porosidad efectiva para las diferentes clasificaciones litológicas; en donde se observa que las rocas ígneas, metamórficas y sedimentarias tienen una distribución sesgada hacia la derecha con forma leptocúrtica, mientras que la distribución de los sedimentos no consolidados tiende a la simetría con una forma mesocúrtica.

Los sedimentos no consolidados son multimodales con varios picos en valores cercanos a 8, 18 y 25%, otro valor en 35% también podría considerarse una moda. En el caso de las rocas ígneas, sedimentarias y metamórficas, las frecuencias tienden hacia los valores de 1% y 2% de porosidad efectiva en los tres casos.

Correlación porosidad efectiva

La figura 4.24 muestra las correlaciones y los modelos de regresión de la porosidad efectiva en función de cada propiedad hidrogeológica (porosidad total, conductividad hidráulica, coeficiente de almacenamiento, rendimiento específico y compresibilidad). Se puede concluir que para los modelos B, C, D, E los modelos lineales son bastante pobres, ni siquiera en datos con transformación logarítmica.

El modelo del gráfico A (porosidad efectiva vs porosidad total) muestra ser el que tiene un mejor ajuste con un r^2 del 0.43.

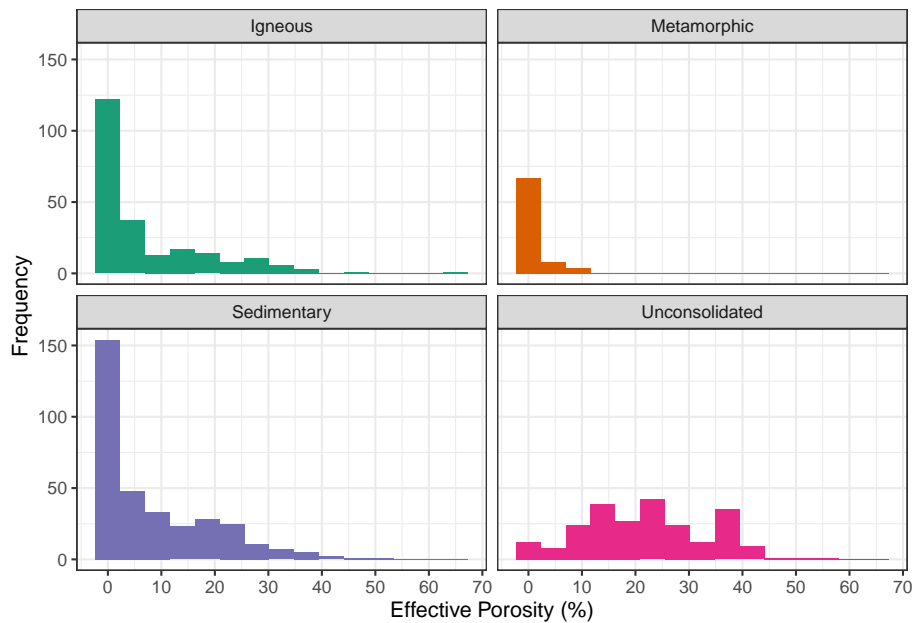


Figura 4.23: Histogramas de la porosidad efectiva (%) para las diferentes clasificaciones litológicas. Histogramas construidos a partir de 15 clases

4.3.2.1 Rocas ígneas

Las rocas ígneas en general tienen porosidades efectivas (%) bastante bajas a excepción de algunos casos como la ignimbrita y la toba, que como se muestra en la figura 4.25 pueden llegar a alcanzar valores extremos de 38 y 45% de porosidad efectiva. Excluyendo a la toba y la ignimbrita, la porosidad efectiva no supera el 15%, a excepción de valores atípicos.

De la misma forma que la porosidad total en este tipo de rocas, los valores límite y valores anómalos se deben a la presencia de rocas extremadamente fracturadas e intemperizadas que propician el aumento de la porosidad efectiva en este tipo de rocas.

Las tobas del valle central del norte de Costa Rica expuestas por Foster et al. (1985) son un claro ejemplo de valores anómalos de porosidad efectiva, llegando a registrar hasta un 65%. Otros ejemplos de valores anómalos se presentan en los basaltos y granitos extremadamente intemperizados expuestos por Gupta and Rao (2001), donde se alcanzan registros mayores al 15% de porosidad efectiva en este tipo de rocas.

La tabla 4.21 presenta los valores puntuales de los diagramas de caja. Se puede resumir que la litología con el valor de porosidad efectiva (%) más bajo es el basalto con 0.001%, mientras que la toba y la ignimbrita tienen los valores más

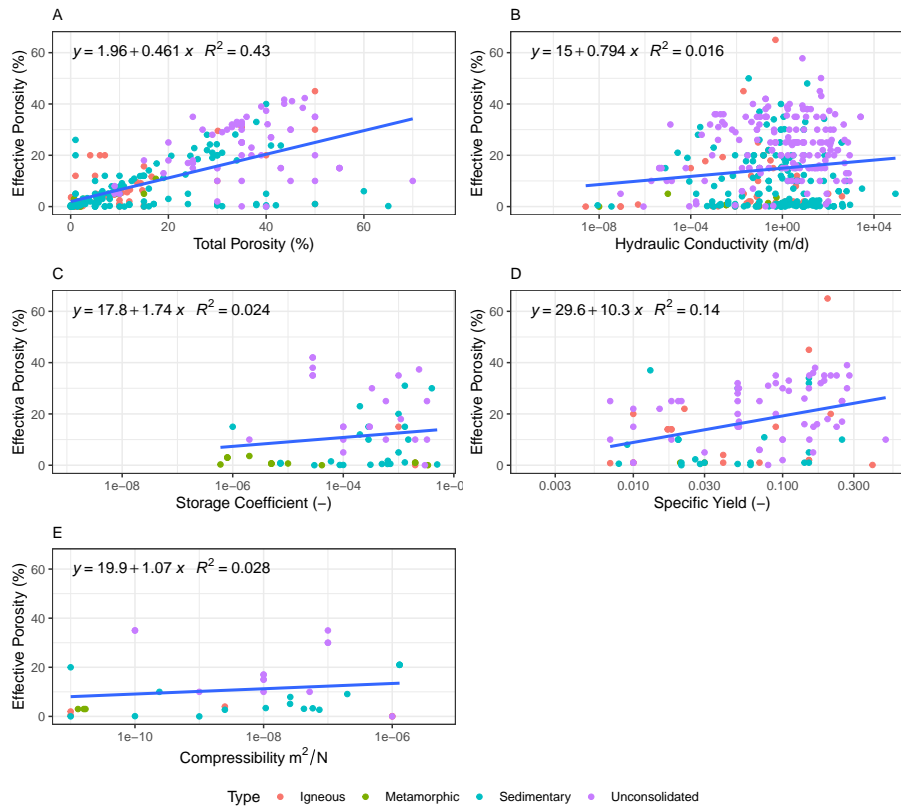


Figura 4.24: Diagramas de correlación de la porosidad efectiva en función de las diferentes propiedades hidrogeológicas. A: Porosidad Efectiva vs Porosidad Total. B: Porosidad Efectiva vs Conductividad Hidráulica (transformación logarítmica en base 10 del eje x). C: Porosidad Efectiva vs Coeficiente de Almacenamiento (transformación logarítmica en base 10 del eje x). D: Porosidad Efectiva vs Rendimiento Específico (transformación logarítmica en base 10 del eje x). E: Porosidad Efectiva vs Compresibilidad (transformación logarítmica en base 10 del eje x)

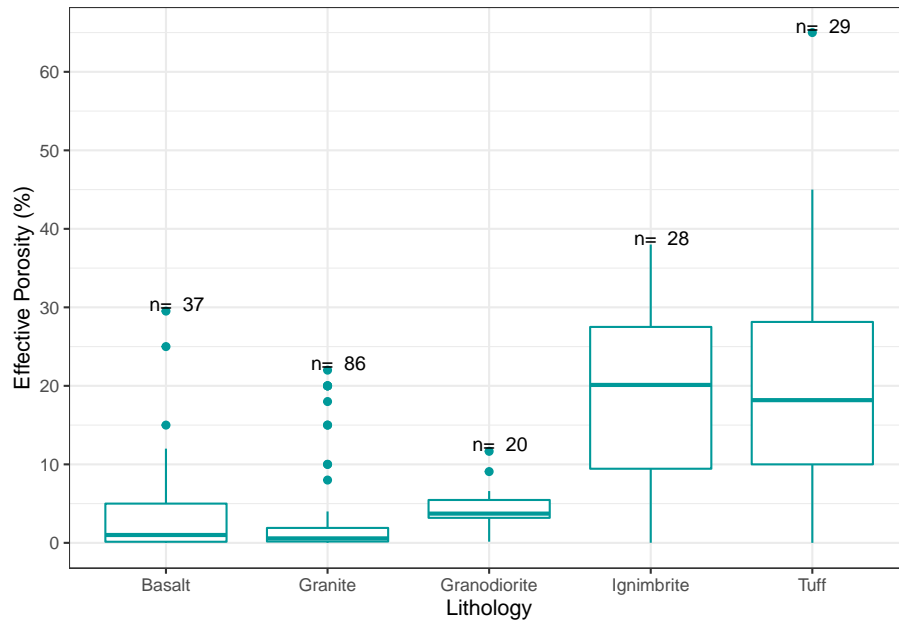


Figura 4.25: Diagramas de caja de la porosidad efectiva (%) en diferentes rocas ígneas.

Tabla 4.21: Valores mínimos, máximos y cuartiles de la porosidad efectiva (%)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Basalt	37	0.0001	0.140	1.000	5.000	29.52
Granite	86	0.0010	0.172	0.570	1.903	22.00
Granodiorite	20	0.1500	3.185	3.725	5.462	11.66
Ignimbrite	28	0.0100	9.440	20.115	27.505	38.00
Tuff	29	0.0010	10.000	18.180	28.140	65.00

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la porosidad efectiva (%) en diferentes rocas ígneas. La columna **no. datos** corresponde con el número de observaciones.

altos con 65 y 38% de porosidad efectiva.

Las estimaciones más altas de la mediana son para la ignimbrita con un 20.11% y la toba con un 18.18% de porosidad efectiva. El registro más bajo de la mediana lo tiene el granito con un 0.57%.

La tabla 4.22 es un resumen de las comparaciones de diferentes medidas de tendencia central de la porosidad efectiva (%) en rocas ígneas. Puede observarse

Tabla 4.22: Comparación de las medidas de tendencia central de la porosidad efectiva (%)

Litología	Media	Media_geométrica	Mediana
Basalt	4.648	0.747	1.000
Granite	2.720	0.554	0.570
Granodiorite	4.205	2.624	3.725
Ignimbrite	19.008	11.572	20.115
Tuff	19.566	7.198	18.180

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la porosidad efectiva (%) en diferentes rocas ígneas.

que, en el caso del basalto, granito y granodiorita, la mediana y la media geométrica resultan ser más parecidas en comparación con la media aritmética. En la ignimbrita y la toba ocurre que la media aritmética resulta ser más parecida la mediana que a la media geométrica.

La tabla 4.23 nos permite analizar las medidas de dispersión de los datos. Se puede inferir de acuerdo con los valores de la desviación estándar que la toba es la roca que tiene una mayor dispersión en su conjunto de datos de porosidad efectiva, mientras que el granito es la litología con la menor dispersión entre sus datos.

Analizando las desviaciones medias absolutas de un punto central se puede observar que en todos los casos la desviación de los datos respecto a la mediana es menor que la desviación respecto a la media aritmética y la desviación respecto a la media geométrica. La gran dispersión que existe entre los datos de porosidad efectiva en rocas ígneas se debe a que dentro de cada litología se presentan ciertas características particulares como lo son la presencia de fracturas e intemperismo que afectan en gran medida provocando que los datos de una misma litología tengan un gran rango de valores de porosidad efectiva.

Los histogramas de las rocas ígneas fracturadas y/o intemperizadas (figura 4.26) tienden a ser muy diversos y cambiar para cada tipo de litología, en las distribuciones del basalto y el granito se tienen sesgos positivos y forma platocúrtica, con las mayores frecuencias en estimaciones entre el 0 y el 1% de porosidad efectiva en ambos casos.

La granodiorita tiende a tener cierta simetría, con una mayor frecuencia en estimaciones cercanas al 4% de porosidad efectiva, mientras que el histograma de la ignimbrita tiene una tendencia hacia un sesgo negativo, con las mayores frecuencias en datos cercanos al 25%. Finalmente, la toba presenta un ligero sesgo positivo y se caracteriza por ser multimodal con las mayores frecuencias en valores cercanos a 2, 15 y 30% de porosidad efectiva.

Tabla 4.23: Medidas de dispersión de la porosidad efectiva (%)

Litología	Rango	IQR	σ	DM	DMd	DMg
Basalt	29.520	4.860	6.924	4.966	4.232	4.307
Granite	21.999	1.730	5.510	3.486	2.541	2.541
Granodiorite	11.510	2.277	2.905	2.098	2.057	2.519
Ignimbrite	37.990	18.065	11.505	9.932	9.866	11.466
Tuff	64.999	18.140	15.119	11.428	11.362	15.135

Nota:

Tabla de comparación de las medidas de dispersión de la porosidad efectiva (%) en diferentes rocas ígneas. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

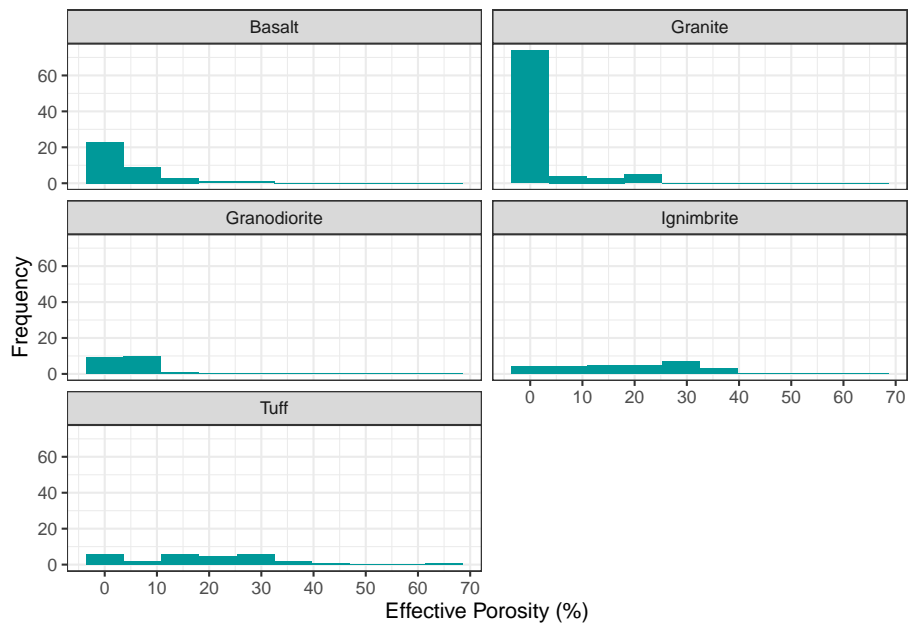


Figura 4.26: Histogramas de la porosidad efectiva (%) en diferentes rocas ígneas. Histogramas construidos a partir de 10 clases

4.3.2.2 Rocas Metamórficas

Las rocas metamórficas al igual que las rocas ígneas generan porosidades efectivas muy bajas, salvo algunos valores anómalos. En la figura 4.27 se puede observar que los valores atípicos pueden llegar a alcanzar hasta un 10% de porosidad en los dos ejemplos de rocas metamórficas. La porosidad efectiva en este tipo de rocas varía en un rango muy pequeño, un rango menor al 1%.

La presencia de datos anómalos se debe a la existencia de registros de rocas metamórficas con un alto grado de intemperismo y fracturamiento que genera porosidades secundarias bastante elevadas. Un ejemplo son las cuarcitas completamente intemperizadas de Delhi, India expuestas por Gupta and Rao (2001) donde se registra hasta un 10% de porosidad efectiva en este tipo de rocas. Otro ejemplo son los esquistos fracturados e intemperizados expuestos por Taylor and Eggleton (2001) donde se registra hasta un 10% de porosidad efectiva.

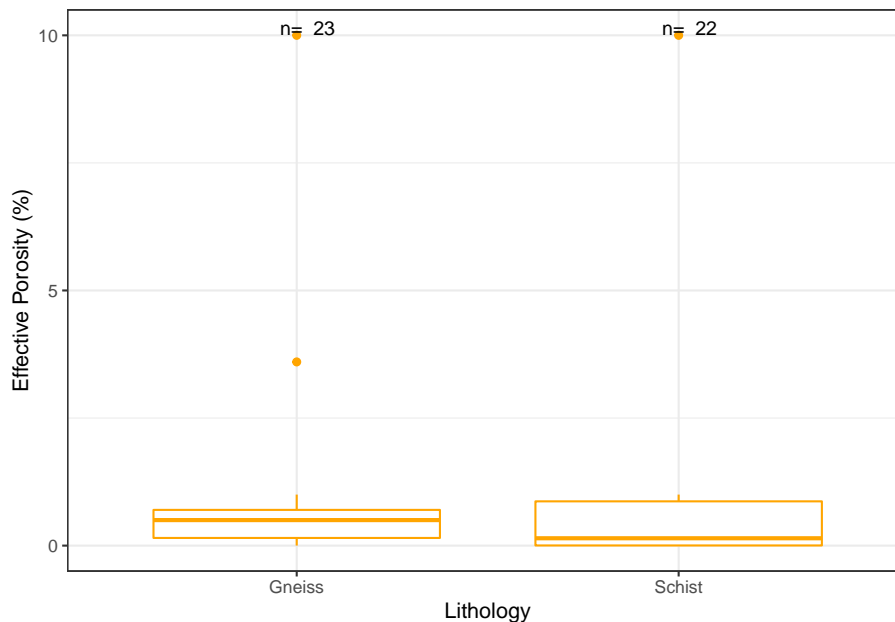


Figura 4.27: Diagramas de caja de la porosidad efectiva (%) en diferentes rocas metamórficas.

En la tabla 4.24 se muestran los valores puntuales de los diagramas de caja. En la tabla, ambas rocas se caracterizan por tener porosidades efectivas muy bajas, la que tiene un menor valor de la mediana es el gneiss con 0.5% mientras que el esquisto presenta un valor de la mediana de 0.91%.

Analizando las medidas de tendencia central (tabla 4.25) se puede observar que en ambos casos las medidas de tendencia central son muy diferentes entre sí,

Tabla 4.24: Valores mínimos, máximos y cuartiles de la porosidad efectiva (%)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Gneiss	23	1e-03	0.150	0.500	0.700	10
Schist	22	1e-05	0.003	0.144	0.867	10

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la porosidad efectiva (%) en diferentes rocas metamórficas. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.25: Comparación de las medidas de tendencia central de la porosidad efectiva (%)

Litología	Media	Media_geométrica	Mediana
Gneiss	0.976	0.247	0.500
Schist	0.810	0.030	0.144

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la porosidad efectiva (%) en diferentes rocas metamórficas.

siendo la mediana y la media geométrica las que varían en menor grado, una de otra, en comparación con la media aritmética.

Como puede observarse en la tabla 4.26 la desviación estándar es mayor en el gneis, lo que indica una mayor dispersión entre sus datos. La dispersión de los datos se debe a la existencia de rocas fracturadas e intemperizadas dentro del subgrupo de datos de cada roca.

En ambos casos las desviaciones de los datos respecto a la mediana son menores que las desviaciones respecto a la media geométrica y la media aritmética.

En la figura 4.28 se muestran los histogramas de las rocas metamórficas. En ambos casos las distribuciones tienen un sesgo positivo; para el gneis y para el esquisto las mayores frecuencias se concentran en estimaciones cercanas a 1% de porosidad efectiva.

4.3.2.3 Rocas Sedimentarias

Las rocas sedimentarias varían en un amplio rango de estimaciones de porosidad efectiva, pudiendo alcanzar hasta un 50% de porosidad efectiva en situaciones extremas en rocas como los carbonatos y la arenisca. En la figura 4.29 se puede identificar a la arenisca como la única litología cuyas estimaciones son mayores al 35% de porosidad efectiva sin tener en cuenta valores atípicos, mientras que los

Tabla 4.26: Medidas de dispersión de la porosidad efectiva (%)

Litología	Rango	IQR	σ	DM	DMd	DMg
Gneiss	9.999	0.550	2.101	1.019	0.809	0.848
Schist	10.000	0.864	2.095	0.914	0.802	0.804

Nota:

Tabla de comparación de las medidas de dispersión de la porosidad efectiva (%) en diferentes rocas metamórficas. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (Dmg).

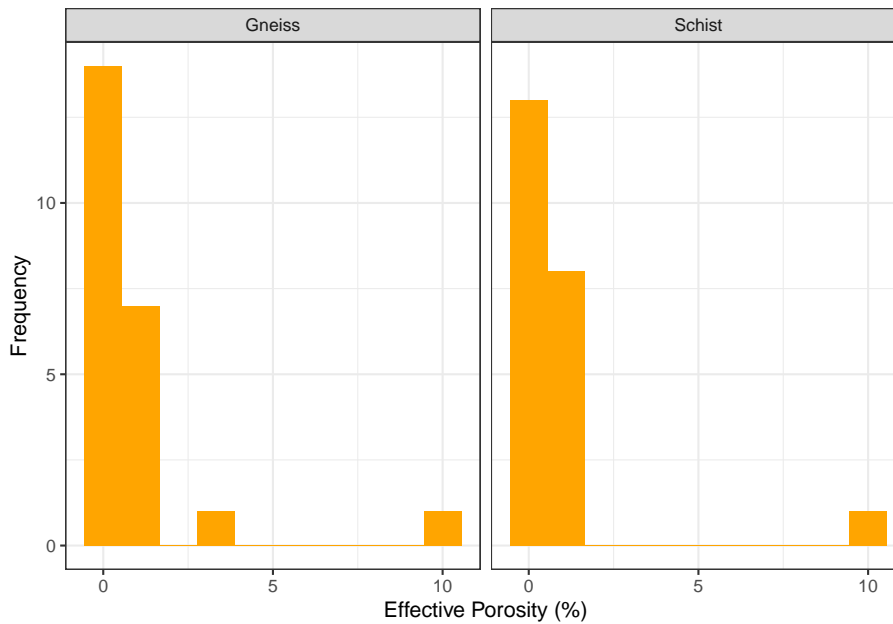


Figura 4.28: Histogramas de la porosidad efectiva (%) en diferentes rocas metamórficas. Histogramas construidos a partir de 10 clases

carbonatos no superan el 25% a excepción de casos anómalos. El yeso también puede llegar a alcanzar valores importantes de porosidad efectiva.

La existencia de grandes rangos de porosidad efectiva, así como la existencia de valores anómalos dentro de una misma litología se debe a diferentes estructuras como el karst y las fracturas que aumentan en gran medida las estimaciones de porosidad efectiva. Algunos ejemplos de valores anómalos producto de karstificación se presentan en los carbonatos de Dublin, Polonia donde Motyka (1998) registra más del 40% de porosidad efectiva en este tipo de rocas.

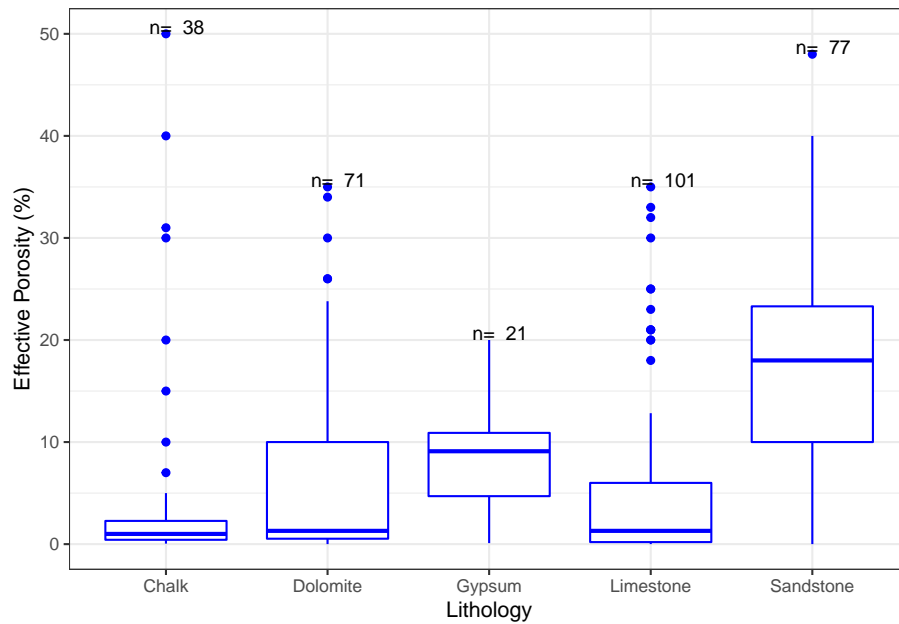


Figura 4.29: Diagramas de caja de la porosidad efectiva (%) en diferentes rocas sedimentarias.

A manera de resumen, en la tabla 4.27 se presenta la información de los diagramas de caja, en donde es posible apreciar mejor las estimaciones máximas y mínimas que registra cada roca, así como el valor de sus cuartiles. Las areniscas son las rocas que tienen un mayor valor de la mediana con 18% en comparación con las otras rocas sedimentarias, mientras que la creta es la roca que registra el menor valor de la mediana con 1% de porosidad efectiva.

La tabla 4.28 muestra la comparación de las diferentes medidas de tendencia central, obtenidas para cada tipo de roca. En esta tabla se puede observar que, en el caso de los carbonatos, la mediana y la media geométrica son más parecidas entre sí, que con la media aritmética. En el caso de la arenisca y el yeso la mediana es más parecida a la media aritmética que a la media geométrica.

Analizando las desviaciones estándar de la tabla 4.29 se puede inferir que las

Tabla 4.27: Valores mínimos, máximos y cuartiles de la porosidad efectiva (%)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Chalk	38	0.050	0.425	1.0	2.275	50
Dolomite	71	0.001	0.530	1.3	10.000	35
Gypsum	21	0.100	4.700	9.1	10.900	20
Limestone	101	0.001	0.200	1.3	6.000	35
Sandstone	77	0.001	10.000	18.0	23.300	48

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la porosidad efectiva (%) en diferentes rocas sedimentarias. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.28: Comparación de las medidas de tendencia central de la porosidad efectiva (%)

Litología	Media	Media_geométrica	Mediana
Chalk	6.060	1.130	1.0
Dolomite	6.309	1.494	1.3
Gypsum	8.438	5.829	9.1
Limestone	5.445	1.145	1.3
Sandstone	17.172	6.893	18.0

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la porosidad efectiva (%) en diferentes rocas sedimentarias.

areniscas son las rocas que tienen una mayor dispersión entre sus datos, mientras que el yeso presenta una menor dispersión. La dispersión de los datos se produce por la existencia de diferentes estructuras y texturas dentro la misma litología que producen un aumento o reducción importante en el valor de la porosidad efectiva.

Analizando las desviaciones medias absolutas de un punto central se puede concluir que en todos los casos las desviaciones de los datos respecto a la mediana son menores que las desviaciones respecto a la media geométrica y a la media aritmética.

Los histogramas de la figura 4.30 representan el comportamiento de las distribuciones de datos para cada roca sedimentaria. Las calizas, dolomitas y cretas tienen un comportamiento muy parecido a una distribución unimodal y sesgo positivo, donde las mayores frecuencias se presentan en registros entre el 1 y 2% de porosidad efectiva.

Tabla 4.29: Medidas de dispersión de la porosidad efectiva (%)

Litología	Rango	IQR	σ	DM	DMd	DMg
Chalk	49.950	1.85	12.075	8.133	5.703	5.731
Dolomite	34.999	9.47	9.102	6.967	5.755	5.760
Gypsum	19.900	6.20	5.483	4.303	4.271	4.755
Limestone	34.999	5.80	8.409	6.085	5.012	5.017
Sandstone	47.999	13.30	10.817	8.570	8.529	12.438

Nota:

Tabla de comparación de las medidas de dispersión de la porosidad efectiva (%) en diferentes rocas sedimentarias. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

Para el caso de las areniscas la distribución cambia bastante dando como resultado un histograma bimodal con un ligero sesgo positivo, las mayores frecuencias de esta distribución se concentran en estimaciones cercanas al 2 y al 20% de porosidad efectiva. Otra litología que difiere de las demás es el yeso; el cual tiene pocos datos, pero se infiere cierta simetría y altas frecuencias en 12% de porosidad efectiva.

4.3.2.4 Sedimentos no consolidados

Los sedimentos no consolidados tienen estimaciones de porosidad efectiva muy elevadas en comparación con las rocas. En sedimentos no consolidados las arenas y los limos superan el 40% de porosidad efectiva sin la presencia de valores anómalos.

En la figura 4.31 se puede observar el comportamiento de los diagramas de caja para los diferentes tipos de sedimento; estimando que no existen valores anómalos a excepción de las arcillas y las gravas, mientras que el resto de los sedimentos confirman que valores elevados de porosidad efectiva ocurren sin que existan situaciones anómalas, superando el 40% de porosidad total en limos, arenas, combinación de gravas arenas, y combinación de arenas, limos y arcillas.

La tabla 4.30 muestra los valores puntuales de las estimaciones mínimas y máximas, así como los cuartiles de cada uno de los tipos de sedimentos expuestos. En esta tabla es posible apreciar el comportamiento de las medianas; siendo la combinación de arena, limo y arcilla la que alcanza el mayor registro de la mediana con 28.5% de porosidad efectiva, mientras que la arcilla es la que registra el menor valor de la mediana con 10%. Los limos son los sedimentos que registran el valor más alto de todas las litologías, con una porosidad efectiva del

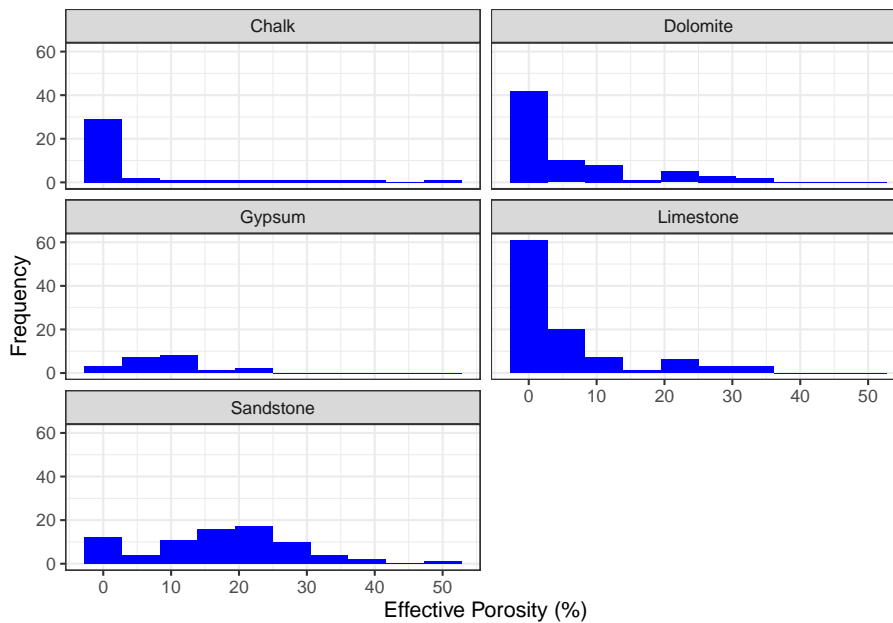


Figura 4.30: Histogramas de la porosidad efectiva (%) en diferentes rocas sedimentarias. Histogramas construidos a partir de 10 clases

50.1% mientras que la arcilla registra el valor más bajo con 0.01% de porosidad efectiva.

A manera de resumen la tabla 4.31 muestra la comparación de diferentes medidas de tendencia central, en donde se puede apreciar que las arcillas y arenas tienen un valor de la mediana más parecido al valor de la media aritmética que al valor de la media geométrica. En el caso de la grava el valor de la mediana es más parecido a la media geométrica que a la media aritmética. Para el caso de la combinación de arena, limo y arcilla, las tres medidas de tendencia central son muy parecidas y no varían en más de una unidad. Finalmente, los limos y la combinación de arenas y gravas tienen medidas de tendencia central muy diferentes que varían en más de una unidad.

Analizando las medidas de dispersión de la tabla 4.32 se puede apreciar que los limos son los sedimentos que tienen un mayor valor de la desviación estándar en relación con los otros tipos de sedimento lo que indica una mayor dispersión entre sus datos de porosidad efectiva (%), por otro lado, las combinaciones de arenas, limos y arcillas tienen la desviación estándar más baja lo que indica una menor dispersión entre sus datos en comparación con las otras litologías.

Las desviaciones medias absolutas de un punto central muestran que para la combinación de arenas, limos y arcillas las desviaciones son bastante parecidas.

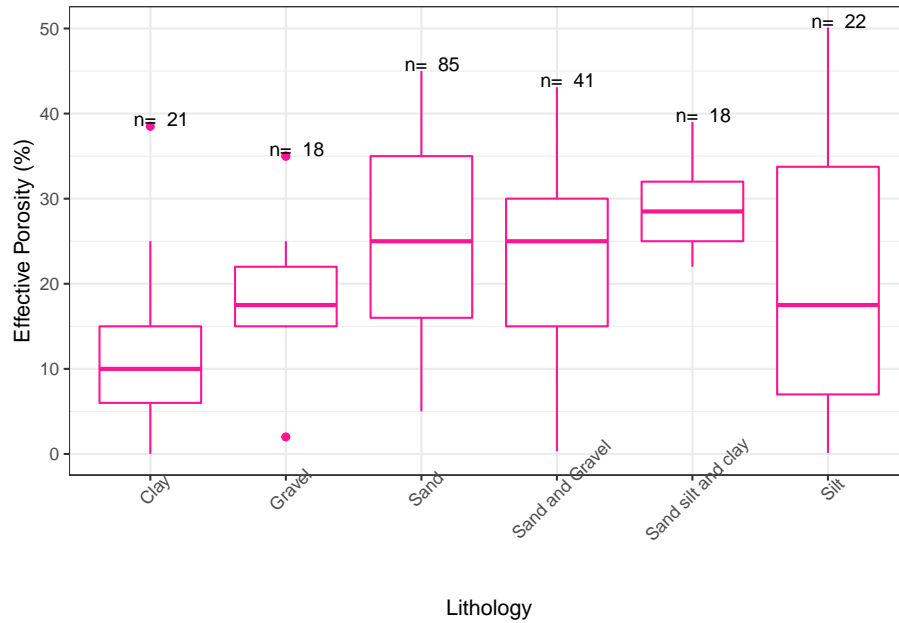


Figura 4.31: Diagramas de caja de la porosidad efectiva (%) en diferentes sedimentos no consolidados.

Tabla 4.30: Valores mínimos, máximos y cuartiles de la porosidad efectiva (%)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Clay	21	0.01	6	10.0	15.00	38.5
Gravel	18	2.00	15	17.5	22.00	35.0
Sand	85	5.00	16	25.0	35.00	45.0
Sand and Gravel	41	0.30	15	25.0	30.00	43.1
Sand silt and clay	18	22.00	25	28.5	32.00	39.0
Silt	22	0.10	7	17.5	33.75	50.1

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la porosidad efectiva (%) en diferentes sedimentos no consolidados. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.31: Comparación de las medidas de tendencia central de la porosidad efectiva (%)

Litología	Media	Media_geométrica	Mediana
Clay	11.715	6.255	10.0
Gravel	19.222	17.092	17.5
Sand	24.644	22.348	25.0
Sand and Gravel	22.955	18.289	25.0
Sand silt and clay	28.667	28.353	28.5
Silt	20.059	9.914	17.5

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la porosidad efectiva (%) en diferentes sedimentos no consolidados.

Tabla 4.32: Medidas de dispersión de la porosidad efectiva (%)

Litología	Rango	IQR	σ	DM	DMd	DMg
Clay	38.49	9.00	8.960	6.483	6.238	7.652
Gravel	33.00	7.00	7.705	5.580	5.444	5.444
Sand	40.00	19.00	9.887	8.214	8.201	8.434
Sand and Gravel	42.80	15.00	10.710	8.853	8.604	9.673
Sand silt and clay	17.00	7.00	4.419	3.667	3.667	3.667
Silt	50.00	26.75	14.959	12.415	12.405	14.262

Nota:

Tabla de comparación de las medidas de dispersión de la porosidad efectiva (%) en diferentes sedimentos no consolidados. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

En la combinación de arenas y gravas la desviación de los datos respecto a la mediana resulta ser la menor en comparación con las desviaciones respecto a la media geométrica y la media aritmética. En las arcillas, arenas y limos la desviación respecto a la mediana resulta ser menor que las desviaciones respecto a la media aritmética y la media geométrica. Finalmente, las gravas tienen desviaciones respecto a la mediana y respecto a la media geométrica que son similares y son menores que la desviación respecto a la media aritmética.

La figura 4.32 muestra los diferentes histogramas para los diferentes sedimentos no consolidados, con diferentes formas y tendencias dependiendo del tipo de sedimento. La distribución de las gravas y la combinación de arenas y gravas tienden a tener cierta simetría con mayores frecuencias en estimaciones cercanas

al 15 y 25% respectivamente.

Las arcillas y la combinación de arenas, limos y arcillas tienen distribuciones bimodales con un ligero sesgo positivo, la mayor frecuencia en estos dos casos se encuentra en valores cercanos al 8 y 18% para arcillas y 15 y 35% para la combinación de arenas, limos y arcillas.

Las arenas tienen una distribución multimodal donde las mayores frecuencias se concentran en estimaciones cercanas a 18, 20 y 40% de porosidad efectiva. Finalmente, los limos también muestran ser multimodales con las mayores frecuencias en datos cercanos a al 2, 20 y 35% de porosidad efectiva.

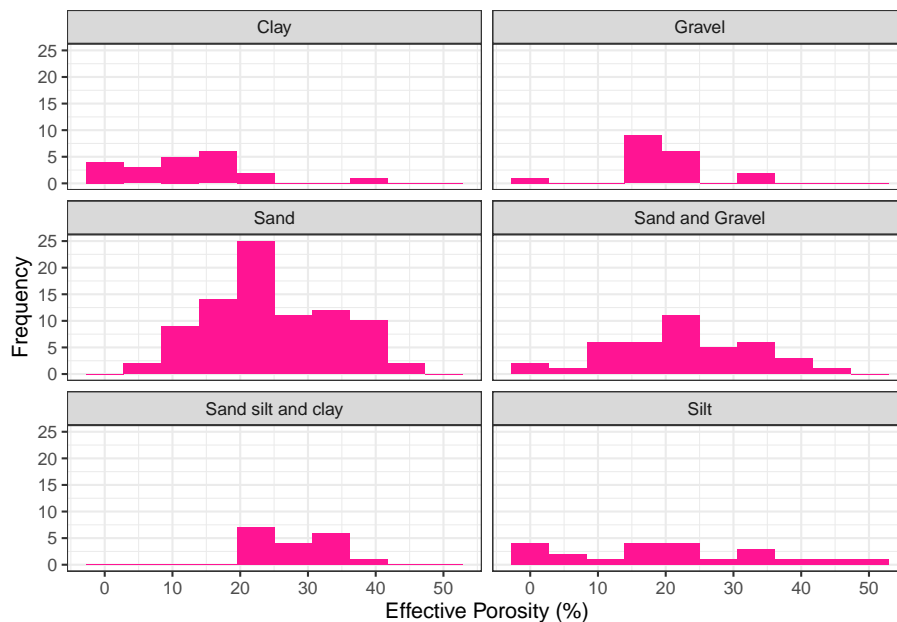


Figura 4.32: Histogramas de la porosidad efectiva (%) en diferentes sedimentos no consolidados. Histogramas construidos a partir de 10 clases

4.3.3 Conductividad Hidráulica

La conductividad hidráulica (m/d) ocurre en un amplio rango de valores para todos los tipos de litologías cambiando en varios órdenes de magnitud. Para todos los tipos de litología el rango de conductividad hidráulica varía en más de 9 órdenes de magnitud, siendo las rocas metamórficas las que varían en menor grado.

En la figura 4.33 se puede observar que las rocas sedimentarias y los sedimentos no consolidados varían en más de 16 órdenes de magnitud y superan estimaciones mayores a $1 \times 10^{+4}$ (m/d) de conductividad hidráulica. Para el

caso de los sedimentos no consolidados los valores mínimos registran hasta 1×10^{-10} (m/d); y en las rocas sedimentarias los valores mínimos registran hasta el 1×10^{-11} (m/d) de conductividad hidráulica.

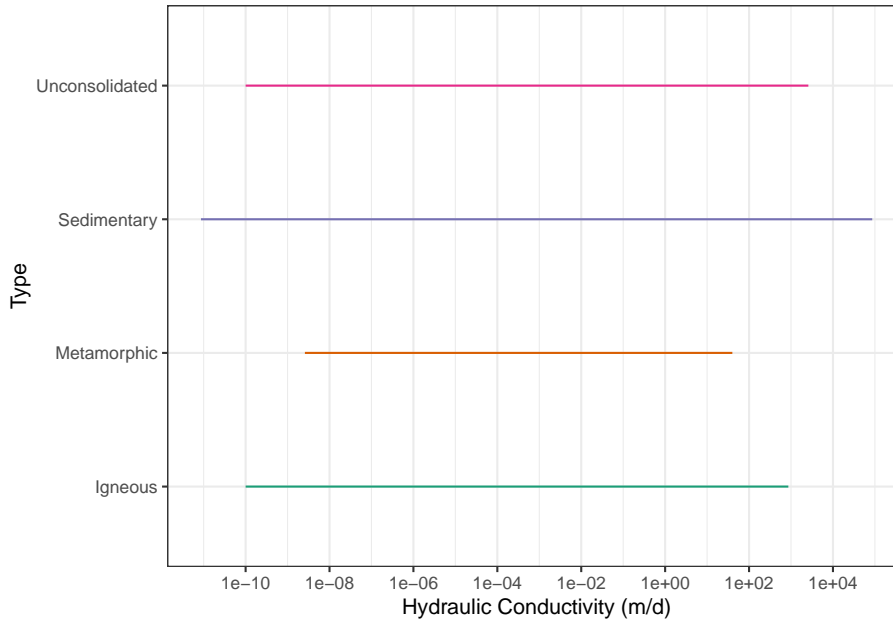


Figura 4.33: Diagrama de rangos en escala logarítmica en base 10 de la conductividad hidráulica (m/d) para los diferentes tipos de litología.

La conductividad hidráulica en rocas ígneas puede cambiar hasta en 14 órdenes de magnitud, con estimaciones que van desde 1×10^{-10} hasta superar los 500 (m/d) en casos extremos. En rocas metamórficas la conductividad hidráulica varía hasta en 9 órdenes de magnitud, con valores que van desde 1×10^{-8} (m/d) hasta superar los 20 (m/d) en casos atípicos.

Como ya se mencionó el elevado rango en el que varía la conductividad hidráulica (m/d) se debe por una parte a las propiedades de la roca como su porosidad, la cual aumenta por la presencia de fracturas. Por otra parte, existen otros parámetros que no son intrínsecos de la litología como lo es el gradiente hidráulico. Por lo tanto, un mismo tipo de litología puede tener un rango muy elevado de valores de conductividad hidráulica.

Por ejemplo, los trabajos de Kruseman et al. (1970) muestran que la conductividad hidráulica es de hasta 300 (m/d) en rocas extremadamente fracturadas y en rocas volcánica hasta 1000 (m/d).

Para visualizar mejor los rangos, la figura 4.34 contiene las litologías más representativas de cada tipo, en la cual se puede estimar que, en rocas ígneas, el

basalto presenta valores muy elevados, alcanzando hasta 1000 (m/d), mientras que el resto de las rocas ígneas supera por poco los 100 (m/d). A su vez el granito registra las menores estimaciones con 1×10^{-10} (m/d).

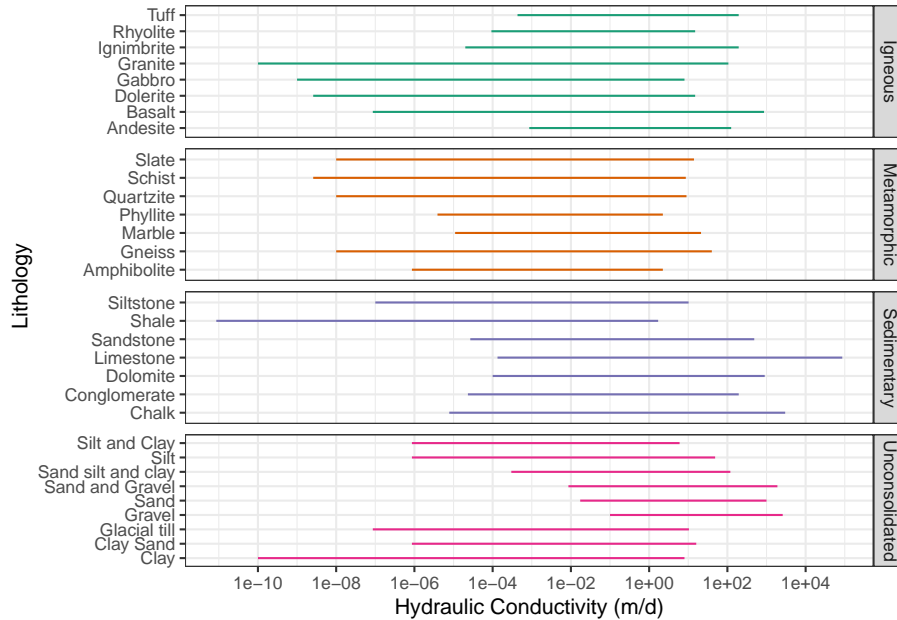


Figura 4.34: Diagrama de rangos en escala logarítmica en base 10 de la conductividad hidráulica (m/d) para diferentes litologías.

En el caso de las rocas metamórficas, el gneis es la roca que supera los 50 m/d, mientras que la mayoría de las rocas metamórficas no supera los 10 (m/d). La pizarra y el esquisto son litologías que registran las menores estimaciones, con valores menores a 1×10^{-8} (m/d).

En rocas sedimentarias los carbonatos registraron los valores más elevados de conductividad hidráulica (m/d). Las calizas en casos anómalos superan la estimación de $1 \times 10^{+4}$ (m/d), en el caso particular de la creta se alcanzó un valor de hasta 3000 (m/d). Por su parte las rocas como la lutita y la arenisca son las que registran las estimaciones más bajas con 1×10^{-11} y 1×10^{-10} (m/d) respectivamente.

Finalmente, los rangos de los sedimentos no consolidados como los limos y las arenas alcanzaron valores superiores a $1 \times 10^{+4}$ (m/d), mientras que la arcilla fue la litología que registró las estimaciones más bajas con 1×10^{-10} (m/d).

La figura 4.35 muestra los diferentes diagramas de caja de los tipos de litologías. Se puede observar que en el caso de rocas ígneas se consideran datos anómalos todas aquellas medidas mayores a 10 (m/d), en rocas metamórficas

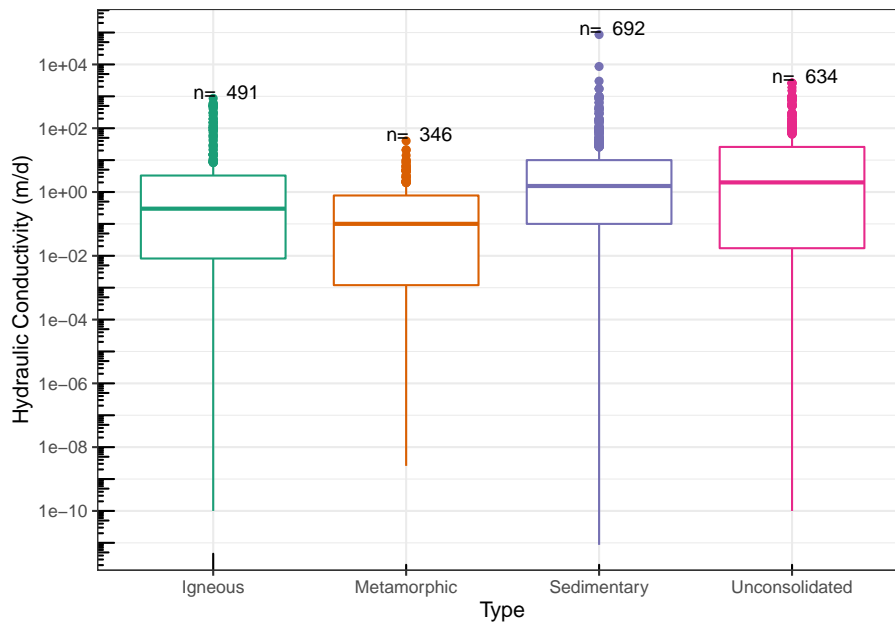


Figura 4.35: Diagramas de caja en escala logarítmica en base 10 de la conductividad hidráulica (m/d) para las diferentes clasificaciones litológicas.

aquellas estimaciones mayores a 2 (m/d) son consideradas anómalas. En rocas sedimentarias los rangos son mucho mayores y son considerados datos anómalos los registros de conductividad hidráulica mayores a 20 (m/d). Finalmente, en sedimentos no consolidados los registros de conductividad hidráulica mayores a 40 (m/d) son considerados anómalos.

La tabla 4.33 muestra los valores puntuales de los diagramas de caja, en la cual podemos observar que las rocas sedimentarias son las que registraron los menores valores con $8.64e^{-12}$ (m/d), mientras que los valores mayores entran presentes en sedimentos no consolidados con 86400 (m/d) de conductividad hidráulica.

Los sedimentos no consolidados son los que presentan un mayor valor de la mediana con 2 (m/d), mientras que las rocas metamórficas registraron el menor valor de la mediana con 0.1 (m/d).

La tabla 4.34 muestra las diferentes medidas de tendencia central de la conductividad hidráulica (m/d). En la tabla puede observarse que para todos los tipos de rocas la media aritmética difiere por mucho de la mediana y la media geométrica; esto se debe a la presencia de una gran cantidad de valores anómalos dentro del grupo de datos que afectan en gran medida la media aritmética.

La tabla 4.35 está compuesta por las diferentes medidas de dispersión de

Tabla 4.33: Valores mínimos, máximos y cuartiles de la conductividad hidráulica (m/d)

Tipo	No.Datos	min.	Q1	Mediana	Q3	max.
Igneous	491	1.00e-10	0.008	0.300	3.285	864
Metamorphic	346	2.59e-09	0.001	0.100	0.778	40
Sedimentary	692	8.64e-12	0.100	1.542	10.000	86400
Unconsolidated	634	1.00e-10	0.017	2.000	25.920	2592

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la conductividad hidráulica (m/d) para las diferentes clasificaciones litológicas. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.34: Comparación de las medidas de tendencia central de la conductividad hidráulica (m/d)

Tipo	Media	Media geométrica	Mediana
Igneous	25.836	0.078	0.3000
Metamorphic	1.003	0.017	0.1005
Sedimentary	174.257	0.607	1.5420
Unconsolidated	76.013	0.446	2.0000

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la conductividad hidráulica (m/d) en los diferentes tipos de litologías.

los datos de conductividad hidráulica para los diferentes tipos de litologías. Analizando las desviaciones estándar del logaritmo de los datos de conductividad hidráulica se puede inferir que las rocas sedimentarias presentan una menor dispersión entre sus datos, a diferencia de las rocas ígneas que son las que presentan una mayor dispersión entre sus datos.

Analizando las desviaciones medias absolutas se puede afirmar que en todos los casos, existe una menor desviación de los datos respecto a la mediana, en comparación con las desviaciones respecto a la media aritmética y la media geométrica.

La gran dispersión que existe entre los datos de un mismo tipo de litología se debe por una parte la gran cantidad de órdenes de magnitud en los que varía la conductividad hidráulica y por otra parte a la existencia de ciertas litologías con diferentes propiedades intrínsecas que se encuentran dentro de un mismo tipo de litología.

Los histogramas presentados en la figura 4.36 muestran el comportamiento de la distribución de los datos de la conductividad hidráulica (m/d). Se puede afirmar

Tabla 4.35: Medidas de dispersión de la conductividad hidráulica (m/d)

Tipo	Rango	IQR	$\sigma_{\log(K)}$	DM	DMd	DMg
Igneous	864	3.277	2.427	43.625	25.791	25.808
Metamorphic	40	0.777	2.253	1.343	0.987	0.997
Sedimentary	86400	9.900	1.992	314.603	173.907	174.010
Unconsolidated	2592	25.903	2.382	114.352	75.743	75.881

Nota:

Nota: Tabla de comparación de las medidas de dispersión de la conductividad hidráulica (m/d) en los diferentes tipos de litologías. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar del logaritmo en base 10 de los datos de conductividad hidráulica ($\sigma_{\log(K)}$). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

que todos los casos presentan un ligero sesgo negativo y forma leptocúrtica.

Para el caso de rocas ígneas las mayores frecuencias se concentran en valores cercanos a 0.1 y 5 (m/d). Las rocas metamórficas tienen alta frecuencia en datos cercanos a 0.5 (m/d). Por su parte las rocas sedimentarias presentan frecuencias elevadas en estimaciones cercanas a 10 (m/d). Finalmente, para los sedimentos no consolidados las mayores frecuencias se encuentran en valores cercanos a 5 y 50 (m/d).

Correlación conductividad hidráulica

La tabla de correlaciones (4.37) muestra los diferentes modelos de regresión de la conductividad hidráulica en función de las diferentes propiedades hidrogeológicas (porosidad total, porosidad efectiva, coeficiente de almacenamiento, rendimiento específico y compresibilidad). Se puede concluir que no existe un modelo lineal, aun con métodos linealizables mediante transformación de datos, que explique un porcentaje importante de la proporción de la varianza de la variable dependiente, ya que los coeficientes de determinación son muy bajos, indicando un modelo de regresión pobre, producto de una baja relación entre las variables.

4.3.3.1 Rocas ígneas

La conductividad hidráulica (m/d) en rocas ígneas varía en un gran rango de órdenes de magnitud, la figura 4.38 muestra los diagramas de caja de la conductividad hidráulica para las rocas más representativas del grupo. Para el caso del basalto, toba, dolerita y granito se observa un mayor volumen de datos atípicos, siendo para el basalto los valores mayores a 15 (m/d); para el granito y dolerita los mayores a 5 (m/d); y para la toba los mayores a 3 (m/d).

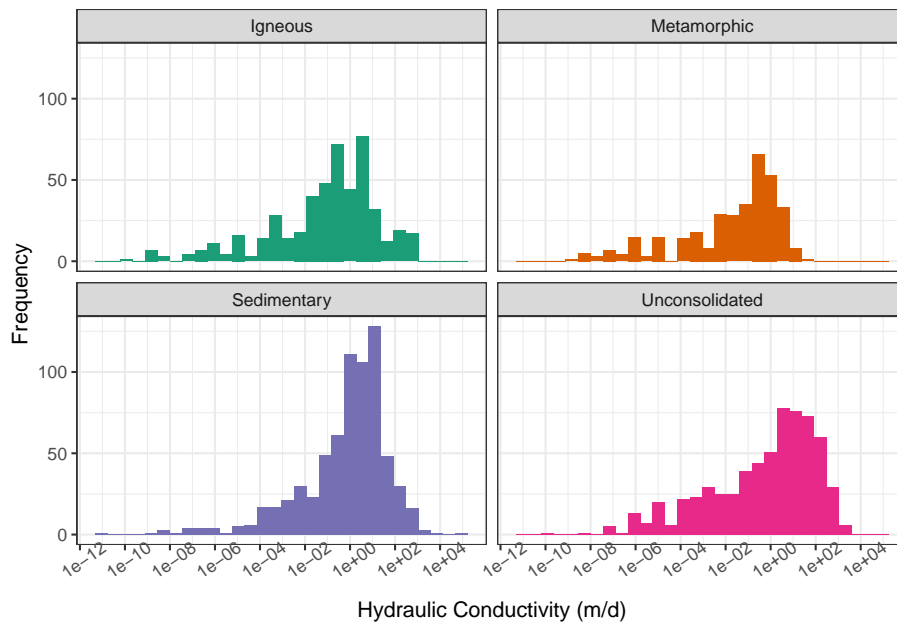


Figura 4.36: Histogramas en escala logarítmica en base 10 de la conductividad hidráulica (m/d) para las diferentes clasificaciones litológicas. Histogramas contruidos a partir de 30 clases.

En todos los tipos de litología se muestra una tendencia muy parecida entre los diagramas de caja, con medianas que varían entre 0.1 y 1 (m/d), a excepción de la dolerita cuya mediana es menor que 0.1 (m/d).

La presencia de valores atípicos dentro de una misma litología se debe al grado de fracturamiento e intemperismo que presentan las rocas. Un ejemplo de registros extremos de conductividad hidráulica son los basaltos fracturados de Maui, Hawái donde se estiman conductividades hidráulicas de más de 500 (m/d) (Rotzoll et al., 2007). Otro ejemplo son las andesitas fracturadas de Jeju Island, Korea, donde Won et al. (2006) estima más de 100 (m/d) de conductividad hidráulica en este tipo de rocas.

La tabla 4.36 muestra una comparación de los valores mínimos y máximos, así como las medianas y cuartiles de diferentes rocas ígneas, donde se puede observar que rocas fracturadas como el basalto alcanzan el valor más alto de todas las rocas ígneas con 864 (m/d), mientras que el valor más bajo está presente en el granito con 1×10^{-10} .

En cuanto al análisis de la tendencia de las medianas, se puede apreciar que ninguna roca supera 1 (m/d), siendo la toba la que mayor media registra con 0.86 (m/d), a diferencia de la dolerita que tiene la menor mediana con 0.051

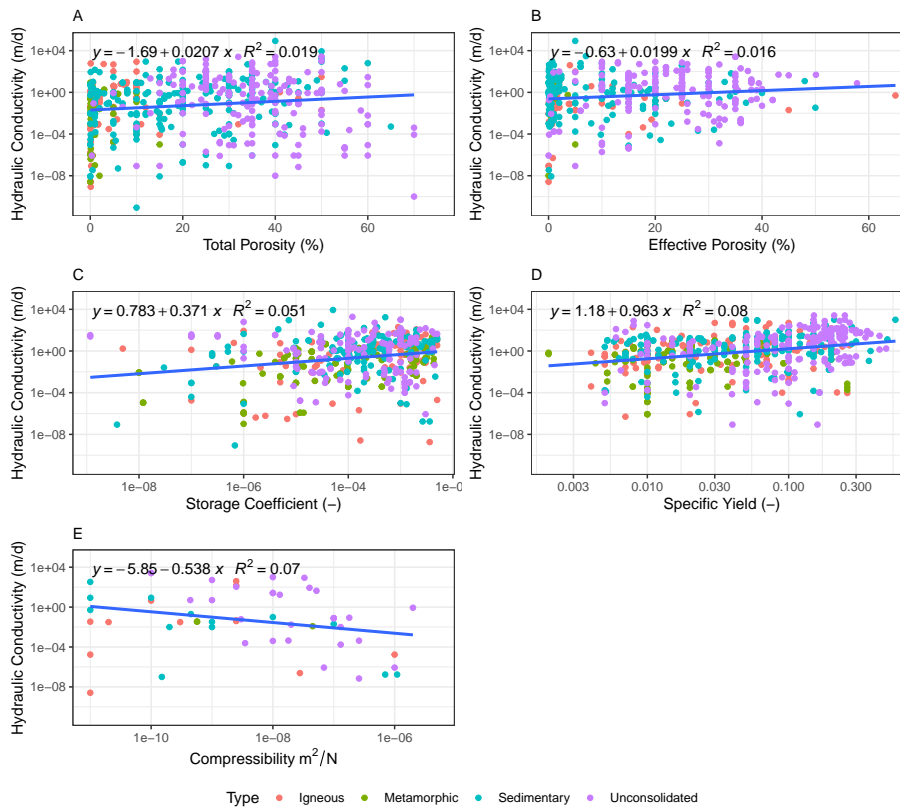


Figura 4.37: Diagramas de correlación de la conductividad hidráulica en función de las diferentes propiedades hidrogeológicas. A: Conductividad hidráulica vs Porosidad Total (transformación logarítmica en base 10 del eje y). B: Conductividad Hidráulica vs Porosidad Efectiva (transformación logarítmica en base 10 del eje y). C: Conductividad hidráulica vs Coeficiente de Almacenamiento (transformación logarítmica en base 10 del eje y, y el eje x). D: Conductividad hidráulica vs Rendimiento Específico (transformación logarítmica en base 10 del eje y y el eje x); E: Conductividad hidráulica vs Compresibilidad (transformación logarítmica en base 10 del eje y y eje x)

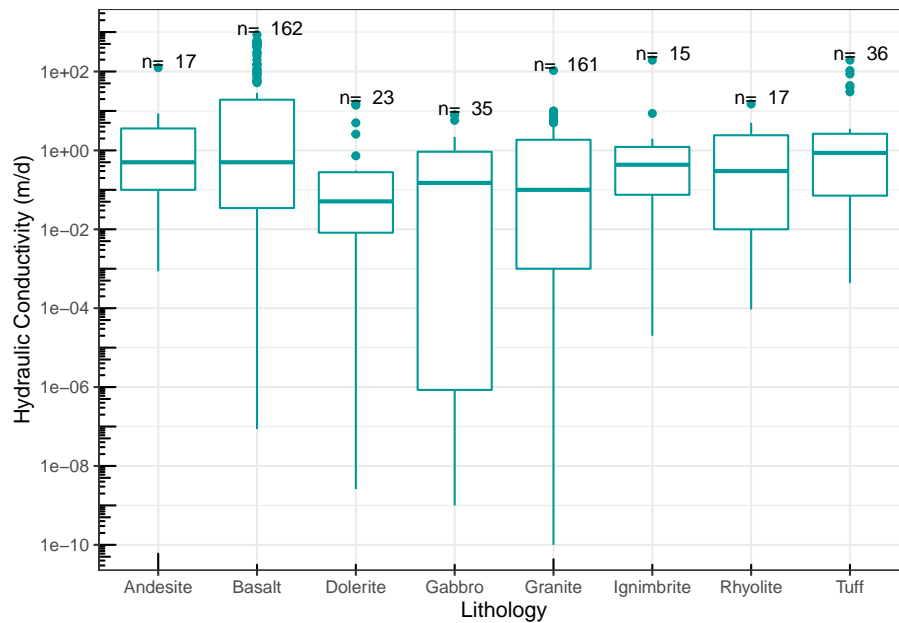


Figura 4.38: Diagramas de caja en escala logarítmica en base 10 de la conductividad hidráulica (m/d) de diferentes rocas ígneas.

m/d.

Las medidas de tendencia central se presentan en la tabla 4.37. En todos los casos se puede apreciar que la media aritmética se diferencia por mucho de la mediana y la media geométrica. La media aritmética es muy sensible a la presencia de datos anómalos por lo que no es un buen indicador en este tipo de rocas; ya que la presencia de rocas fracturadas puede aumentar en gran medida los registros de la conductividad hidráulica, generando valores atípicos.

La tabla 4.38 muestra las diferentes medidas de dispersión de los valores de conductividad hidráulica (m/d). Analizando las desviaciones estándar del logaritmo de los datos de la conductividad hidráulica se puede inferir que el gabbro es la roca que tiene una menor dispersión entre sus datos, por el contrario, el basalto es la roca que tiene el mayor rango y la mayor dispersión entre sus datos.

Los valores de dispersión tan elevados son resultado de la presencia de datos anómalos y datos extremos producto de rocas extremadamente fracturadas e intemperizadas dentro de los datos de una misma roca que aumentan en la conductividad hidráulica varios órdenes de magnitud.

El análisis de la comparación de las desviaciones medias absolutas de un punto central, de los datos de la conductividad hidráulica en rocas ígneas, muestra que

Tabla 4.36: Valores mínimos, máximos y cuartiles de la conductividad hidráulica (m/d)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Andesite	17	8.64e-04	1.00e-01	0.500	3.600	124.9
Basalt	162	8.64e-08	3.46e-02	0.500	19.275	864.0
Dolerite	23	2.59e-09	8.20e-03	0.051	0.280	15.0
Gabbro	35	1.00e-09	8.47e-07	0.150	0.925	8.0
Granite	161	1.00e-10	1.00e-03	0.100	1.860	106.0
Ignimbrite	15	2.00e-05	7.50e-02	0.432	1.225	194.4
Rhyolite	17	9.30e-05	1.00e-02	0.300	2.420	15.0
Tuff	36	4.32e-04	7.13e-02	0.864	2.643	194.4

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la conductividad hidráulica (m/d) en diferentes rocas ígneas. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.37: Comparación de las medidas de tendencia central de la conductividad hidráulica (m/d)

Litología	Media	Media_geométrica	Mediana
Andesite	9.366	0.504	0.500
Basalt	69.639	0.351	0.500
Dolerite	1.682	0.021	0.051
Gabbro	1.018	0.002	0.150
Granite	2.086	0.020	0.100
Ignimbrite	13.969	0.202	0.432
Rhyolite	1.952	0.096	0.300
Tuff	14.619	0.397	0.864

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la conductividad hidráulica (m/d) en diferentes rocas ígneas.

Tabla 4.38: Medidas de dispersión de la conductividad hidráulica (m/d)

Litología	Rango	IQR	$\sigma_{\log(K)}$	DM	DMd	DMg
Andesite	124.899	3.500	1.259	13.592	9.223	9.224
Basalt	864.000	19.240	2.377	104.749	69.545	69.553
Dolerite	15.000	0.272	2.219	2.597	1.671	1.673
Gabbro	8.000	0.925	3.270	1.297	1.006	1.017
Granite	106.000	1.859	2.520	2.756	2.071	2.079
Ignimbrite	194.400	1.150	1.629	24.057	13.857	13.879
Rhyolite	15.000	2.410	1.648	2.384	1.914	1.926
Tuff	194.400	2.571	1.543	22.942	14.430	14.466

Nota:

Tabla de comparación de las medidas de dispersión de la conductividad hidráulica (m/d) en diferentes rocas ígneas. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar del logaritmo en base 10 de los datos de la conductividad hidráulica (σ_K). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (Dmg).

en todos los casos expuestos la desviación de los datos respecto a la mediana resulta ser menor que la desviación respecto a la media geométrica.

Los histogramas de la figura 4.39 son muy diferentes para cada litología, sin embargo, coinciden en ser distribuciones unimodales con sesgo positivo. Para el caso de la andesita, granito, ignimbrita, gabro y toba las mayores frecuencias se concentran entre 1 y 2 m/d; mientras que la riolita y la dolerita concentran la mayor de sus frecuencias registros cercanos a 0.1 m/d. Finalmente el basalto es la roca que presenta una mayor moda, en donde la mayor frecuencia está en estimaciones cercanas a 10 m/d.

4.3.3.2 Rocas metamórficas

Las rocas metamórficas registran el menor rango en órdenes de magnitud de la conductividad hidráulica (m/d), variando en menos de 10 órdenes de magnitud. En la figura 4.40 se muestran los diagramas de caja de la conductividad hidráulica para las rocas metamórficas más representativas. Rocas como la anfibolita, mármol, filita y esquistos no superan 1 (m/d) de conductividad hidráulica a excepción de datos anómalos. El gneis y la cuarcita registran valores anómalos en estimaciones mayores a 2 (m/d).

Las estimaciones elevadas de la conductividad hidráulica son producto de registros de rocas extremadamente fracturadas e intemperizadas dentro de la base de datos. Como ejemplo se tienen los trabajos de Chinnasamy et al.

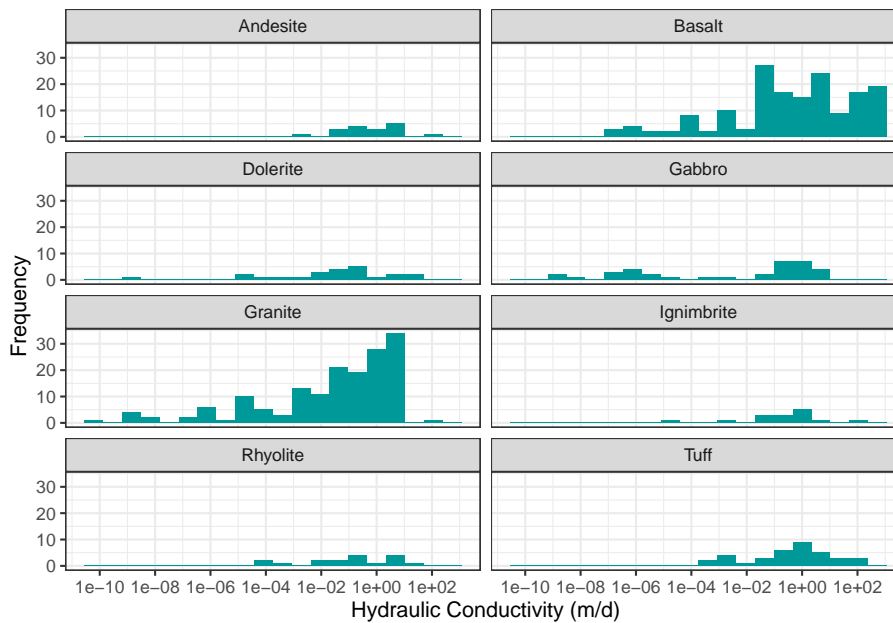


Figura 4.39: Histogramas en escala logarítmica en base 10 de la conductividad hidráulica (m/d) en diferentes rocas ígneas. Histogramas construidos a partir de 20 clases

(2018) donde se registran conductividades hidráulicas de más de 40 (m/d) en complejos metamórficos fracturados compuestos por filita, gneis, esquisto y cuarcita, en ciertas regiones de Rajasthan, India. Otro ejemplo son los complejos metamórficos compuestos de filitas, granulitas, esquistos y pizarras de Dharwars, India donde se registran conductividades hidráulicas de hasta 10 (m/d) (Sharma and Kumar, 2008).

Medidas puntuales de valores máximos, mínimos y cuartiles se presentan en la tabla 4.39. El gneis y el mármol son las rocas con los valores máximos más altos con 44 y 21.19 (m/d) respectivamente; por su parte el esquisto y la anfibolita presentan los valores mínimos más bajos, con estimaciones menores al 8.6×10^{-7} (m/d).

Los valores de las medianas tienden a ser menores a 1 (m/d) en todos los casos, siendo la pizarra la roca que tiene una mayor estimación de la mediana con 0.256 (m/d) y la anfibolita la que tiene un menor registro de la mediana con 0.067 m/d.

La tabla 4.40 muestra la comparación de las medidas de tendencia central de la conductividad hidráulica en diferentes rocas metamórficas. Se puede inferir que la media aritmética es la que más se diferencia de las otras dos medidas en respuesta a su sensibilidad a valores anómalos que afectan en mayor medida su

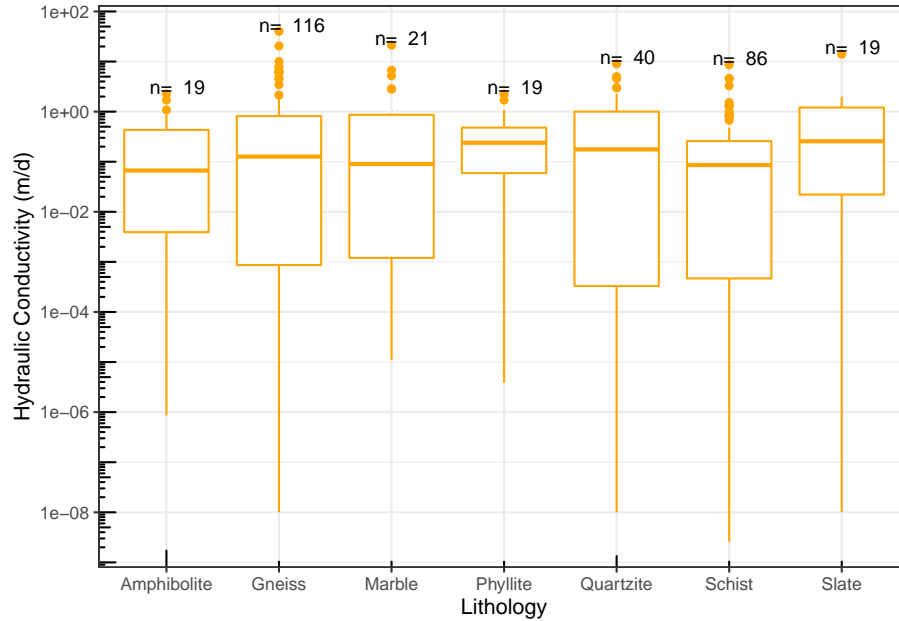


Figura 4.40: Diagramas de caja en escala logarítmica en base 10 de la conductividad hidráulica (m/d) de diferentes rocas metamórficas.

Tabla 4.39: Valores mínimos, máximos y cuartiles de la conductividad hidráulica (m/d)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Amphibolite	19	8.64e-07	3.92e-03	0.067	0.432	2.24
Gneiss	116	1.00e-08	8.64e-04	0.127	0.816	40.00
Marble	21	1.10e-05	1.20e-03	0.090	0.864	21.19
Phyllite	19	3.89e-06	5.92e-02	0.239	0.480	2.24
Quartzite	40	1.00e-08	3.30e-04	0.176	1.000	9.00
Schist	86	2.59e-09	4.67e-04	0.086	0.258	8.64
Slate	19	1.00e-08	2.21e-02	0.256	1.207	14.00

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la conductividad hidráulica (m/d) en diferentes rocas metamórficas. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.40: Comparación de las medidas de tendencia central de la conductividad hidráulica (m/d)

Litología	Media	Media_geométrica	Mediana
Amfibolite	0.405	0.037	0.06670
Gneiss	1.333	0.016	0.12653
Marble	1.978	0.050	0.09000
Phyllite	0.456	0.076	0.23900
Quartzite	1.072	0.017	0.17620
Schist	0.444	0.006	0.08583
Slate	1.289	0.042	0.25600

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la conductividad hidráulica (m/d) en diferentes rocas metamórficas.

resultado.

En la tabla 4.41 se muestran las diferentes medidas de dispersión de los datos de la conductividad hidráulica (m/d) en rocas metamórficas. Analizando la desviación estándar del logaritmo de los datos de la conductividad hidráulica, se puede inferir que la filita es la roca que presenta una menor dispersión entre sus datos, por el contrario, la cuarcita es la litología que presenta una mayor dispersión en comparación con el resto de las rocas metamórficas analizadas.

En el análisis de las desviaciones medias absolutas de un punto central se puede concluir que, en todos los casos de rocas metamórficas analizados, la desviación de los datos respecto a la mediana es menor en comparación con las desviaciones respecto a la media aritmética y media geométrica, siendo la desviación respecto a la media aritmética mayor en comparación con las otras dos desviaciones medias.

Los valores elevados de dispersión en la conductividad hidráulica en rocas metamórficas se deben a la presencia de rocas del mismo tipo dentro de la base de datos que tienen diferentes grados de fracturamiento e intemperismo, propiedades que aumentan en gran medida las estimaciones de la conductividad hidráulica en varios órdenes de magnitud.

Los histogramas de la figura 4.41 describen la distribución de datos de conductividad hidráulica (m/d) para las diferentes rocas metamórficas. En todos los casos se observa un sesgo negativo. Las rocas como el gneis, esquisto, filita y cuarcita tienen distribuciones unimodales con las mayores frecuencias entre 0.1 y 1 (m/d); por su parte el mármol y la anfibolita son bimodales, con las mayores frecuencias en los valores cercanos a 0.001 y 0.8 (m/d) para la anfibolita; y en 1×10^{-4} y 0.5 (m/d) para el mármol.

Tabla 4.41: Medidas de dispersión de la conductividad hidráulica (m/d)

Litología	Rango	IQR	$\sigma_{\log(K)}$	DM	DMd	DMg
Amphibolite	2.24	0.428	1.546	0.455	0.392	0.394
Gneiss	40.00	0.815	2.357	1.821	1.315	1.327
Marble	21.19	0.863	1.744	2.752	1.961	1.967
Phyllite	2.24	0.421	1.472	0.440	0.393	0.415
Quartzite	9.00	1.000	2.468	1.296	1.048	1.066
Schist	8.64	0.258	2.377	0.590	0.437	0.442
Slate	14.00	1.185	2.301	1.541	1.223	1.269

Nota:

Tabla de comparación de las medidas de dispersión de la conductividad hidráulica (-) en diferentes rocas metamórficas. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar del logaritmo en base 10 de los datos de la conductividad hidráulica ($\sigma : \log(K)$). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

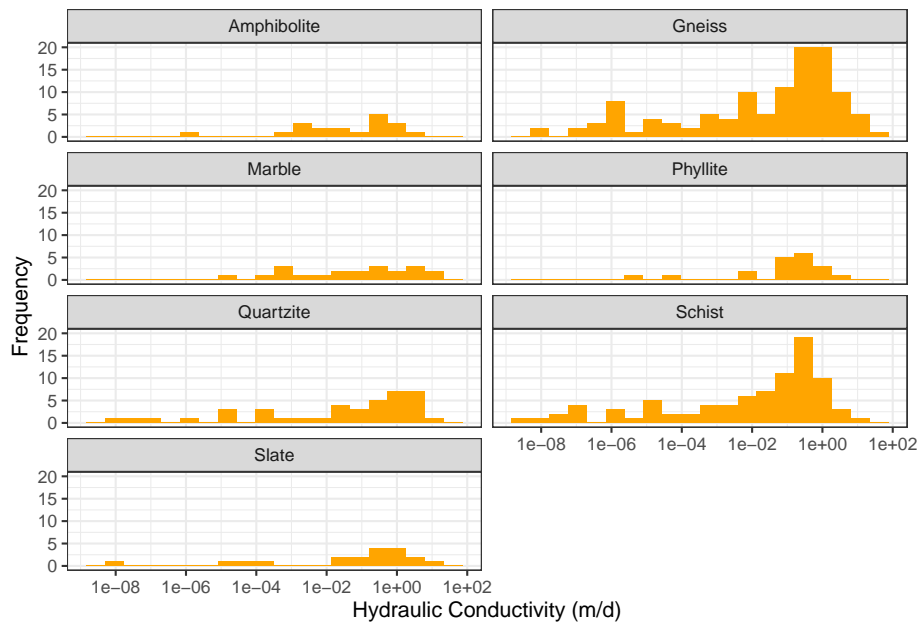


Figura 4.41: Histogramas en escala logarítmica en base 10 de la conductividad hidráulica (m/d) en diferentes rocas metamórficas. Histogramas construidos a partir de 20 clases

4.3.3.3 Rocas sedimentarias

La conductividad hidráulica (m/d) en rocas sedimentarias alcanza valores muy elevados en todos los casos registrados. En la figura 4.42 se muestran los diagramas de caja en una escala logarítmica para las litologías más representativas. En rocas carbonatadas se consideran valores anómalos aquellas estimaciones mayores a 20 (m/d). En rocas como la arenisca y el conglomerado, se consideran estimaciones anómalas aquellos valores mayores a 15 (m/d). Finalmente, en rocas de grano fino como la lutita y la limolita estimaciones mayores a 0.5 (m/d) se consideran anómalas.

La existencia de una gran cantidad de valores anómalos en rocas sedimentarias se debe la presencia de ciertas rocas pertenecientes a una misma litología que tiene diferencias texturales y estructurales importantes que aumentan en gran medida las estimaciones de la conductividad hidráulica, como ejemplo se tienen los carbonatos como la caliza con una mediana de 6 (m/d) que por procesos de disolución aumentan en gran medida los valores de conductividad hidráulica, alcanzando hasta 86400 (m/d) en algunos casos, como los datos registrados por Poehls and Smith (2011).

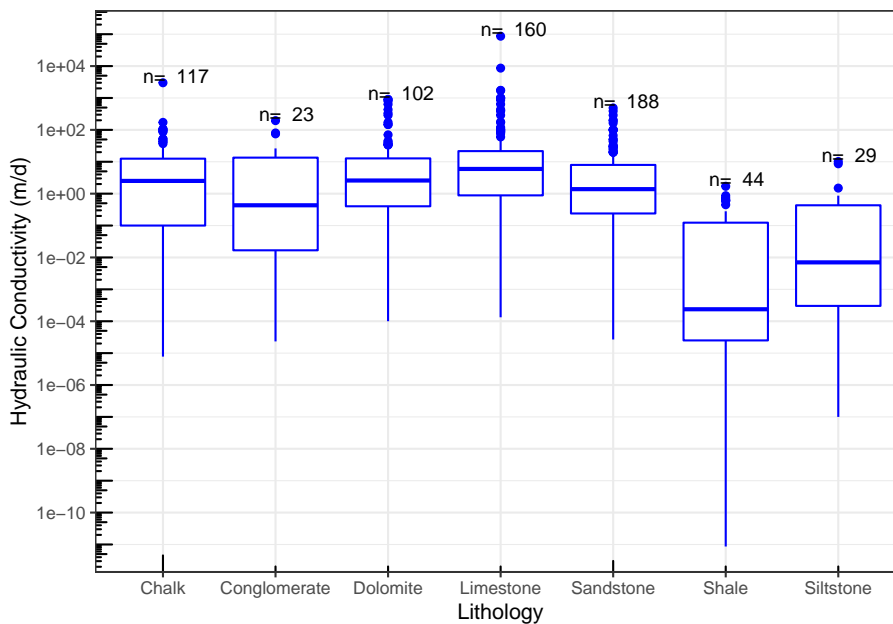


Figura 4.42: Diagramas de caja en escala logarítmica en base 10 de la conductividad hidráulica (m/d) de diferentes rocas sedimentarias.

Los valores puntuales de los diagramas de caja se muestran en la tabla 4.42, donde se puede apreciar que la caliza registra el valor máximo más alto en comparación con las otras rocas, con 86,400 (m/d) en casos anómalos; mientras

Tabla 4.42: Valores mínimos, máximos y cuartiles de la conductividad hidráulica (m/d)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Chalk	117	7.78e-06	1.00e-01	2.5100	12.500	3000.0
Conglomerate	23	2.33e-05	1.68e-02	0.4320	13.485	195.0
Dolomite	102	1.00e-04	4.02e-01	2.5850	12.750	900.0
Limestone	160	1.33e-04	8.83e-01	5.9500	21.600	86400.0
Sandstone	188	2.68e-05	2.39e-01	1.3850	7.963	486.0
Shale	44	8.64e-12	2.50e-05	0.0002	0.123	1.7
Siltstone	29	1.00e-07	3.02e-04	0.0070	0.432	10.2

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la conductividad hidráulica (m/d) en diferentes rocas sedimentarias. La columna **No. datos** corresponde con el número de observaciones.

que la lutita con 8.64×10^{-12} (m/d) registra el valor mínimo más bajo de conductividad hidráulica, comparada con el resto de las rocas sedimentarias registradas.

Analizando el valor de las medianas de cada roca, se estima que la caliza es la roca que tiene un mayor registro de la mediana con 6 (m/d), mientras que la lutita es la que tiene el menor registro con 2×10^{-4} (m/d).

La comparación de medidas de tendencia central de los datos de conductividad hidráulica (m/d) se presenta en la tabla 4.43. En rocas sedimentarias la diferencia entre la media aritmética y las otras dos medidas de tendencia central es bastante elevada. En los carbonatos este fenómeno está más marcado, por la presencia de valores anómalos tan elevados que afectan en gran medida la estimación de la media aritmética.

La tabla 4.44 presenta la comparación de las medidas de dispersión de los datos de la conductividad hidráulica (m/d) en rocas sedimentarias. Se puede inferir de acuerdo con los valores de la desviación estándar del logaritmo de los datos de la conductividad hidráulica que la arenisca es la roca que tiene una menor dispersión entre sus datos en comparación con las otras litologías registradas; por el contrario, la lutita es la que presenta una mayor dispersión entre sus datos registrados de conductividad hidráulica en comparación con las otras rocas.

En el análisis de las desviaciones medias absolutas de un punto central se infirió que todas las rocas sedimentarias registradas en la base de datos tienden a tener una menor desviación de sus datos respecto a la mediana en comparación con las desviaciones respecto a la media aritmética y media geométrica.

La gran dispersión entre los datos, sobre todo en carbonatos y en caso especial en la caliza, se debe a la existencia de valores anómalos y valores extremos

Tabla 4.43: Comparación de las medidas de tendencia central de la conductividad hidráulica (m/d)

Litología	Media	Media_geométrica	Mediana
Chalk	38.197	0.8057	2.5100
Conglomerate	19.186	0.3143	0.4320
Dolomite	53.093	1.7920	2.5850
Limestone	668.009	4.5569	5.9500
Sandstone	16.226	1.0522	1.3850
Shale	0.155	0.0004	0.0002
Siltstone	0.848	0.0047	0.0070

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la conductividad hidráulica (m/d) en diferentes rocas sedimentarias.

Tabla 4.44: Medidas de dispersión de la conductividad hidráulica (m/d)

Litología	Rango	IQR	$\sigma_{\log(K)}$	DM	DMd	DMg
Chalk	3000.0	12.400	1.721	58.654	37.609	37.840
Conglomerate	195.0	13.468	2.043	26.285	19.075	19.087
Dolomite	900.0	12.348	1.413	83.132	52.467	52.497
Limestone	86400.0	20.716	1.282	1212.699	666.531	666.585
Sandstone	486.0	7.723	1.285	23.375	15.809	15.824
Shale	1.7	0.123	2.726	0.216	0.155	0.155
Siltstone	10.2	0.432	2.435	1.228	0.846	0.847

Nota:

Tabla de comparación de las medidas de dispersión de la conductividad hidráulica (m/d) en diferentes rocas sedimentarias. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar del logaritmo en base 10 de los datos de la conductividad hidráulica ($\sigma_{\log(K)}$). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

que afectan en gran medida la distribución. Los casos extremos y anómalos se deben a propiedades intrínsecas y procesos secundarios que afectan a las rocas produciendo aumentos significativos en los valores de la conductividad hidráulica, la cual varía en una gran cantidad de órdenes de magnitud.

Los histogramas de la figura 4.43 muestran la distribución de los datos de la conductividad hidráulica (m/d) de las diferentes rocas sedimentarias. En todos los casos a excepción del conglomerado los histogramas tienen un sesgo negativo

y son unimodales, con las mayores frecuencias en valores cercanos a 10 (m/d) para la Creta; 15 (m/d) para la dolomita; 30 (m/d) para la caliza; 1 (m/d) para la arenisca y lutita; y 1×10^{-4} (m/d) para la lutita. El conglomerado es muy diferente en cuanto a la distribución de sus datos y tiende a ser bimodal con las mayores frecuencias en 1×10^{-4} y 1 (m/d).

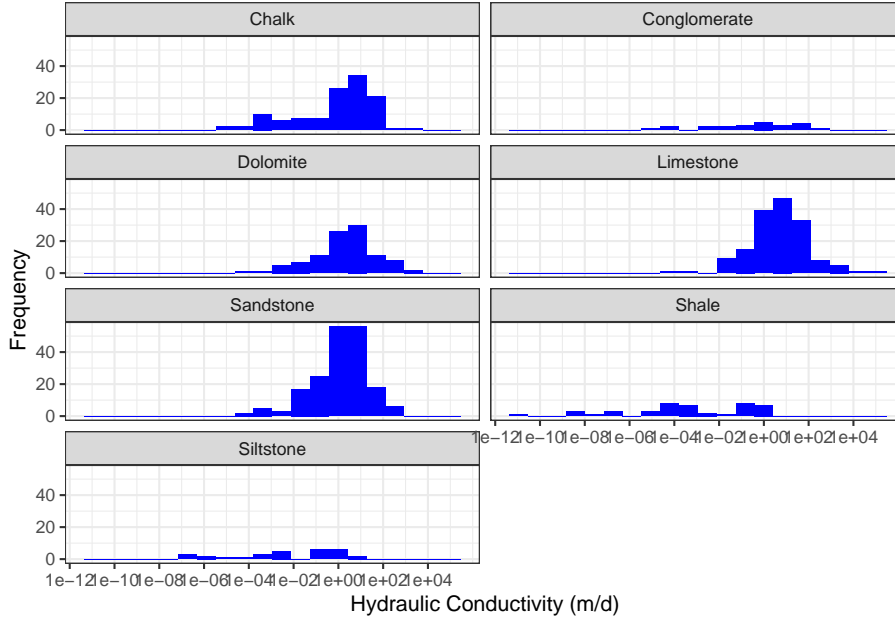


Figura 4.43: Histogramas en escala logarítmica en base 10 de la conductividad hidráulica (m/d) en diferentes rocas sedimentarias. Histogramas construidos a partir de 20 clases

4.3.3.4 Sedimentos no consolidados

Los sedimentos no consolidados se caracterizan por tener estimaciones bastante altas de conductividad hidráulica (m/d). En la figura 4.44 se muestran los diagramas de caja de los principales sedimentos no consolidados registrados en la base de datos. Se puede observar que, a diferencia de las rocas consolidadas, los sedimentos varían en gran medida, respecto a su composición y textura.

Analizando las medianas en la figura 4.44 se puede apreciar que éstas disminuyen en relación con el tamaño de grano de los diferentes sedimentos siendo las gravas las que tiene un mayor valor de la mediana, siguiendo por las arenas, limos y arcillas.

Sedimentos como las arcillas, sedimentos de origen glacial y la combinación de limos y arcillas, presentan valores anómalos en aquellas estimaciones mayores a 0.5 (m/d); en el caso de las gravas las estimaciones mayores a 1000 (m/d)

son anómalas; para las arenas y la combinación de arena y grava, estimaciones mayores a 100 (m/d) se aprecian como anómalas. En los limos aquellas mediciones mayores a 1 (m/d) son anómalas y para la combinación de arenas, limos y arcillas se consideran anómalos aquellos valores mayores a 20 (m/d). Finalmente, en las arenas arcillosas no se presentan valores anómalos en este análisis.

Ejemplos de valores extremos se dan principalmente en gravas y en las combinaciones de gravas y arenas, un ejemplo son los datos expuestos por Domenico et al. (1998) mostrando conductividades hidráulicas mayores a 2000 (m/d) en gravas.

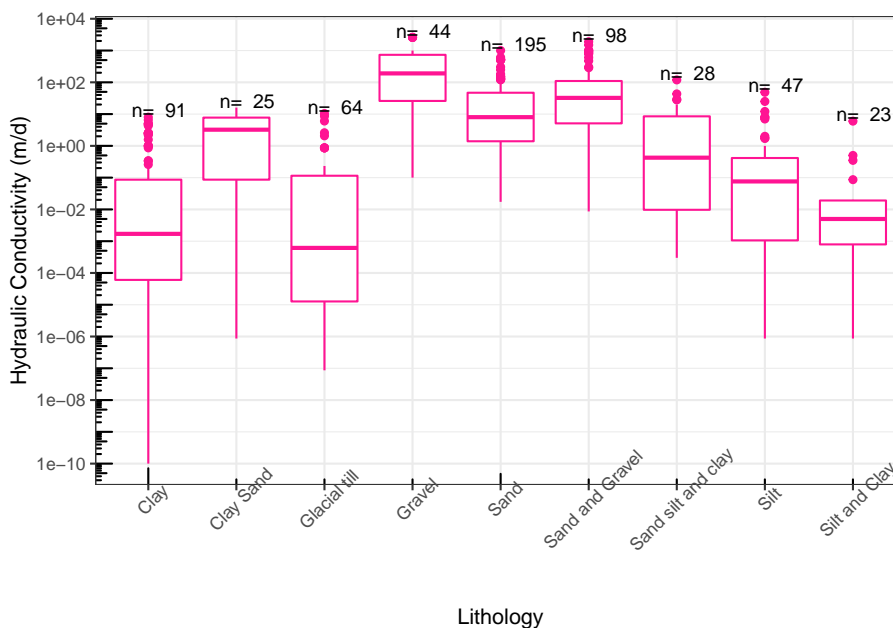


Figura 4.44: Diagramas de caja en escala logarítmica en base 10 de la conductividad hidráulica (m/d) de diferentes sedimentos no consolidados.

La tabla 4.45 presenta un resumen de los datos de los diagramas de caja. En esta tabla se puede observar que las gravas son los sedimentos que pueden llegar a alcanzar los registros máximos más altos con 2592 (m/d), en comparación con los otros sedimentos; mientras que el valor mínimo más bajo registrado es 1×10^{-10} (m/d), correspondiente con las arcillas.

La mayor estimación de la mediana la tiene la grava con 192.72 (m/d) y el menor valor lo registra la arcilla con 2.17×10^{-3} (m/d) de conductividad hidráulica.

La tabla 4.46 muestra las comparaciones de las medidas de tendencia central de la conductividad hidráulica (m/d) en sedimentos no consolidados. Se puede

Tabla 4.45: Valores mínimos, máximos y cuartiles de la conductividad hidráulica (m/d)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Clay	91	1.00e-10	6.05e-05	0.0017	0.086	8.0
Clay Sand	25	8.64e-07	8.64e-02	3.2300	7.750	16.0
Glacial till	64	8.64e-08	1.27e-05	0.0006	0.115	10.4
Gravel	44	1.00e-01	2.59e+01	190.7200	732.885	2592.0
Sand	195	1.73e-02	1.40e+00	8.0000	47.000	1000.0
Sand and Gravel	98	8.64e-03	5.10e+00	32.3500	110.000	1900.0
Sand silt and clay	28	3.00e-04	9.72e-03	0.4250	8.500	119.0
Silt	47	8.64e-07	1.06e-03	0.0760	0.415	48.6
Silt and Clay	23	8.64e-07	7.93e-04	0.0050	0.019	6.0

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la conductividad hidráulica (m/d) en diferentes sedimentos no consolidados. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.46: Comparación de las medidas de tendencia central de la conductividad hidráulica (m/d)

Litología	Media	Media_geométrica	Mediana
Clay	0.609	0.0016	0.0017
Clay Sand	4.284	0.4085	3.2300
Glacial till	0.545	0.0008	0.0006
Gravel	486.658	98.2648	190.7200
Sand	52.873	6.8483	8.0000
Sand and Gravel	149.356	18.5625	32.3500
Sand silt and clay	10.200	0.3092	0.4250
Silt	2.531	0.0252	0.0760
Silt and Clay	0.307	0.0041	0.0050

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la conductividad hidráulica (m/d) en diferentes sedimentos no consolidados.

inferir que la media aritmética tiende a ser más elevada que la media geométrica y la mediana, lo que refleja la sensibilidad que tiene el cálculo de la media aritmética a los valores anómalos presentes y al número de magnitudes en los que se presenta la conductividad hidráulica.

Las medidas de dispersión de los datos normalizados de la conductividad hidráulica se presentan en la tabla 4.47. Analizando la desviación estándar del

Tabla 4.47: Medidas de dispersión de la conductividad hidráulica (m/d)

Litología	Rango	IQR	$\sigma_{\log(K)}$	DM	DMd	DMg
Clay	8.000	0.086	2.334	0.951	0.608	0.608
Clay Sand	16.000	7.664	1.786	3.987	3.859	4.120
Glacial till	10.400	0.115	2.206	0.885	0.545	0.545
Gravel	2591.900	706.965	1.051	515.965	443.708	447.883
Sand	999.983	45.605	1.077	65.274	50.641	50.673
Sand and Gravel	1899.991	104.895	1.159	185.700	141.626	142.435
Sand silt and clay	119.000	8.490	1.656	13.443	10.123	10.126
Silt	48.600	0.414	1.920	3.979	2.522	2.524
Silt and Clay	6.000	0.018	1.436	0.516	0.305	0.305

Nota:

Tabla de comparación de las medidas de dispersión de la conductividad hidráulica (m/d) en diferentes sedimentos no consolidados. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar del logaritmo en base 10 de los datos de la conductividad hidráulica ($\sigma_{\log(K)}$). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

logaritmo de la conductividad hidráulica se puede concluir que las gravas a pesar del gran rango de valores que presentan tienen una menor dispersión entre los datos que componen la muestra, en comparación de los otros sedimentos, por el contrario las arcillas son los sedimentos que tienen la mayor dispersión entre sus datos.

Con el análisis de las desviaciones medias absolutas respecto a un punto central se puede concluir que para todos los tipos de sedimento analizados la desviación de los datos respecto a la mediana resulta ser la menor en comparación con las desviaciones respecto a la media aritmética y media geométrica. También se puede inferir que en todos los tipos de sedimento analizados la desviación respecto a la mediana es más parecida a la desviación respecto a la media geométrica, a diferencia de la media aritmética que es la que más difiere de las otras dos medidas de dispersión.

Los altos valores de medidas de dispersión en especial en casos como las arenas, gravas y combinación de arena y gravas se deben a la gran cantidad de órdenes de magnitud en los que se maneja la conductividad hidráulica en sedimentos no consolidados.

La distribución de los datos de conductividad hidráulica para cada tipo de sedimento se muestra en la figura 4.45, donde se observa que en los casos arcillas y la combinación de limos y arcillas, tienen histogramas con cierta simetría y la mayor frecuencia en valores cercanos a 1×10^{-2} (m/d) para ambos casos. Para

el caso de la combinación de arenas, limos y arcilla se tiene una distribución bastante diferente a las demás, con un histograma bimodal en donde las mayores frecuencias se presentan en 5×10^{-3} y en 10 (m/d).

Litologías como las gravas, combinación de arenas y gravas, arenas y arena arcillosas tienen histogramas unimodales con un sesgo negativo, en donde las mayores frecuencias se concentran en valores cercanos a: 5 (m/d) para arenas arcillosas; 10 (m/d) para arenas; 100 (m/d) para combinación de arenas y gravas; y 150 (m/d) para gravas.

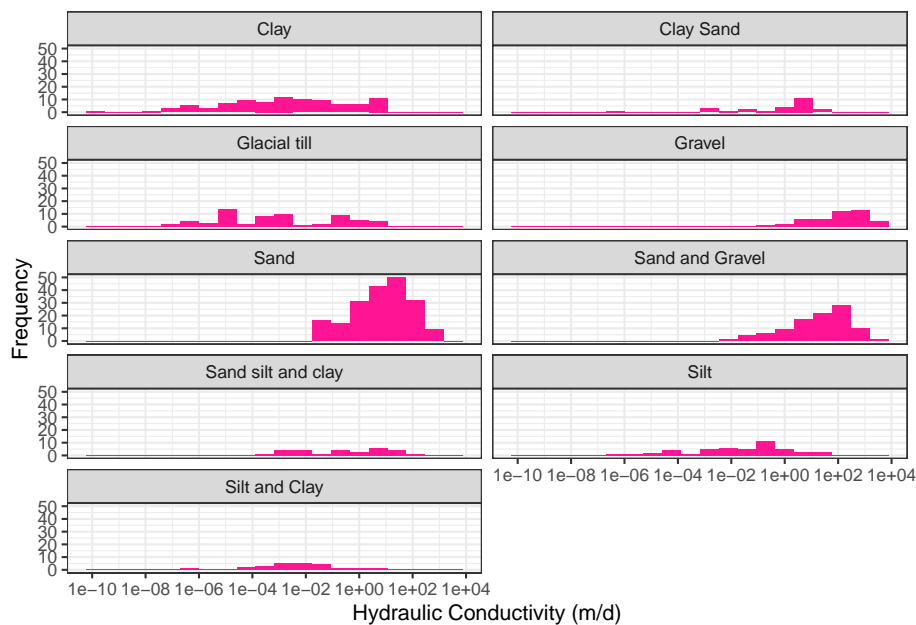


Figura 4.45: Histogramas en escala logarítmica en base 10 de la conductividad hidráulica (m/d) en diferentes sedimentos no consolidados. Histogramas construidos a partir de 20 clases

4.3.4 Coeficiente de almacenamiento

El coeficiente de almacenamiento (-) varía en más de 6 órdenes de magnitud para todos los tipos de litología (figura 4.46), siendo los sedimentos no consolidados los que mayor rango abarcan, desde valores cercanos a 1×10^{-9} (-) hasta superar el 1×10^{-3} (-). Mientras que las rocas metamórficas varían en menor rango desde 1×10^{-8} (-) hasta superar el 1×10^{-3} (-) de coeficiente de almacenamiento.

En la figura 4.47 se muestran los diagramas de rangos del coeficiente de almacenamiento para cada una de las litologías más representativas, en donde se puede observar que, en el caso de las rocas ígneas, el granito registra los menores valores del coeficiente de almacenamiento, siendo menores que 1×10^{-8}

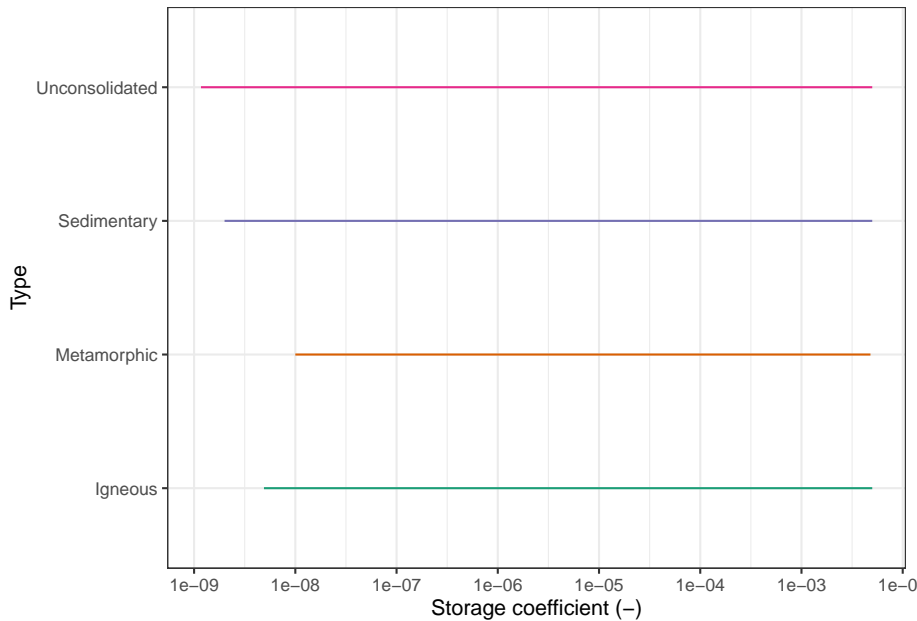


Figura 4.46: Diagrama de rangos en escala logarítmica en base 10 del coeficiente de almacenamiento (-) para los diferentes tipos de litología.

(-). Las rocas metamórficas tienen un comportamiento parecido, con registros mínimos de 1×10^{-8} (-), a excepción de la anfibolita cuyo valor mínimo no es menor que 1×10^{-6} (-).

Los mayores registros de rocas sedimentarias superan 1×10^{-3} (-) en todas las litologías registradas, mientras que los registros mínimos están presentes en rocas como la arenisca y la caliza con estimaciones menores a 1×10^{-7} .

La arena y grava como sedimentos no consolidados son los que mayor rango del coeficiente de almacenamiento presentan, en comparación del resto de rocas y sedimentos, con valores que van desde estimaciones cercanas a 1×10^{-9} (-) hasta valores que superan el 1×10^{-3} (-).

En la figura 4.48 se muestran los diagramas de caja en escala logarítmica del coeficiente de almacenamiento (-) para las diferentes clasificaciones litológicas. En el diagrama de caja se pueden apreciar los valores extremos, siendo en rocas ígneas y metamórficas aquellos valores mayores a 4×10^{-3} (-) considerados anómalos, mientras que en rocas sedimentarias y sedimentos no consolidados se consideran anómalos todas aquellas estimaciones mayores a 3×10^{-3} (-) de coeficiente de almacenamiento.

En la tabla 4.48 se muestran los valores puntuales de los valores máximos y mínimos, así como los cuartiles de los datos del coeficiente de almacenamiento (-)

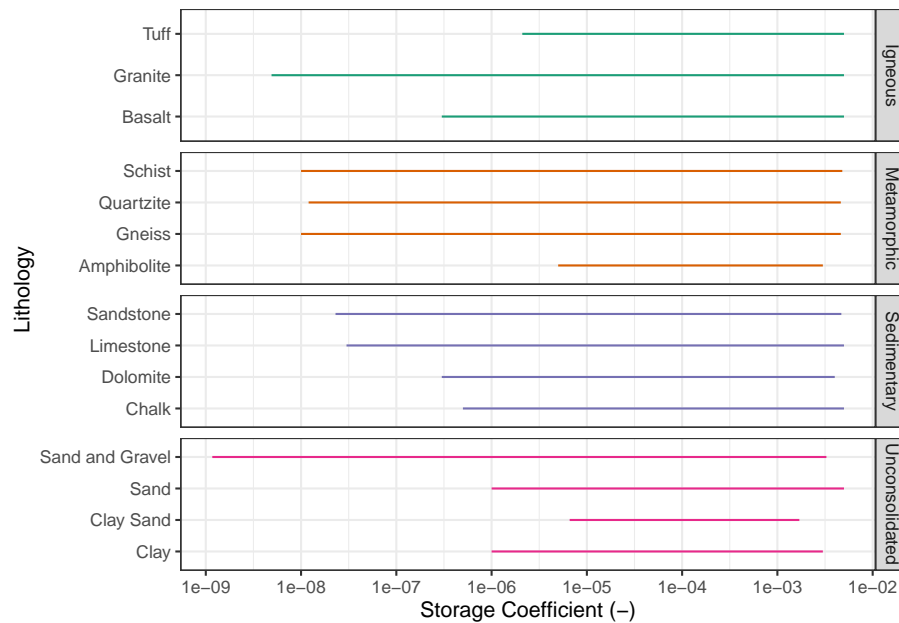


Figura 4.47: Diagrama de rangos en escala logarítmica en base 10 del coeficiente de almacenamiento (-) para diferentes litologías.

Tabla 4.48: Valores mínimos, máximos y cuartiles del coeficiente de almacenamiento (-)

Type	No.Datos	min.	Q1	Mediana	Q3	max.
Igneous	173	4.90e-09	2.30e-05	5.00e-04	2.00e-03	5.00e-03
Metamorphic	230	1.00e-08	1.00e-05	1.46e-04	1.86e-03	4.80e-03
Sedimentary	282	2.00e-09	6.00e-05	2.00e-04	1.00e-03	5.00e-03
Unconsolidated	190	1.17e-09	3.30e-05	2.55e-04	1.00e-03	5.00e-03

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 del coeficiente de almacenamiento (-) para las diferentes clasificaciones litológicas. La columna **No. datos** corresponde con el número de observaciones.

) para cada una de las clasificaciones litológicas. Los sedimentos no consolidados son las litologías que registran los valores mínimos más bajos con 1.17×10^{-9} , en comparación con los otros tipos de litologías. Las rocas ígneas, sedimentarias y sedimentos no consolidados registran estimaciones máximas de 5×10^{-3} . Las rocas ígneas tienen el registro más alto de la mediana con 5×10^{-4} , mientras que las rocas metamórficas registran el menor valor de la mediana con 1.46×10^{-4} .

Una comparación de las diferentes medidas de tendencia central de los datos se

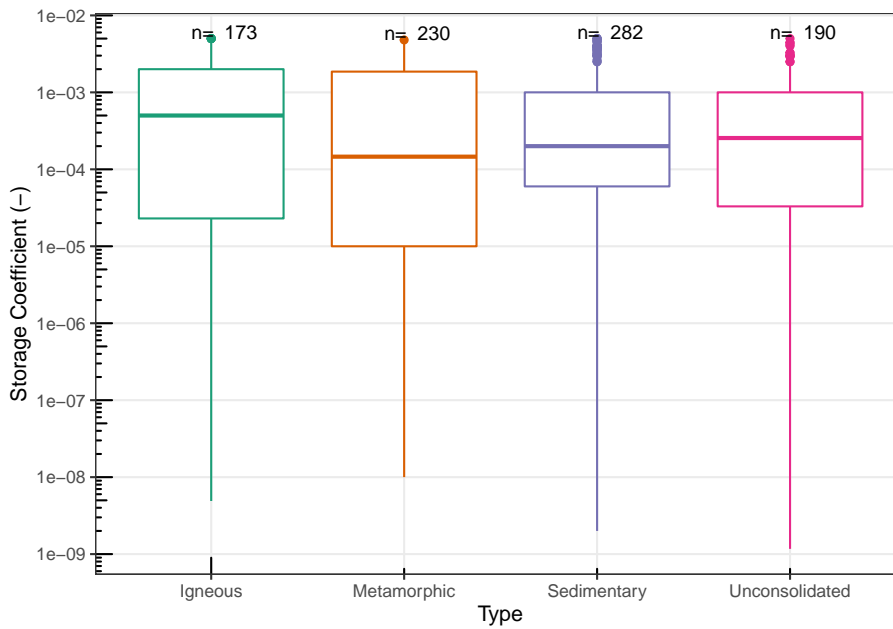


Figura 4.48: Diagramas de caja en escala logarítmica del coeficiente de almacenamiento (-) para las diferentes clasificaciones litológicas.

Tabla 4.49: Comparación de las medidas de tendencia central del coeficiente de almacenamiento (-)

Tipo	Media	Media_geométrica	Mediana
Igneous	1.09e-03	1.54e-04	5.00e-04
Metamorphic	9.78e-04	9.98e-05	1.46e-04
Sedimentary	8.46e-04	1.53e-04	2.00e-04
Unconsolidated	7.28e-04	1.55e-04	2.55e-04

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) del coeficiente de almacenamiento (-) en los diferentes tipos de litologías.

presenta en la tabla 4.49. En rocas ígneas la mediana es más parecida a la media geométrica ya que se encuentran dentro del mismo orden de magnitud. Para el caso de las rocas metamórficas la mediana es más parecida a la media aritmética que a la media geométrica. En rocas sedimentarias y sedimentos no consolidados las tres medidas de tendencia central se encuentran dentro del mismo orden de magnitud y en ambos casos la mediana es más parecida a la media geométrica que a la media aritmética.

Tabla 4.50: Medidas de dispersión del coeficiente de almacenamiento (-)

Tipo	Rango	IQR	σ	DM	DMd	DMg
Igneous	5.00e-03	1.98e-03	1.33e-03	1.08e-03	9.94e-04	1.02e-03
Metamorphic	4.80e-03	1.85e-03	1.36e-03	1.15e-03	9.52e-04	9.52e-04
Sedimentary	5.00e-03	9.40e-04	1.26e-03	9.32e-04	7.76e-04	7.81e-04
Unconsolidated	5.00e-03	9.67e-04	9.89e-04	7.34e-04	6.59e-04	6.66e-04

Nota:

Tabla de comparación de las medidas de dispersión del coeficiente de almacenamiento (-) en los diferentes tipos de litologías. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

Analizando las medidas de la desviación estándar de los datos del coeficiente de almacenamiento (-) de la tabla 4.50 se puede inferir que los sedimentos no consolidados son las litologías que presentan una menor dispersión en sus datos del coeficiente de almacenamiento. Por su parte las rocas metamórficas tienen una mayor dispersión entre sus datos.

Las desviaciones medias absolutas indican que todos los tipos de litologías tienden a tener una menor desviación de sus datos respecto a la mediana en comparación con las desviaciones respecto a la media aritmética y la media geométrica.

Los histogramas en escala logarítmica del coeficiente de almacenamiento para las diferentes clasificaciones de litológicas se presentan en la figura 4.49. Se puede observar que en todos los casos las distribuciones tienen un sesgo negativo, son unimodales y con forma leptocúrtica. Aunque existen histogramas con varios picos indicando varias posibles modas, esto debido a las diferentes litologías dentro de un mismo tipo de litología.

Las rocas ígneas están sesgadas hacia valores altos de coeficiente de almacenamiento y tienen una mayor frecuencia estimaciones cercanas a 1×10^{-3} (-). Las rocas metamórficas tienen un comportamiento muy parecido a las rocas ígneas con el mismo sesgo y la mayor frecuencia en el valor de 5×10^{-3} (-). Por su parte las rocas sedimentarias tienen una mayor frecuencia en la estimación de 1×10^{-4} (-). Finalmente, las mayores frecuencias en sedimentos no consolidados se concentran en los valores cercanos a 1×10^{-3} (-) de coeficiente de almacenamiento.

Correlación Coeficiente de Almacenamiento

La figura 4.50 muestra las diferentes correlaciones, así como los modelos de regresión del coeficiente de almacenamiento en función de las diferentes

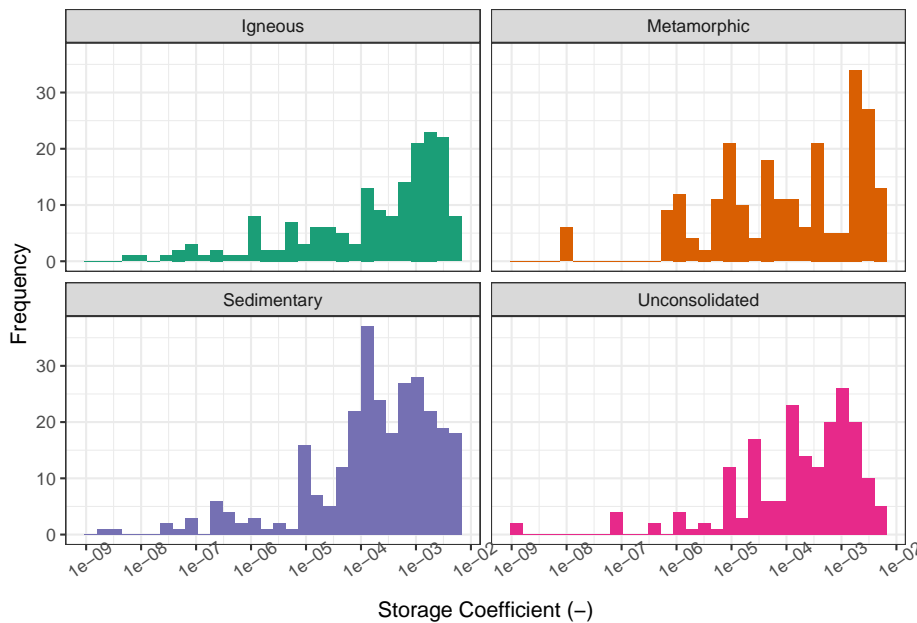


Figura 4.49: Histogramas en escala logarítmica en base 10 del coeficiente de almacenamiento (-) para las diferentes clasificaciones litológicas. Histogramas contruidos a partir de 30 clases

propiedades hidrogeológicas, con el fin de identificar relación entre las variables. Se puede inferir que los modelos de regresión lineal resultan ser muy pobres aun aplicando transformaciones logarítmicas para linealizar los datos.

Un modelo que resalta por tener un mejor ajuste con comparación con los otros casos expuestos en los diagramas de correlaciones es el modelo E (coeficiente de almacenamiento vs compresibilidad) con un coeficiente de determinación de 0.48, no obstante, este modelo también resalta por tener una menor cantidad de datos con los que se realizó el modelo.

4.3.4.1 Rocas ígneas

La figura 4.51 muestra los diagramas de caja en escala logarítmica del coeficiente de almacenamiento (-), en las rocas ígneas más representativas de la base de datos. En esta figura se aprecia que, la existencia de valores anómalos es mínima y se presenta en aquellas estimaciones mayores a 4×10^{-3} en basaltos y granitos. Los ejemplos de valores anómalos de 0.005 (-) registrados para el coeficiente de almacenamiento, corresponden con las rocas volcánicas en Martinica, expuestas por Vittecoq et al. (2015).

El resumen de las medidas puntuales de los diagramas de caja se explica en la

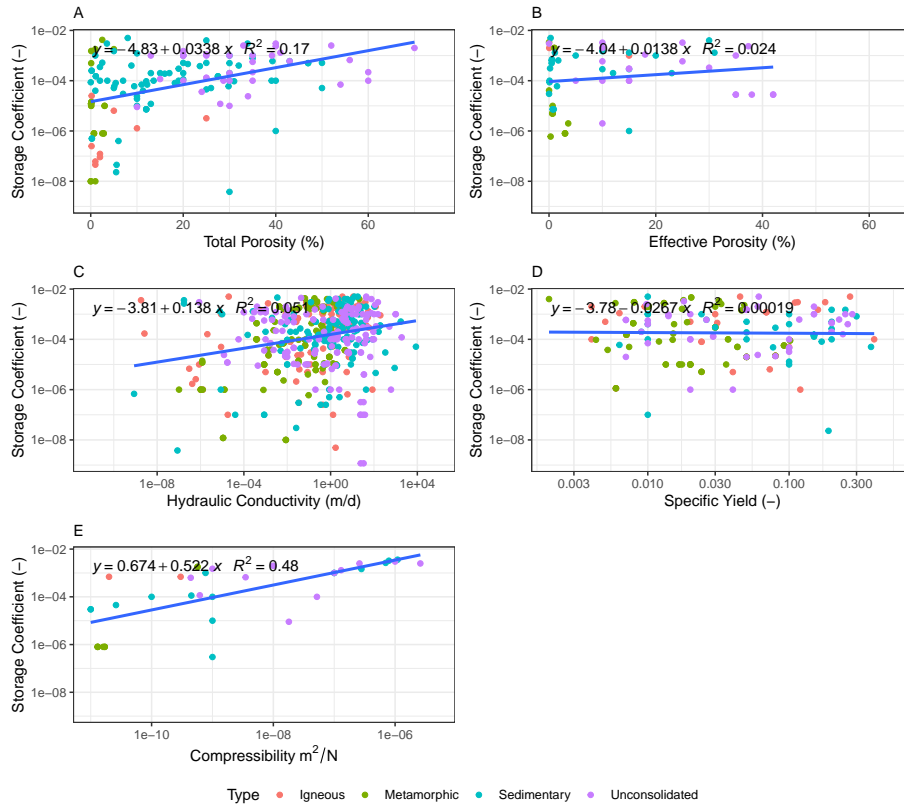


Figura 4.50: Diagramas de correlación del coeficiente de almacenamiento en función de las diferentes propiedades hidrogeológicas. A: Coeficiente de Almacenamiento vs Porosidad Total (transformación logarítmica en base 10 del eje y). B: Coeficiente de Almacenamiento vs Porosidad Efectiva (transformación logarítmica en base 10 del eje y). C: Coeficiente de Almacenamiento vs Conductividad Hidráulica (transformación logarítmica en base 10 del eje y y el eje x). D: Coeficiente de Almacenamiento vs Rendimiento Específico (transformación logarítmica en base 10 del eje y, y el eje x). E: Coeficiente de Almacenamiento vs Compresibilidad (transformación logarítmica en base 10 del eje y y eje x)

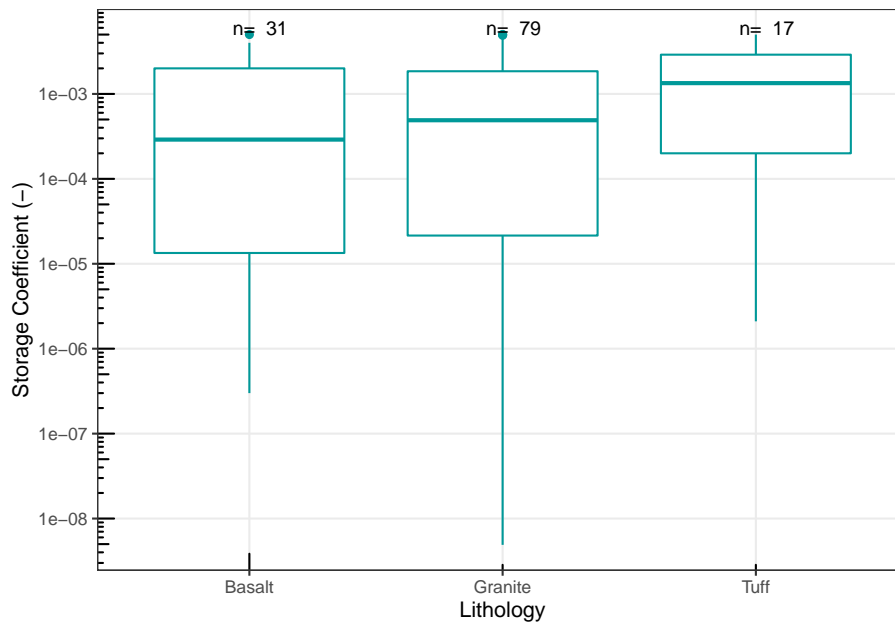


Figura 4.51: Diagramas de caja en escala logarítmica en base 10 del coeficiente de almacenamiento (-) de diferentes rocas ígneas.

Tabla 4.51: Valores mínimos, máximos y cuartiles del coeficiente de almacenamiento (-)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Basalt	31	3.00e-07	1.34e-05	2.90e-04	2.00e-03	5.00e-03
Granite	79	4.90e-09	2.15e-05	4.90e-04	1.85e-03	5.00e-03
Tuff	17	2.10e-06	2.00e-04	1.34e-03	2.90e-03	5.00e-03

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 del coeficiente de almacenamiento (-) en diferentes rocas ígneas. La columna **No. datos** corresponde con el número de observaciones.

tabla 4.51. Los valores máximos del coeficiente de almacenamiento registrados por las tres rocas es 5×10^{-3} (-), mientras que el valor mínimo está presente en el granito con 4.9×10^{-9} . La toba es la roca que presenta el mayor valor de la mediana con 1.34×10^{-3} (-), mientras que el basalto tiene el menor valor de la median en comparación con las otras rocas, con 2.9×10^{-4} (-).

La tabla 4.52 muestra la comparación de las medidas de tenencia central de los datos del coeficiente de almacenamiento (-) en rocas ígneas. En rocas como el basalto y el granito, la mediana tiende a ser más parecida a la media geométrica

Tabla 4.52: Comparación de las medidas de tendencia central del coeficiente de almacenamiento (-)

Litología	Media	Media_geométrica	Mediana
Basalt	1.10e-03	1.56e-04	2.90e-04
Granite	1.09e-03	1.04e-04	4.90e-04
Tuff	1.56e-03	3.94e-04	1.34e-03

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) del coeficiente de almacenamiento (-) en diferentes rocas ígneas.

Tabla 4.53: Medidas de dispersión del coeficiente de almacenamiento (-)

Litología	Rango	IQR	σ	DM	DMd	DMg
Basalt	5.00e-03	1.99e-03	1.36e-03	1.10e-03	1.04e-03	1.05e-03
Granite	5.00e-03	1.83e-03	1.40e-03	1.13e-03	1.02e-03	1.05e-03
Tuff	5.00e-03	2.70e-03	1.46e-03	1.18e-03	1.16e-03	1.37e-03

Nota:

Tabla de comparación de las medidas de dispersión del coeficiente de almacenamiento (-) en diferentes rocas ígneas. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

que a la media aritmética ya que tanto la media geométrica y la mediana se encuentran dentro del mismo orden de magnitud. La toba a diferencia del basalto y el granito tiene una mediana más parecida a la media aritmética que a la media geométrica.

Los coeficientes de variación del coeficiente de almacenamiento (-) (tabla 4.53) muestran que la toba es la roca que tiene una mayor dispersión entre sus datos, mientras que el basalto tiene la menor dispersión entre sus datos del coeficiente de almacenamiento. Analizando las desviaciones medias absolutas de un punto central se infiere que en los tres casos la desviación de los datos respecto a la mediana resulta ser la menor.

Analizando los histogramas del coeficiente de almacenamiento (-) en rocas ígneas (figura 4.52) se infieren distribuciones unimodales con sesgos negativos. En los tres casos las mayores frecuencias se concentran en datos entre 1×10^{-3} y 2×10^{-3} (-) de coeficiente de almacenamiento.

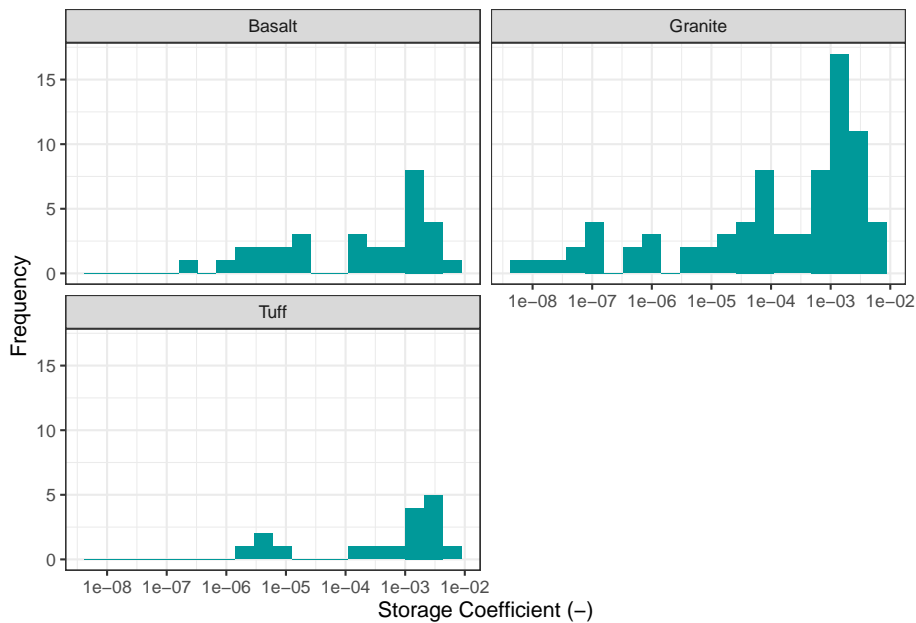


Figura 4.52: Histogramas en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en rocas ígneas. Histogramas construidos a partir de 20 clases

4.3.4.2 Rocas metamórficas

En la figura 4.53 se presentan los diagramas de caja del coeficiente de almacenamiento (-) en escala logarítmica para diferentes rocas metamórficas. Analizando la gráfica de la figura se puede apreciar la presencia de valores anómalos en rocas como la anfibolita y el gneis, siendo a aquellas estimaciones mayores a 1×10^{-3} (-) consideradas anómalas. Por su parte la cuarcita y el esquisto no muestran valores anómalos en este análisis. En rocas metamórficas la presencia de valores elevados de coeficiente de almacenamiento se debe a algunas propiedades texturales como la foliación.

Algunos ejemplos de estimaciones anómalas del coeficiente de almacenamiento con más de 0.001 (-) están presentes en estudios de agua subterránea en terrenos metamórficos en Tamil Nadu, India (Kumar et al., 2016).

La tabla 4.54 muestra el resumen de los diagramas de caja. El esquisto es la roca que registra el valor máximo de 4.8×10^{-3} (-), siendo el más alto en comparación con las otras rocas metamórficas, mientras que el valor más bajo lo registran el esquisto y el gneis con 1×10^{-8} (-). El valor más alto de la mediana del coeficiente de almacenamiento está presente en el esquisto con 7.5×10^{-4} (-), mientras que el valor más bajo de la mediana lo registra el gneis con 4.1×10^{-5} .

Las medidas de tendencia central de los datos del coeficiente de almacenamiento (-) se presentan en la tabla 4.55. Se puede apreciar que la anfibolita tiene medidas

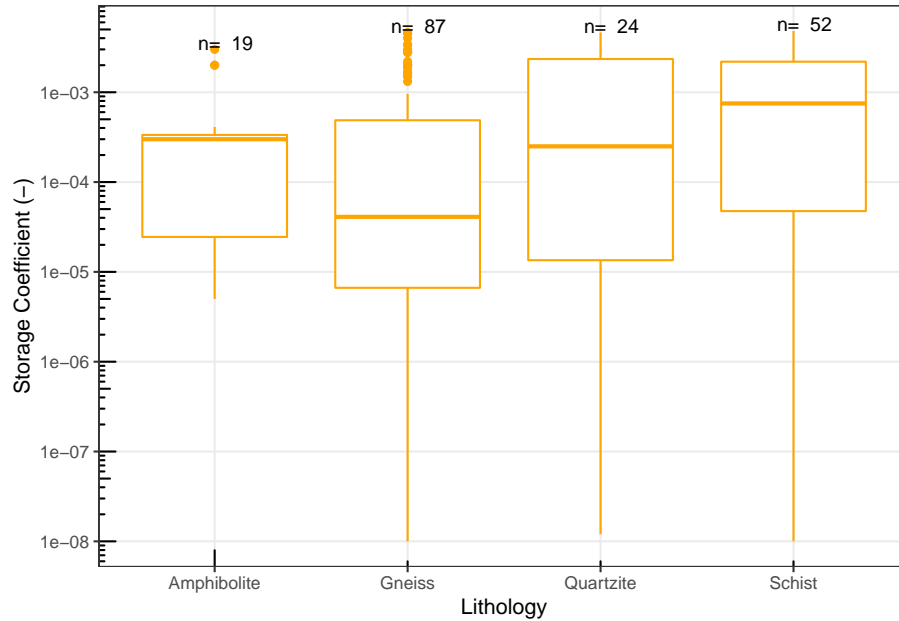


Figura 4.53: Diagramas de caja en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en diferentes rocas metamórficas.

Tabla 4.54: Valores mínimos, máximos y cuartiles del coeficiente de almacenamiento (-)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Amphibolite	19	5.00e-06	2.45e-05	3.00e-04	3.35e-04	3.00e-03
Gneiss	87	1.00e-08	6.65e-06	4.10e-05	4.88e-04	4.63e-03
Quartzite	24	1.20e-08	1.35e-05	2.50e-04	2.34e-03	4.63e-03
Schist	52	1.00e-08	4.75e-05	7.50e-04	2.19e-03	4.80e-03

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 del coeficiente de almacenamiento (-) en diferentes rocas metamórficas. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.55: Comparación de las medidas de tendencia central del coeficiente de almacenamiento (-)

Litología	Media	Media_geométrica	Mediana
Amphibolite	5.16e-04	1.26e-04	3.00e-04
Gneiss	6.52e-04	5.00e-05	4.10e-05
Quartzite	1.30e-03	1.67e-04	2.50e-04
Schist	1.39e-03	2.10e-04	7.50e-04

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) del coeficiente de almacenamiento (-) en diferentes rocas metamórficas.

Tabla 4.56: Medidas de dispersión del coeficiente de almacenamiento (-)

Litología	Rango	IQR	σ	DM	DMd	DMg
Amphibolite	3.00e-03	3.11e-04	8.43e-04	5.74e-04	4.45e-04	4.69e-04
Gneiss	4.63e-03	4.81e-04	1.15e-03	8.80e-04	6.43e-04	6.43e-04
Quartzite	4.63e-03	2.33e-03	1.56e-03	1.34e-03	1.26e-03	1.26e-03
Schist	4.80e-03	2.14e-03	1.52e-03	1.30e-03	1.28e-03	1.31e-03

Nota:

Tabla de comparación de las medidas de dispersión del coeficiente de almacenamiento (-) en diferentes rocas metamórficas. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

de tendencia central bastante parecidas. Por su parte el gneis, cuarcita y esquisto tienen medianas que son más parecidas sus respectivas medias geométricas que a las medias aritméticas.

La tabla 4.55 muestra las medidas de dispersión del coeficiente de almacenamiento (-) en rocas metamórficas. Se puede inferir de acuerdo con las mediciones de la desviación estándar que existe una mayor dispersión entre los datos de la cuarcita, por su parte la anfibolita es la roca que presenta una menor dispersión entre sus datos.

La desviación de los datos respecto a la mediana resulta ser menor que las desviaciones respecto a la media geométrica y la media aritmética en todos los casos presentados en la tabla.

En la figura 4.53 se presentan los histogramas del coeficiente de almacenamiento de cada una de las rocas metamórficas. En todos los casos, a excepción del gneis,

los histogramas son unimodales con sesgo negativo y las mayores frecuencias se presentan en valores cercanos a: 8×10^{-4} (-) para la anfibolita; 5×10^{-3} (-) en la cuarcita y el esquisto. Para el caso del gneis el histograma es bimodal con las mayores frecuencias en 8×10^{-6} y 5×10^{-3} (-) de coeficiente de almacenamiento.

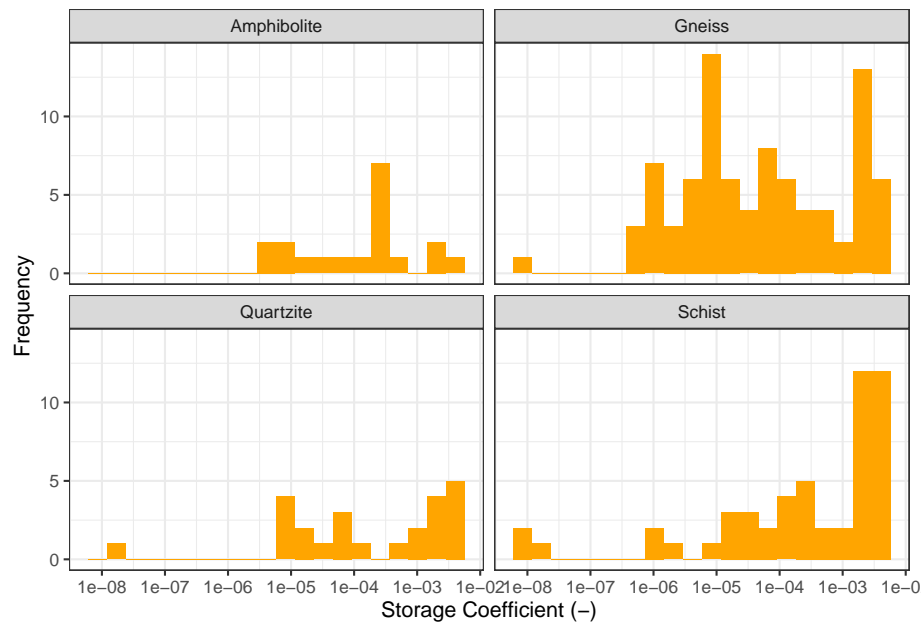


Figura 4.54: Histogramas en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en rocas metamórficas. Histogramas construidos a partir de 20 clases

4.3.4.3 Rocas sedimentarias

Los diagramas de caja del coeficiente de almacenamiento (-) en escala logarítmica para las rocas sedimentarias más representativas de la base de datos se presentan en la figura 4.54. Los diagramas muestran un comportamiento bastante similar en la caliza, dolomita y arenisca; con sus respectivas medianas dentro de un rango de 1×10^{-4} (-) a 5×10^{-4} (-). En el caso de la creta, el comportamiento difiere de las demás litologías, con una media que supera 1×10^{-3} (-) de coeficiente de almacenamiento.

Valores anómalos están presentes en las calizas, dolomitas y areniscas, siendo en los tres casos aquellas estimaciones mayores a 3×10^{-3} consideradas anómalas. Como ejemplo se tienen estimaciones mayores a 3×10^{-3} (-) en carbonatos de Inglaterra registradas por MacDonald and Allen (2001).

Los valores máximos, mínimos y cuartiles del coeficiente de almacenamiento (-) se muestran en la tabla 4.57. El valor máximo está presente en la creta y la

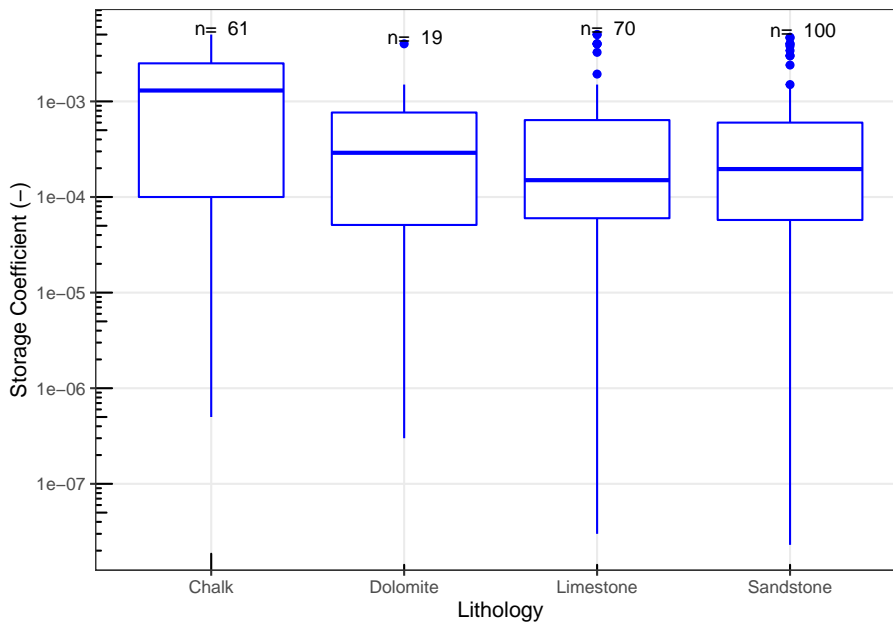


Figura 4.55: Diagramas de caja en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en diferentes rocas sedimentarias.

Tabla 4.57: Valores mínimos, máximos y cuartiles del coeficiente de almacenamiento (-)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Chalk	61	5.00e-07	1.00e-04	1.30e-03	2.50e-03	5.00e-03
Dolomite	19	3.00e-07	5.10e-05	2.90e-04	7.65e-04	4.00e-03
Limestone	70	3.00e-08	6.00e-05	1.50e-04	6.38e-04	5.00e-03
Sandstone	100	2.30e-08	5.75e-05	1.96e-04	6.00e-04	4.69e-03

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 del coeficiente de almacenamiento (-) en diferentes rocas sedimentarias. La columna **No. datos** corresponde con el número de observaciones.

caliza con 5×10^{-3} , por su parte la arenisca es la roca que registra el valor mínimo más bajo de todas las rocas sedimentarias registradas en la base de datos con 2.3×10^{-8} (-). La creta es la roca con la median más alta, registrando un valor de 1.3×10^{-3} ; por el contrario, la caliza presenta la mediana más baja con 1.5×10^{-4} (-).

La tabla 4.58 muestra la comparación de las diferentes medidas de tendencia central de los datos del coeficiente de almacenamiento (-) en rocas metamórficas.

Tabla 4.58: Comparación de las medidas de tendencia central del coeficiente de almacenamiento (-)

Litología	Media	Media geométrica	Mediana
Chalk	1.56e-03	5.10e-04	1.30e-03
Dolomite	6.02e-04	1.59e-04	2.90e-04
Limestone	6.78e-04	1.27e-04	1.50e-04
Sandstone	6.13e-04	1.24e-04	1.96e-04

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) del coeficiente de almacenamiento (-) en diferentes rocas sedimentarias.

Tabla 4.59: Medidas de dispersión del coeficiente de almacenamiento (-)

Litología	Rango	IQR	σ	DM	DMd	DMg
Chalk	5.00e-03	2.40e-03	1.49e-03	1.27e-03	1.26e-03	1.36e-03
Dolomite	4.00e-03	7.14e-04	9.42e-04	6.02e-04	5.34e-04	5.41e-04
Limestone	5.00e-03	5.77e-04	1.18e-03	7.47e-04	6.23e-04	6.24e-04
Sandstone	4.69e-03	5.42e-04	1.05e-03	6.88e-04	5.49e-04	5.55e-04

Nota:

Tabla de comparación de las medidas de dispersión del coeficiente de almacenamiento (-) en diferentes rocas sedimentarias. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

Analizando la tabla se puede inferir que la creta tiene una mediana más parecida a la media aritmética que a la media geométrica. En las otras rocas presentes en la tabla las tres de tendencia central resultan ser bastante parecidas entre sí.

La tabla 4.59 muestra las comparaciones de las medidas de dispersión de los datos del coeficiente de almacenamiento (-) en rocas sedimentarias. En respuesta a las lecturas de la desviación estándar se concluye que la dolomita es la roca que presenta una menor dispersión entre sus datos, por su parte la creta es la roca sedimentaria con una mayor dispersión en comparación con las otras rocas sedimentarias registradas.

En todas las rocas presentes en la tabla se puede inferir que las desviaciones de los datos respecto a la mediana son menores en comparación con las desviaciones respecto a la media aritmética y la media geométrica, siendo la media aritmética la que difiere más de las tres desviaciones medias absolutas de un punto central.

Las distribuciones del coeficiente de almacenamiento (-) (figura 4.56) de las diferentes rocas sedimentarias muestran ser unimodales con sesgo negativo en todos los casos, con las mayores frecuencias en 5×10^{-3} , 8×10^{-4} , 8×10^{-4} y 1×10^{-4} (-) de coeficiente de almacenamiento para la Creta, dolomita, caliza y arenisca respectivamente.

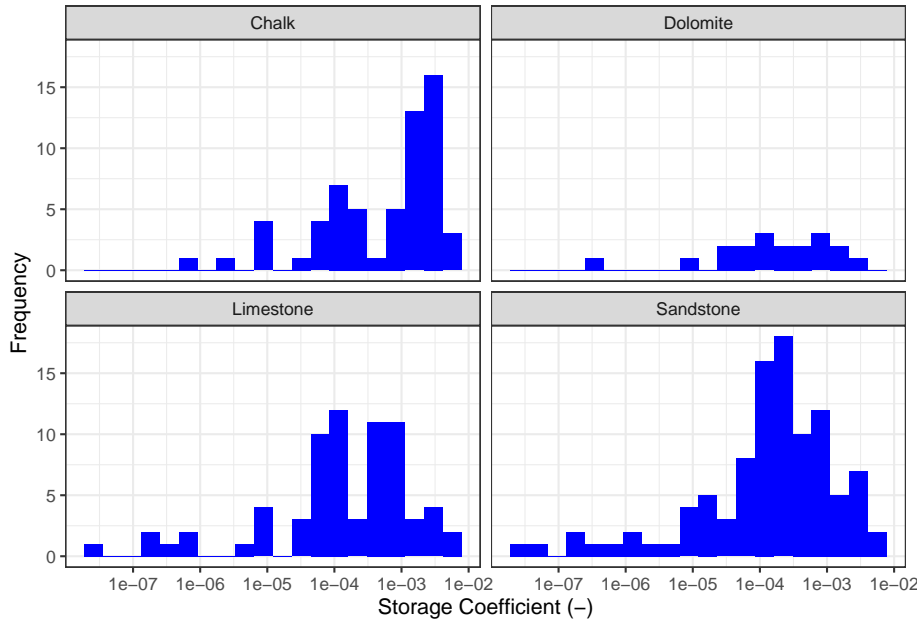


Figura 4.56: Histogramas en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en rocas sedimentarias. Histogramas construidos a partir de 20 clases

4.3.4.4 Sedimentos no consolidados

En la figura 4.57 se muestran los diagramas de caja en escala logarítmica del coeficiente de almacenamiento (-) para los tipos de sedimento no consolidados más representativos de la base de datos. En los diagramas de caja es posible apreciar la presencia de valores anómalos en todos los casos expuestos, estos datos anómalos aparecen en estimaciones mayores a 2×10^{-3} (-). Ejemplos de valores anómalos son las arenas de la Llanura costera del golfo, EE. UU. registrados por Huang et al. (2012).

Los valores puntuales de los diagramas de caja se presentan en la tabla 4.60. Las arenas son los sedimentos que presentan el registro máximo más alto en comparación con los otros sedimentos analizados, con un valor de 5×10^{-3} (-); el registro más bajo está representado por la combinación de arenas y gravas con 1.17×10^{-9} (-). El valor de la mediana más alto de los sedimentos presentados

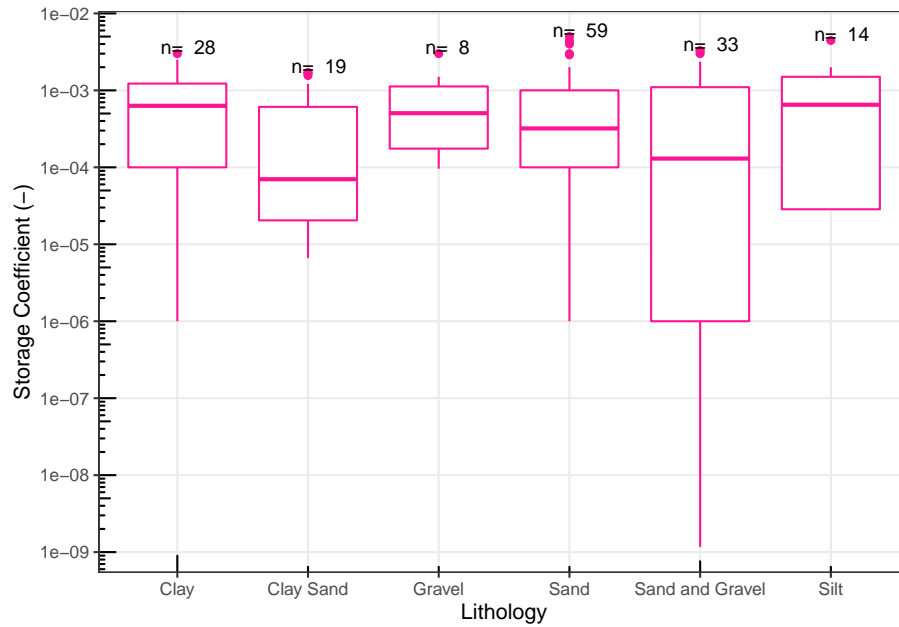


Figura 4.57: Diagramas de caja en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en diferentes sedimentos no consolidados.

corresponde con los limos con 6.5×10^{-4} (-), mientras que el registro más bajo de la mediana lo tienen las arenas arcillosas con 7×10^{-5} (-).

La tabla 4.61 muestra la comparación de las diferentes medidas de tendencia central de los datos del coeficiente de almacenamiento (-) en sedimentos no consolidados. Se puede inferir que las arenas arcillosas tienen una media aritmética más parecida a la media geométrica que a la mediana, en el caso de la combinación de arenas y gravas la media aritmética resulta ser más parecida a la mediana que a la media geométrica. El resto de las litologías analizadas en la tabla tienen una mediana más parecida a la media geométrica que a la media aritmética.

El resumen de las medidas de dispersión del coeficiente de almacenamiento (-) de los datos normalizados se presenta en la tabla 4.62 en donde se puede apreciar de acuerdo con el análisis de la desviación estándar que los limos tienen una mayor dispersión entre sus datos del coeficiente de almacenamiento en comparación con el resto de las litologías, por su parte las arenas arcillosas son los sedimentos con la menor dispersión entre sus datos.

En todos los casos presentes en la tabla de dispersión se tienen desviaciones medias absolutas que tienden a ser bastante parecidas, variando en el mismo orden de magnitud, sin embargo, se puede concluir que la desviación de los

Tabla 4.60: Valores mínimos, máximos y cuartiles del coeficiente de almacenamiento (-)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Clay	28	1.00e-06	1.00e-04	6.30e-04	1.22e-03	3.00e-03
Clay Sand	19	6.60e-06	2.05e-05	7.00e-05	6.10e-04	1.70e-03
Gravel	8	9.60e-05	1.75e-04	5.05e-04	1.13e-03	3.00e-03
Sand	59	1.00e-06	1.00e-04	3.20e-04	1.00e-03	5.00e-03
Sand and Gravel	33	1.17e-09	1.00e-06	1.30e-04	1.10e-03	3.27e-03
Silt	14	2.80e-05	2.85e-05	6.50e-04	1.50e-03	4.45e-03

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 del coeficiente de almacenamiento (-) en diferentes sedimentos no consolidados. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.61: Comparación de las medidas de tendencia central del coeficiente de almacenamiento (-)

Litología	Media	Media geométrica	Mediana
Clay	8.36e-04	2.92e-04	6.30e-04
Clay Sand	4.57e-04	1.13e-04	7.00e-05
Gravel	8.63e-04	4.53e-04	5.05e-04
Sand	7.52e-04	2.26e-04	3.20e-04
Sand and Gravel	7.75e-04	3.54e-05	1.30e-04
Silt	1.01e-03	2.47e-04	6.50e-04

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) del coeficiente de almacenamiento (-) en diferentes sedimentos no consolidados.

Tabla 4.62: Medidas de dispersión del coeficiente de almacenamiento (-)

Litología	Rango	IQR	σ	DM	DMd	DMg
Clay	3.00e-03	1.12e-03	8.54e-04	6.79e-04	6.62e-04	7.11e-04
Clay Sand	1.69e-03	5.90e-04	6.17e-04	4.93e-04	4.30e-04	4.33e-04
Gravel	2.90e-03	9.50e-04	9.97e-04	7.28e-04	6.94e-04	6.94e-04
Sand	5.00e-03	9.00e-04	1.10e-03	7.44e-04	6.51e-04	6.57e-04
Sand and Gravel	3.27e-03	1.10e-03	1.06e-03	8.61e-04	7.43e-04	7.64e-04
Silt	4.42e-03	1.47e-03	1.27e-03	9.70e-04	9.70e-04	9.70e-04

Nota:

Tabla de comparación de las medidas de dispersión del coeficiente de almacenamiento (-) en diferentes sedimentos no consolidados. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

datos respecto a la mediana resulta ser la menor en comparación con las otras dos desviaciones, para cada tipo de sedimento.

Los histogramas de la figura 4.58 muestran un comportamiento unimodal con sesgo negativo para el caso de las arenas, arcillas y combinación de arenas y gravas en donde las mayores frecuencias se presentan en valores cercanos a 8×10^{-4} , 1×10^{-5} , y 1×10^{-5} (-) respectivamente.

Las arenas arcillosas tienen una distribución bimodal en donde las mayores frecuencias se encuentran en 1×10^{-5} y 1×10^{-3} (-). La distribución de los limos es poco representativa debido a la falta de datos, sin embargo, puede observarse una mayor frecuencia en datos cercanos a 1×10^{-3} (-) de coeficiente de almacenamiento.

4.3.5 Rendimiento específico

Los rangos del rendimiento específico (-) varían en magnitudes de 3 órdenes de magnitud. Los diagramas de rangos se presentan en la tabla 4.59, estos rangos toman en cuenta todas las rocas expuestas en la base de datos, por lo que incluye valores anómalos dentro de sus resultados.

En la figura 4.60 se pueden observar los rangos del rendimiento específico (-) para algunas de las litologías más representativas. Para el caso de las rocas ígneas, los rangos no varían mucho en relación al tipo de roca. En rocas ígneas el basalto y el granito superan los 0.3 (-) a diferencia. Además, es el granito la roca que alcanza los valores más bajos, con 0.005(-).

Las rocas metamórficas tampoco difieren mucho unas de otras, siendo el gneis el que registra el menor rendimiento específico, con estimaciones cercanas a 0.003(-)

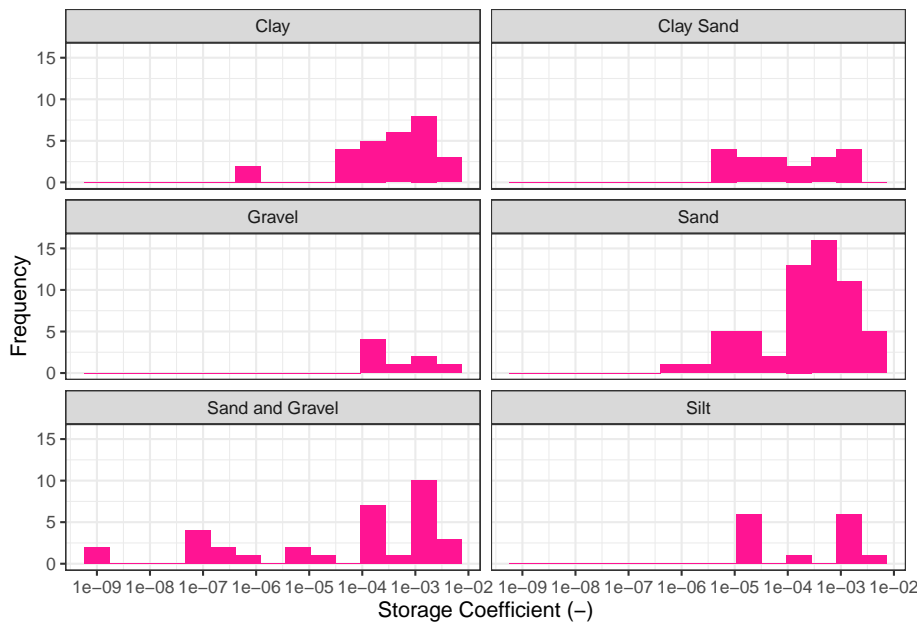


Figura 4.58: Histogramas en escala logarítmica en base 10 del coeficiente de almacenamiento (-) en sedimentos no consolidados. Histogramas construidos a partir de 15 clases

). Mientras que el esquisto y la cuarcita alcanzan estimaciones cercanas a 0.3 (-).

Por el lado de las rocas sedimentarias, la creta es la roca que tiene un menor valor de rendimiento específico y un menor rango en general, con estimaciones que van de 0.003 (-) hasta valores de 0.1 (-). Mientras que la dolomita y la caliza registran los mayores rendimientos específicos con valores cercanos a 0.5 (-).

Finalmente, los sedimentos no consolidados tienen rangos muy variados dependiendo del tipo de sedimentos. Para el caso de las arenas los registros superan el valor de 0.5 (-); también los limos alcanzan valores muy cercanos a 0.5 (-). A su vez las arenas y arcillas muestran las estimaciones más bajas con valores de 0.005 (-).

Los diagramas de caja para del rendimiento específico en escala logarítmica del rendimiento específico (-) para las diferentes clasificaciones litológicas, se presentan en la figura 4.61. En la gráfica de la figura puede apreciarse una cantidad considerable de estimaciones anómalas en todos los casos expuestos. En rocas ígneas y sedimentarias los valores anómalos se presentan en estimaciones mayores a 0.2 (-), en rocas metamórficas las estimaciones mayores a 0.08 (-) se consideran anómalos, para los sedimentos no consolidados las estimaciones

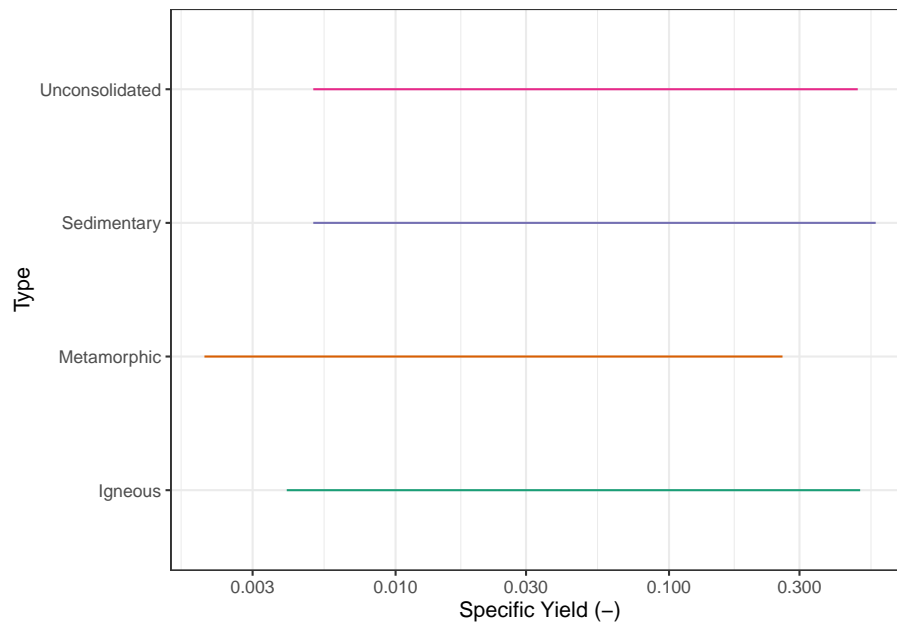


Figura 4.59: Diagrama de rangos en escala logarítmica en base 10 del rendimiento específico (-) para los diferentes tipos de litología.

Tabla 4.63: Valores mínimos, máximos y cuartiles del rendimiento específico (-)

Type	No.Datos	min.	Q1	Mediana	Q3	max.
Igneous	254	0.004	0.010	0.024	0.069	0.50
Metamorphic	278	0.002	0.010	0.017	0.037	0.26
Sedimentary	321	0.005	0.015	0.045	0.100	0.57
Unconsolidated	319	0.005	0.050	0.100	0.200	0.49

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 del rendimiento específico (-) para las diferentes clasificaciones litológicas. La columna **No. datos** corresponde con el número de observaciones.

mayores a 0.4 (-) se grafican como anómalas.

La tabla 4.63 muestra los valores mínimos, máximos, cuartiles, así como las medianas de cada grupo de datos. El valor mínimo está presente en las rocas metamórficas con 0.002 (-), mientras que los valores máximos se registran en sedimentos no consolidados con 0.61 (-) de rendimiento específico. El registro de la mediana más alto es 0.1 y es propio de los sedimentos no consolidados, por su parte el registro más bajo con 0.017 (-) está presente en rocas metamórficas.

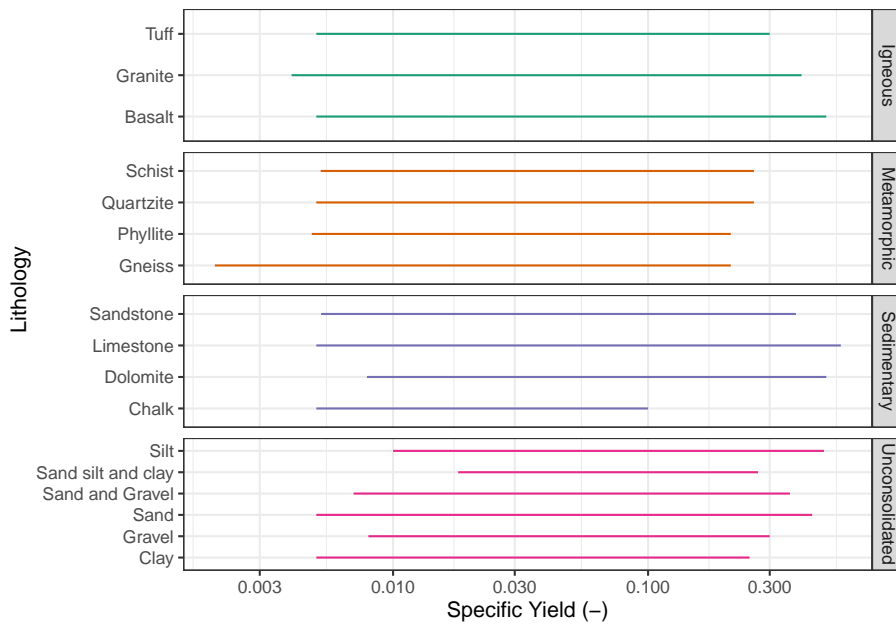


Figura 4.60: Diagrama de rangos en escala logarítmica en base 10 del rendimiento específico (-) para diferentes litologías.

Las mediciones de tendencia central de los datos del rendimiento específico de cada grupo de datos se presentan en la tabla 4.64. Del análisis de comparaciones se puede concluir que para todos los casos expuestos se tiene un mayor parecido entre la mediana y la media geométrica, mientras que la media aritmética tiende a alejarse más de los resultados presentados por las otras dos medidas de tendencia central.

Las medidas de dispersión de los datos del rendimiento específico (-) se muestran en tabla 4.65. La desviación estándar de los sedimentos no consolidados muestra ser más elevada que el resto de los tipos de litología, lo que refleja una mayor dispersión entre sus datos de rendimiento específico, a diferencia de las rocas metamórficas que muestra la menor dispersión entre datos en comparación con los otros tipos. En general se puede concluir que la desviación de los datos respecto a la mediana y respecto a la media geométrica son bajas en comparación la desviación respecto a la media aritmética.

La figura 4.62 muestra los histogramas del rendimiento específico para las diferentes clasificaciones litológicas, en la cual podemos observar comportamientos muy diversos. En todos los casos los histogramas tienden a ser multimodales. Para el caso de los sedimentos no consolidados se observa un claro sesgo negativo, con las mayores frecuencias en valores cercanos a 0.1 y 0.25(-).

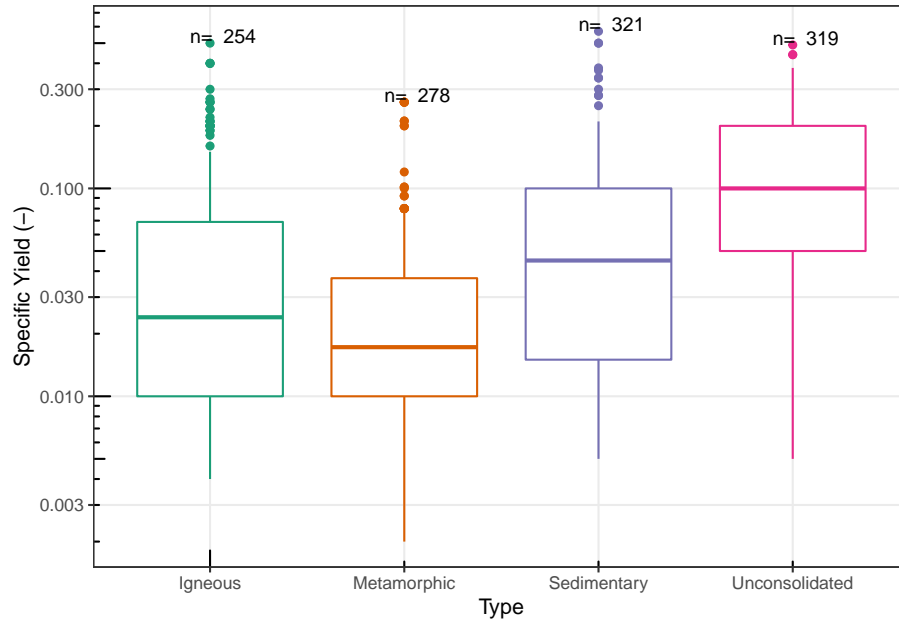


Figura 4.61: Diagramas de caja en escala logarítmica en base 10 del rendimiento específico para las diferentes clasificaciones litológicas.

Tabla 4.64: Comparación de las medidas de tendencia central del rendimiento específico (-)

Tipo	Media	Media_geométrica	Mediana
Igneous	0.057	0.029	0.024
Metamorphic	0.032	0.019	0.017
Sedimentary	0.070	0.039	0.045
Unconsolidated	0.123	0.080	0.100

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) del rendimiento específico (-) en los diferentes tipos de litologías.

Tabla 4.65: Medidas de dispersión del rendimiento específico (-)

Tipo	Rango	IQR	σ	DM	DMd	DMg
Igneous	0.496	0.059	0.077	0.052	0.044	0.044
Metamorphic	0.258	0.027	0.044	0.026	0.022	0.022
Sedimentary	0.565	0.085	0.082	0.057	0.053	0.053
Unconsolidated	0.485	0.150	0.093	0.078	0.077	0.079

Nota:

Tabla de comparación de las medidas de dispersión del rendimiento específico (-) en los diferentes tipos de litologías. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

Las rocas sedimentarias tienen un histograma con un ligero sesgo positivo, en el cual los valores con más frecuencias se encuentran en 0.1 y en 0.0125(-). Las rocas ígneas al igual que las rocas sedimentarias tienden a tener un sesgo positivo, con las mayores frecuencias en las estimaciones cercanas a 0.012 y 0.04(-).

Finalmente, las rocas metamórficas fracturadas tienden a tener una cierta simetría, ligeramente sesgada hacia estimaciones de menor valor. En este tipo de rocas se tiene la mayor frecuencia en valores cercanos a 0.015 (-), aunque un pico importante en 0.05 (-), también acumula una cantidad considerable de datos.

Correlación rendimiento específico

Las figuras de correlación (4.63) muestran la dispersión y los modelos de regresión entre los datos del rendimiento específico en función de cada una de las propiedades hidrogeológicas (porosidad total, porosidad efectiva, conductividad hidráulica, coeficiente de almacenamiento y compresibilidad). Estos modelos a pesar de aplicarse a los datos transformados logarítmicamente resultan ser bastante pobres para todos los casos, lo que indica una pobre correlación entre los datos del rendimiento específico y cada una de las propiedades hidrogeológicas.

4.3.5.1 Rocas ígneas

En la figura 4.64 se muestran los diagramas de caja en escala logarítmica del rendimiento específico (-) de las rocas ígneas más representativas de la base de datos. La toba no presenta ningún valor anómalo de acuerdo con el análisis, mientras que en el caso del basalto aquellas estimaciones mayores a 0.1 son consideradas anómalas, por su parte el granito tiene valores anómalos de

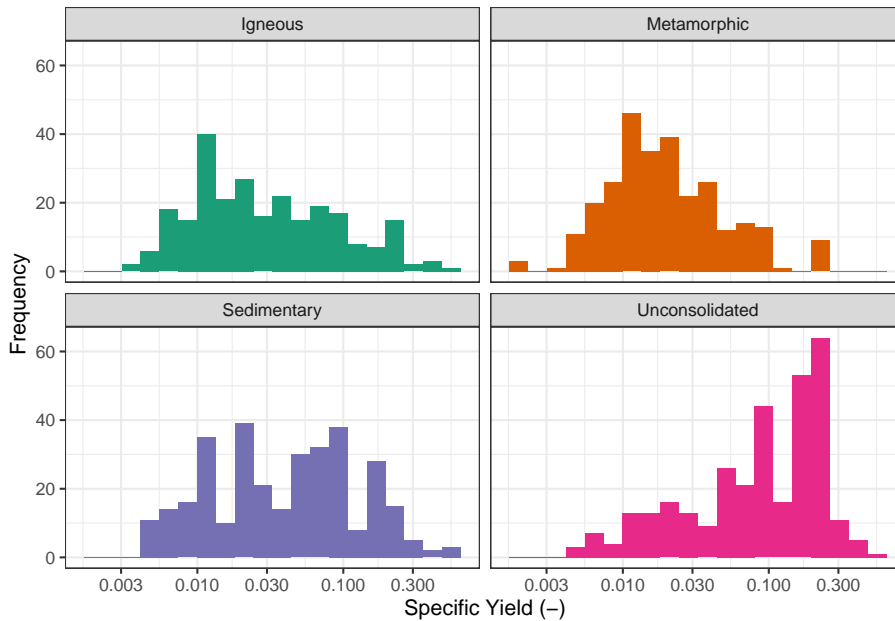


Figura 4.62: Histogramas en escala logarítmica en base 10 del rendimiento específico (-) para las diferentes clasificaciones litológicas. Histogramas contruidos a partir de 20 clases

rendimiento específico en aquellas medidas mayores a 0.15 (-).

Ejemplos de valores anómalos del rendimiento específico, registrados dentro de la base de datos, se presentan en rocas volcánicas fracturadas de la isla Cheju, Corea con estimaciones de hasta 0.5 (-) (Hahn et al., 1997).

La tabla 4.66 presenta las estimaciones mínimas, máximas y cuartiles del rendimiento específico (-) en rocas ígneas. La roca que registra el máximo valor del rendimiento específico es el basalto con 0.5 (-), mientras que el valor más bajo está presente en el granito con 0.004 (-). Analizando las medianas se puede concluir que el mayor valor es 0.11 (-) en las tobas, mientras que el menor valor lo registra el granito con 0.023 (-).

La tabla 4.67 muestra la comparación de las medidas de tendencia central de los datos del rendimiento específico (-) en rocas ígneas. Se puede concluir que en las tres litologías expuestas se tienen medianas que son más parecidas a las medias geométricas, mientras que las medias aritméticas tienden diferir en mayor cantidad.

Las medidas de dispersión del rendimiento específico (-) se muestran en la tabla 4.68. Analizando las desviaciones estándar de los datos se puede concluir que las tobas presentan una mayor dispersión entre sus datos del rendimiento específico,

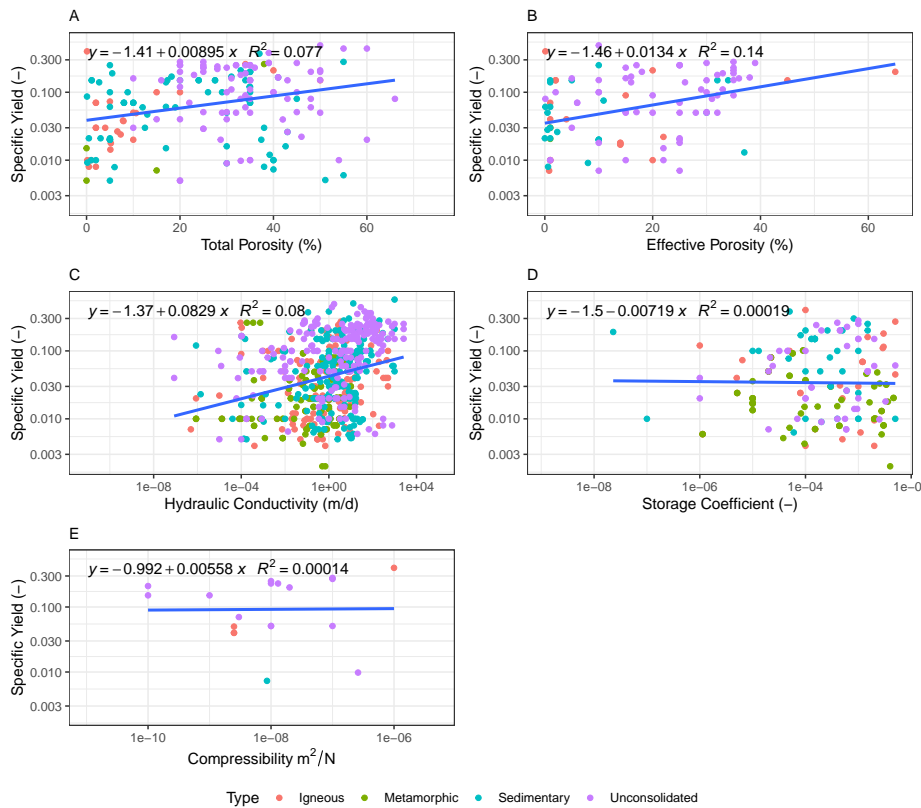


Figura 4.63: Diagramas de correlación del rendimiento específico en función de las diferentes propiedades hidrogeológicas. A: Rendimiento Específico vs Porosidad Total (transformación logarítmica en base 10 del eje y); B: Rendimiento específico vs Porosidad Efectiva (transformación logarítmica en base 10 del eje y); C: Rendimiento Específico vs Conductividad Hidráulica (transformación logarítmica en base 10 del eje y, y el eje x); D: Rendimiento Específico vs Coeficiente de Almacenamiento (transformación logarítmica en base 10 del eje y, y el eje x); E: Rendimiento específico vs Compresibilidad (transformación logarítmica en base 10 del eje y, y eje x)

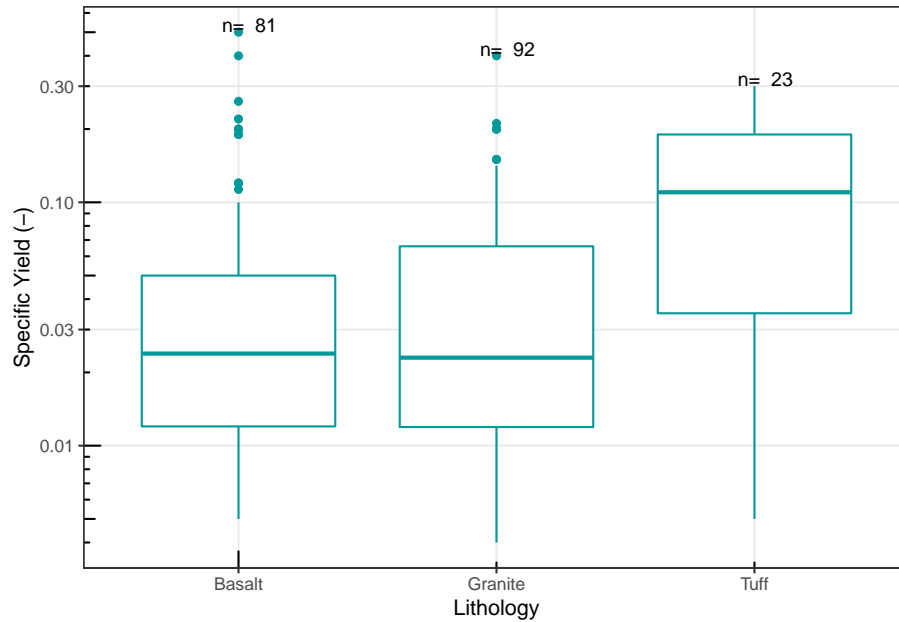


Figura 4.64: Diagramas de caja en escala logarítmica en base 10 del rendimiento específico (-) de diferentes rocas ígneas.

Tabla 4.66: Valores mínimos, máximos y cuartiles del rendimiento específico (-)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Basalt	81	0.005	0.012	0.024	0.050	0.5
Granite	92	0.004	0.012	0.023	0.066	0.4
Tuff	23	0.005	0.035	0.110	0.190	0.3

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 del rendimiento específico (-) en diferentes rocas ígneas. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.67: Comparación de las medidas de tendencia central del rendimiento específico (-)

Litología	Media	Media_geométrica	Mediana
Basalt	0.057	0.029	0.024
Granite	0.048	0.027	0.023
Tuff	0.116	0.068	0.110

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) del rendimiento específico (-) en diferentes rocas ígneas.

Tabla 4.68: Medidas de dispersión del rendimiento específico (-)

Litología	Rango	IQR	σ	DM	DMd	DMg
Basalt	0.495	0.038	0.084	0.053	0.043	0.044
Granite	0.396	0.054	0.061	0.041	0.036	0.036
Tuff	0.295	0.155	0.092	0.078	0.077	0.082

Nota:

Tabla de comparación de las medidas de dispersión del rendimiento específico (-) en diferentes rocas ígneas. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

por su parte los granitos presentan una menor dispersión entre sus datos, en comparación con las otras rocas analizadas.

Las desviaciones medias absolutas de un punto central muestran que la desviación de los datos respecto a la mediana es menor que la desviación respecto a la media geométrica y la media aritmética en todas las rocas ígneas expuestas en la tabla de las medidas de dispersión.

Los histogramas de la figura 4.65 muestran el comportamiento de las distribuciones de datos del rendimiento específico (-). En los casos del basalto y el granito se tiene una distribución con sesgo positivo; siendo el basalto bimodal, con las mayores frecuencias en 0.01 y 0.05(-); y el granito unimodal con la mayor frecuencia en 0.01(-), aunque existe una importante concentración de datos en valores cercanos a 0.05(-).

El caso de la toba es diferente ya que su distribución tiende a tener un sesgo negativo y ser unimodal, con las mayores frecuencias en estimaciones entre los

valores de 0.1 y 0.3 (-) de rendimiento específico.

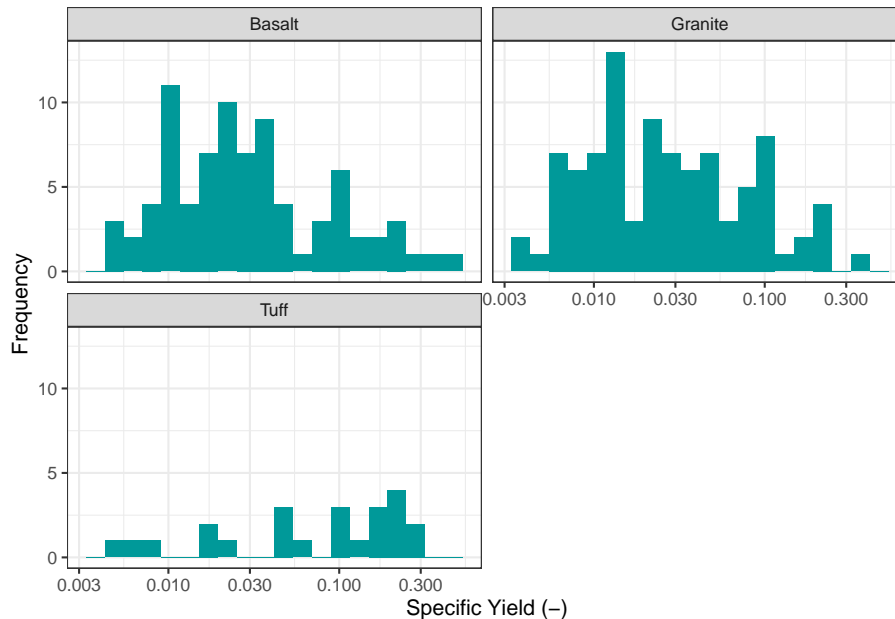


Figura 4.65: Histogramas en escala logarítmica en base 10 del rendimiento específico (-) en diferentes rocas ígneas. Histogramas construidos a partir de 20 clases

4.3.5.2 Rocas metamórficas

Los diagramas de caja en escala logarítmica del rendimiento específico (-) en las rocas metamórficas más representativas de la base de datos se presentan en la figura 4.66. En este diagrama es posible observar la importante cantidad de datos anómalos en todos los casos presentados. En rocas como el gneis, cuarcita y esquisto se tienen estimaciones anómalas en los valores mayores a 0.06 (-), en la filita los valores anómalos aparecen en estimaciones mayores a 0.04 (-) de rendimiento específico.

Ejemplos de valores anómalos del rendimiento específico registrados en la base de datos se presentan en los trabajos de Golian et al. (2018) donde se estiman valores de más de 0.2 (-) de rendimiento específico en complejos metamórficos fracturados.

Los valores puntuales de los diagramas de caja se presentan en la tabla 4.69. El valor máximo más alto de las cuatro litologías expuestas es 0.26 y está presente de las cuarcitas y los esquistos, el valor mínimo lo registra el gneis con 0.002 (-). El análisis de las medianas muestra que el valor más alto registrado es propio

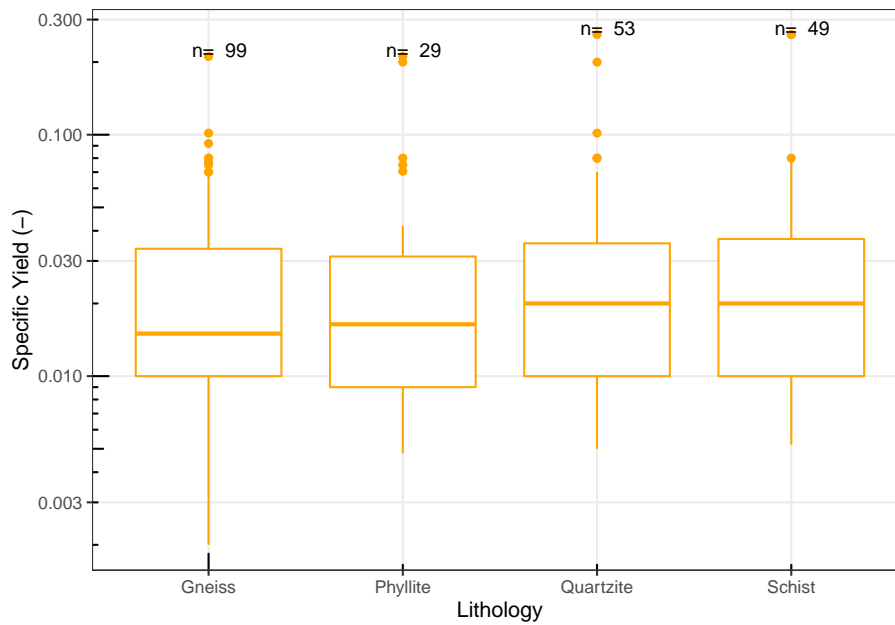


Figura 4.66: Diagramas de caja en escala logarítmica en base 10 del rendimiento específico (-) de diferentes rocas metamórficas.

Tabla 4.69: Valores mínimos, máximos y cuartiles del rendimiento específico (-)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Gneiss	99	0.002	0.010	0.015	0.034	0.211
Phyllite	29	0.005	0.009	0.016	0.031	0.211
Quartzite	53	0.005	0.010	0.020	0.035	0.260
Schist	49	0.005	0.010	0.020	0.037	0.260

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 del rendimiento específico (-) en diferentes rocas metamórficas. La columna **No. datos** corresponde con el número de observaciones.

de las cuarcitas y los esquistos, mientras que el valor más bajo se registra en el gneis.

Las comparaciones de las medidas de tendencia central (tabla 4.70) muestran que en las rocas metamórficas expuestas se tienen valores de la mediana parecidos a los valores de la media geométrica, por su parte la media aritmética es la que más difiere en cuanto a su magnitud.

Tabla 4.70: Comparación de las medidas de tendencia central del rendimiento específico (-)

Litología	Media	Media_geométrica	Mediana
Gneiss	0.026	0.017	0.015
Phyllite	0.035	0.020	0.016
Quartzite	0.033	0.020	0.020
Schist	0.033	0.019	0.020

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) del rendimiento específico (-) en diferentes rocas metamórficas.

Tabla 4.71: Medidas de dispersión del rendimiento específico (-)

Litología	Rango	IQR	σ	DM	DMd	DMg
Gneiss	0.209	0.024	0.029	0.020	0.017	0.017
Phyllite	0.206	0.022	0.051	0.032	0.026	0.026
Quartzite	0.255	0.025	0.045	0.025	0.022	0.022
Schist	0.255	0.027	0.051	0.027	0.023	0.023

Nota:

Tabla de comparación de las medidas de dispersión del rendimiento específico (-) en diferentes rocas metamórficas. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

Analizando las desviaciones estándar de los datos del rendimiento específico (tabla 4.71) se puede inferir que de las cuatro rocas expuestas el gneis presenta una menor dispersión, por el contrario, el esquisto tiene una mayor dispersión entre sus datos del rendimiento específico.

En los cuatro casos de rocas metamórficas expuestas en la tabla de las medidas de dispersión se tiene un comportamiento parecido, en el cual las desviaciones de los datos respecto a la mediana son bastante parecidas a las desviaciones respecto a la media geométrica, siendo estas dos desviaciones menores que la desviación respecto a la media aritmética.

Analizando los histogramas de cada tipo de roca metamórfica expuesta (figura 4.67) se puede concluir que todas las distribuciones son unimodales con sesgo positivo, el gneis, filita y la cuarcita tienen las mayores frecuencias en 0.01 (-),

mientras que el esquistoso tiene sus mayores frecuencias en 0.02(-) de rendimiento específico.

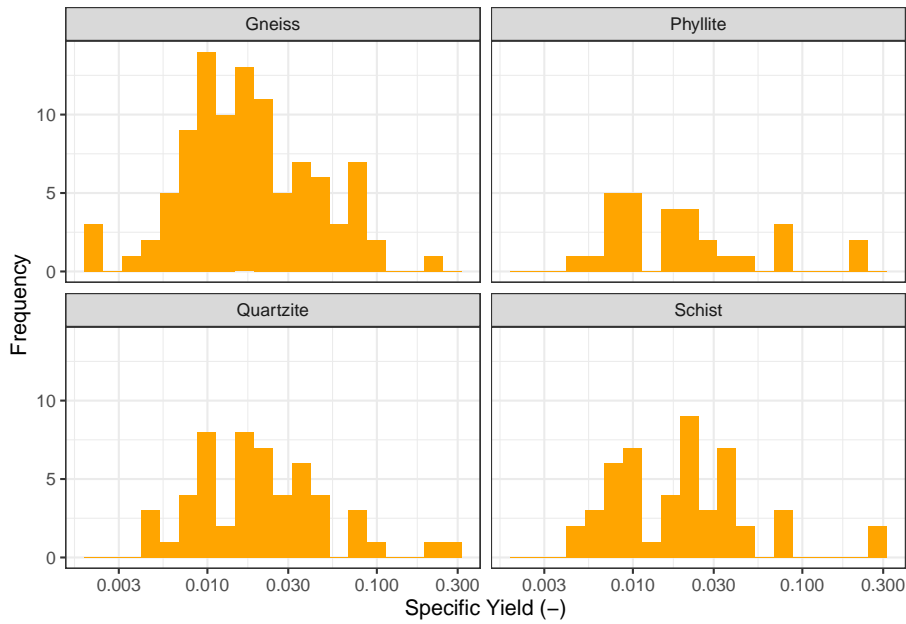


Figura 4.67: Histogramas en escala logarítmica en base 10 del rendimiento específico (-) en diferentes rocas metamórficas. Histogramas construidos a partir de 20 clases

4.3.5.3 Rocas sedimentarias

El rendimiento específico tiende a variar en gran medida de acuerdo con el tipo de roca sedimentaria como se observa en la figura 4.68. En rocas como la creta y la arenisca no se presenta ningún valor anómalo entre su conjunto de datos. La caliza y la dolomita presentan estimaciones anómalas en aquellos valores superiores a 0.15 y 0.1 (-) respectivamente.

Algunos ejemplos de valores anómalos del rendimiento específico se presentan en los trabajos de Bolster et al. (2001) con calizas fracturadas que superan los valores de 0.57 (-).

Las medidas puntuales de los valores mínimos, máximos, cuartiles y medianas del rendimiento específico, registradas para cada roca sedimentaria, se especifican en la tabla 4.72. El registro máximo de todas las rocas sedimentarias lo presenta la caliza con 0.60 (-), por su parte el registro mínimo de las cuatro rocas expuestas es 0.005 (-) y está presente en areniscas, calizas y cretas.

El valor de la mediana más alto registrado en las cuatro rocas está presente en

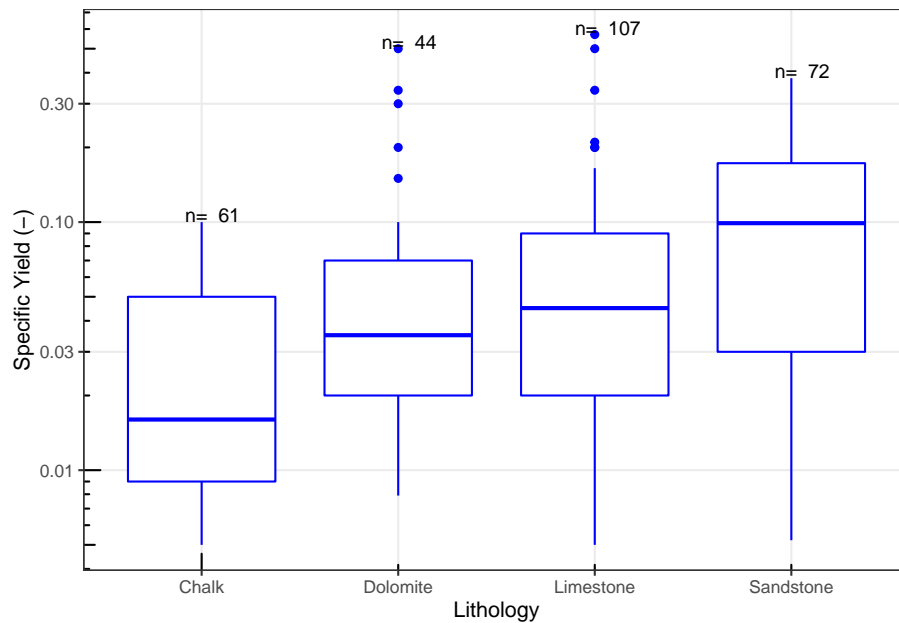


Figura 4.68: Diagramas de caja en escala logarítmica en base 10 del rendimiento específico (-) de diferentes rocas sedimentarias.

Tabla 4.72: Valores mínimos, máximos y cuartiles del rendimiento específico (-)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Chalk	61	0.005	0.009	0.016	0.050	0.10
Dolomite	44	0.008	0.020	0.035	0.070	0.50
Limestone	107	0.005	0.020	0.045	0.090	0.57
Sandstone	72	0.005	0.030	0.099	0.173	0.38

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 del rendimiento específico (-) en diferentes rocas sedimentarias. La columna **No. datos** corresponde con el número de observaciones.

las areniscas con 0.099 (-), en las cretas se registra la estimación de la mediana más baja en comparación con las otras rocas sedimentarias analizadas.

Las medidas de tendencia central de los datos del rendimiento específico (-) se comparan en la tabla 4.73. En los carbonatos se tiene un comportamiento parecido en el cual la mediana tiende a ser más parecida a la media geométrica que a la media aritmética; en el caso de la arenisca se presenta una mediana más parecida a la media aritmética que a la media geométrica.

Tabla 4.73: Comparación de las medidas de tendencia central del rendimiento específico (-)

Litología	Media	Media_geométrica	Mediana
Chalk	0.030	0.019	0.016
Dolomite	0.068	0.039	0.035
Limestone	0.070	0.040	0.045
Sandstone	0.106	0.065	0.099

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) del rendimiento específico (-) en diferentes rocas sedimentarias.

Tabla 4.74: Medidas de dispersión del rendimiento específico (-)

Litología	Rango	IQR	σ	DM	DMd	DMg
Chalk	0.095	0.041	0.028	0.024	0.021	0.022
Dolomite	0.492	0.050	0.096	0.057	0.048	0.049
Limestone	0.565	0.070	0.087	0.056	0.051	0.051
Sandstone	0.375	0.143	0.088	0.071	0.069	0.072

Nota:

Tabla de comparación de las medidas de dispersión del rendimiento específico (-) en diferentes rocas sedimentarias. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (Dmg).

La tabla 4.74 compara las diferentes medidas de dispersión de los datos del rendimiento específico para las rocas más representativas de la base de datos. Analizando las desviaciones estándar de las diferentes rocas sedimentarias se infiere que las calizas tienen una mayor dispersión entre sus datos, por su parte la creta es la roca sedimentaria con la menor dispersión entre los datos del rendimiento específico.

El análisis de las desviaciones medias absolutas de un punto central presentes en la tabla de dispersiones, muestra que las diferentes rocas sedimentarias tienen una menor desviación de sus datos respecto a su mediana en comparación con sus desviaciones respecto a la media geométrica y media aritmética.

Las distribuciones para cada roca sedimentaria (figura 4.69) tienden a ser bastante diferentes, variando para cada tipo de roca. En el caso de la arenisca

se tiene una distribución con sesgo negativo con altas concentraciones en varios puntos, siendo el más importante la concentración en estimaciones cercanas a 0.1 (-). Para el caso de la caliza se tiene una distribución bimodal con las mayores concentraciones en valores cercanos a 0.01 y 0.12 (-).

La creta y la dolomita tienen distribuciones bimodales, en el caso de la creta las mayores concentraciones se presentan en valores cercanos a 0.01 y 0.08 (-); mientras que la dolomita tiene altas frecuencias en estimaciones cercanas a 0.05 y 0.08 (-) de rendimiento específico.

La presencia de multimodalidad sobre todo en carbonatos se debe a la presencia de rocas con diferentes propiedades texturales y estructurales que pertenecen al mismo tipo de rocas.

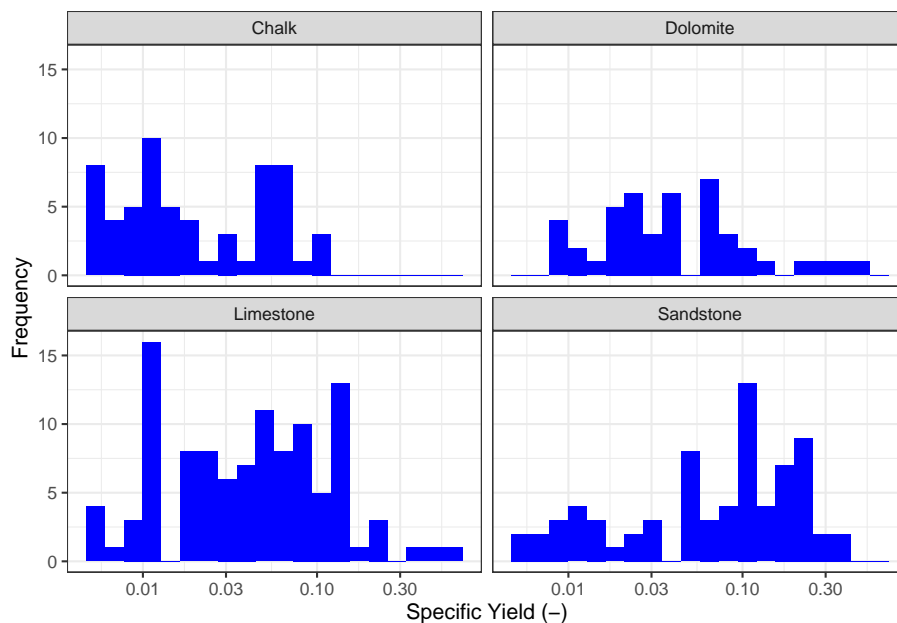


Figura 4.69: Histogramas en escala logarítmica en base 10 del rendimiento específico (-) en diferentes rocas sedimentarias. Histogramas construidos a partir de 20 clases

4.3.5.4 Sedimentos no consolidados

El rendimiento específico (-) varía en gran medida de acuerdo con el tipo de sedimento, como se observa en los diagramas de caja en escala logarítmica (figura 4.70). La presencia de valores anómalos solo se registra en los siguientes sedimentos: arenas, limos y arcillas. En arcillas y limos los valores mayores a 0.2 (-) se consideran anómalos, para las arenas las estimaciones mayores a 0.4 se consideran anómalas.

Un ejemplo de valores anómalos son los presentados por Cheong et al. (2008) donde se registra mediante modelos numéricos los valores de 0.44 (-) de rendimiento específico en arenas del sistema aluvial de Corea del Sur.

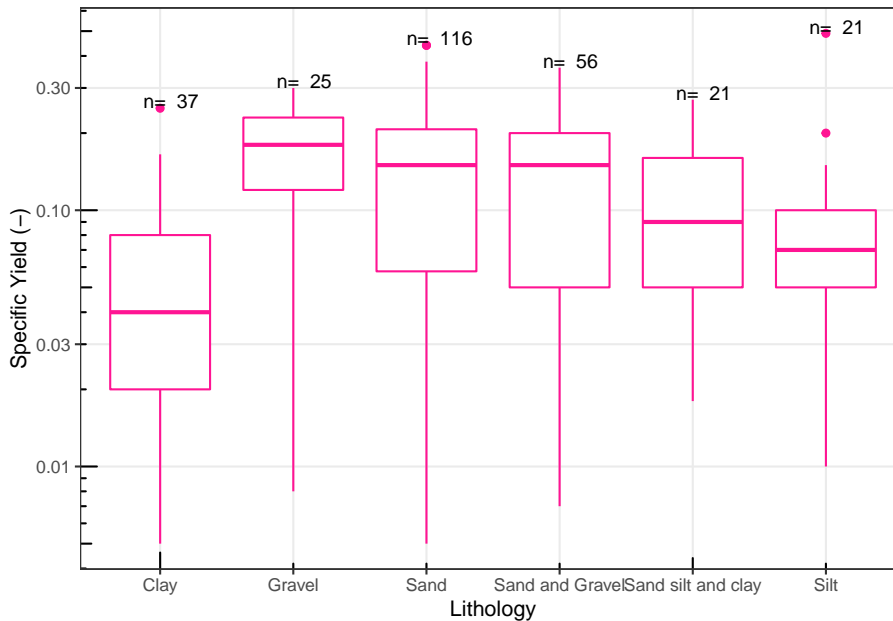


Figura 4.70: Diagramas de caja en escala logarítmica en base 10 del rendimiento específico (-) de diferentes sedimentos no consolidados.

La tabla 4.75 presenta los valores mínimos, máximos y cuartiles de los datos del rendimiento específico de los sedimentos no consolidados más representativos de la base de datos. El valor máximo del rendimiento específico en sedimentos no consolidados está presente en las arenas con 0.61 (-); en el caso de las arcillas y arenas se registra el valor más bajo de los datos del rendimiento específico con 0.005 (-). El registro de la mediana más alto se presenta en las gravas con 0.18 (-), por el contrario, las arcillas registran la mediana más baja con 0.02 (-), en comparación con los otros sedimentos.

En la tabla 4.76 se muestran las comparaciones de las medidas de tendencia central de los datos del rendimiento específico en sedimentos no consolidados. En las gravas, arenas y la combinación de arenas y gravas, se tiene una mediana que tiende a ser más parecida a la media aritmética que a la media geométrica. Para el resto de las litologías, la mediana tiende a ser más parecida a la media aritmética que a la media geométrica.

Las medidas de dispersión de los datos del rendimiento específico se presentan en la tabla 4.77. Analizando las desviaciones estándar de cada tipo de sedimento se puede concluir que las arenas tienen una mayor dispersión entre sus datos

Tabla 4.75: Valores mínimos, máximos y cuartiles del rendimiento específico (-)

Litología	No.Datos	min.	Q1	Mediana	Q3	max.
Clay	37	0.005	0.020	0.04	0.080	0.25
Gravel	25	0.008	0.120	0.18	0.230	0.30
Sand	116	0.005	0.058	0.15	0.207	0.44
Sand and Gravel	56	0.007	0.050	0.15	0.200	0.36
Sand silt and clay	21	0.018	0.050	0.09	0.160	0.27
Silt	21	0.010	0.050	0.07	0.100	0.49

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 del rendimiento específico (-) en diferentes sedimentos no consolidados. La columna **No. datos** corresponde con el número de observaciones.

Tabla 4.76: Comparación de las medidas de tendencia central del rendimiento específico (-)

Litología	Media	Media_geométrica	Mediana
Clay	0.058	0.036	0.04
Gravel	0.164	0.120	0.18
Sand	0.144	0.095	0.15
Sand and Gravel	0.135	0.088	0.15
Sand silt and clay	0.109	0.088	0.09
Silt	0.094	0.067	0.07

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) del rendimiento específico (-) en diferentes sedimentos no consolidados.

de rendimiento específico comparado con el resto de los sedimentos. Las arcillas son los sedimentos que en comparación con los otros sedimentos registran una menor dispersión entre sus datos.

Analizando las desviaciones medias absolutas de un punto central se puede concluir que en todos los casos expuestos en la tabla de dispersiones se tienen desviaciones de los datos menores respecto a la mediana que en comparación a las desviaciones respecto a las medias aritméticas y geométricas.

Los histogramas del rendimiento específico en sedimentos no consolidados son muy diferentes y varían para cada tipo (figura 4.71). En el caso de las arcillas la distribución es bimodal con las mayores frecuencias en 0.02 (-) y en 0.1 (-); la arena es unimodal con sesgo negativo y sus mayores frecuencias se encuentran en el valor de 0.3 (-); la grava también es unimodal con sesgo negativo y sus mayores frecuencias se encuentran en 0.3 (-); por su parte los limos son unimodales con

Tabla 4.77: Medidas de dispersión del rendimiento específico (-)

Litología	Rango	IQR	σ	DM	DMd	DMg
Clay	0.245	0.060	0.057	0.044	0.041	0.041
Gravel	0.292	0.110	0.083	0.068	0.067	0.080
Sand	0.435	0.149	0.100	0.083	0.082	0.089
Sand and Gravel	0.353	0.150	0.095	0.079	0.078	0.087
Sand silt and clay	0.252	0.110	0.072	0.057	0.054	0.054
Silt	0.480	0.050	0.101	0.057	0.052	0.053

Nota:

Tabla de comparación de las medidas de dispersión del rendimiento específico (-) en diferentes sedimentos no consolidados. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

cierta simetría y con una frecuencia alta en 0.08 (-).

La combinación de arena y grava tiene una distribución unimodal con sesgo negativo, donde sus mayores frecuencias se concentran en 0.2 (-); las combinaciones de arenas, limos y arcillas tienen una distribución unimodal, donde no se aprecia un sesgo significativo, y con una frecuencia alta en el valor de 0.3 (-).

4.3.6 Compresibilidad

Los registros de compresibilidad en los diferentes tipos de litologías pueden variar en más de 4 órdenes de magnitud. En la figura 4.72 se muestran los rangos que se pueden llegar a alcanzar las diferentes clasificaciones litológicas, siendo las rocas sedimentarias las que superan los 6 órdenes de magnitud, mientras que las rocas metamórficas son las que varían en menor grado con solo 4 órdenes de magnitud.

Los sedimentos no consolidados registran las mayores estimaciones de compresibilidad, con más de 4×10^{-6} (m^2/N). A su vez las rocas sedimentarias e ígneas registran las estimaciones más bajas con 1×10^{-11} (m^2/N) de compresibilidad.

Los datos de compresibilidad son muy escasos en este estudio, no obstante, se realizaron diagramas de caja en escala logarítmica (figura 4.73) con los datos recuperados de la base de datos. En esta figura se observa que en rocas ígneas los valores mayores a 1×10^{-8} (m^2/N) se consideran anómalos, las rocas metamórficas tienen valores anómalos en las estimaciones mayores a 1×10^{-10} (m^2/N), para las rocas sedimentarias se tiene valores anómalos en las mediciones mayores a 1×10^{-7}

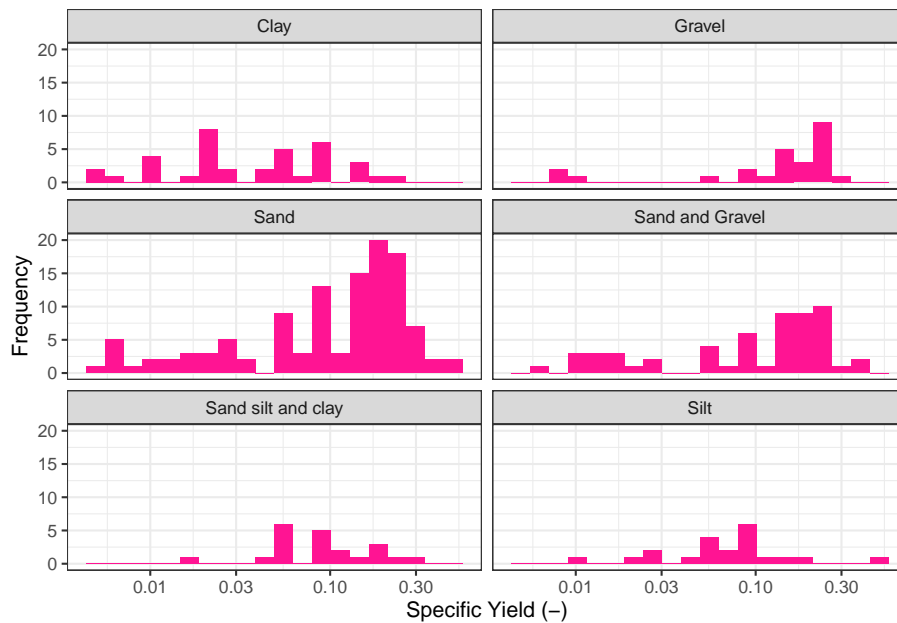


Figura 4.71: Histogramas en escala logarítmica en base 10 del rendimiento específico (-) en diferentes sedimentos no consolidados. Histogramas construidos a partir de 20 clases

y finalmente los sedimentos no consolidados tienen estimaciones anómalas en valores de compresibilidad mayores a 8×10^{-7} (m^2/N).

Las estimaciones máximas, mínimas y los cuartiles se registran en la tabla 4.78. El valor más alto de compresibilidad está presente en los sedimentos no consolidados con 5×10^{-6} (m^2/N), mientras que el valor más bajo de la compresibilidad es 1×10^{-10} (m^2/N) y se registra en rocas ígneas, sedimentarias y metamórficas.

Los sedimentos no consolidados son las litologías que alcanzan un mayor valor de la mediana con 1.2×10^{-7} (m^2/N), mientras que las rocas metamórficas son las que registran un valor menor de la mediana con 1.65×10^{-11} (m^2/N) de compresibilidad.

La tabla 4.79 muestra las medidas de tendencia central de los datos de la compresibilidad en los diferentes tipos de litologías de la base de datos. Analizando la tabla se puede concluir que la mediana y la media geométrica son más parecidas entre sí que con la media aritmética la cual difiere en hasta dos órdenes de magnitud.

La tabla 4.80 muestra las medidas de dispersión de los datos de compresibilidad (m^2/N) en los diferentes tipos de litología de la base de datos. Se puede concluir

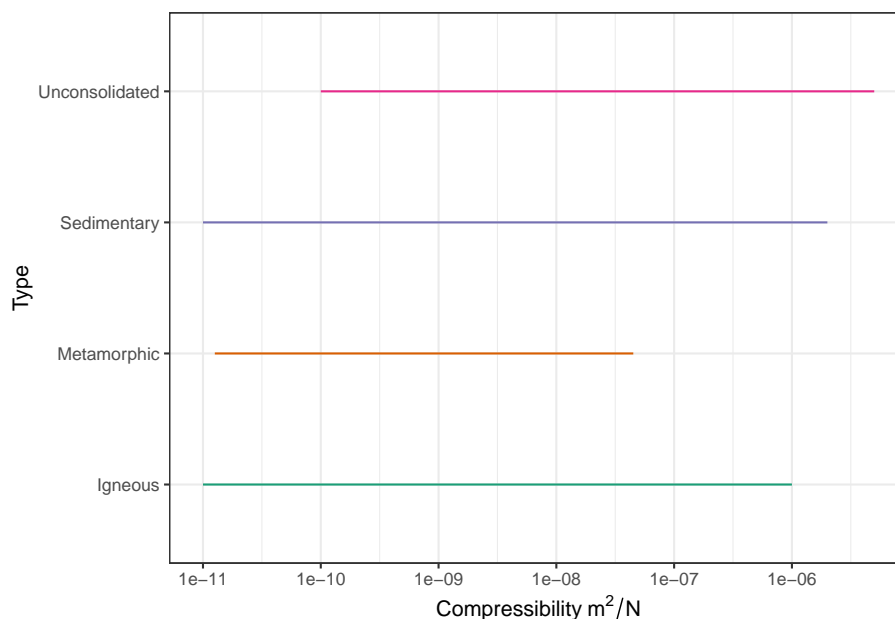


Figura 4.72: Diagrama de rangos en escala logarítmica en base 10 de la compresibilidad (m^2/N) para los diferentes tipos de litología.

Tabla 4.78: Valores mínimos, máximos y cuartiles de la compresibilidad (m^2/N)

Tipo	No.Datos	min.	Q1	Mediana	Q3	max.
Igneous	20	1.00e-11	8.00e-11	2.50e-09	9.35e-09	1.00e-06
Metamorphic	12	1.26e-11	1.30e-11	1.65e-11	1.55e-10	4.50e-08
Sedimentary	49	1.00e-11	4.50e-10	1.28e-09	6.90e-08	2.00e-06
Unconsolidated	109	1.00e-10	1.30e-08	1.20e-07	3.60e-07	5.00e-06

Nota:

Tabla de valores mínimos, máximos y cuartiles Q1, Q2 (Mediana), Q3 de la compresibilidad (m^2/N) para las diferentes clasificaciones litológicas. La columna **No. datos** corresponde con el número de observaciones.

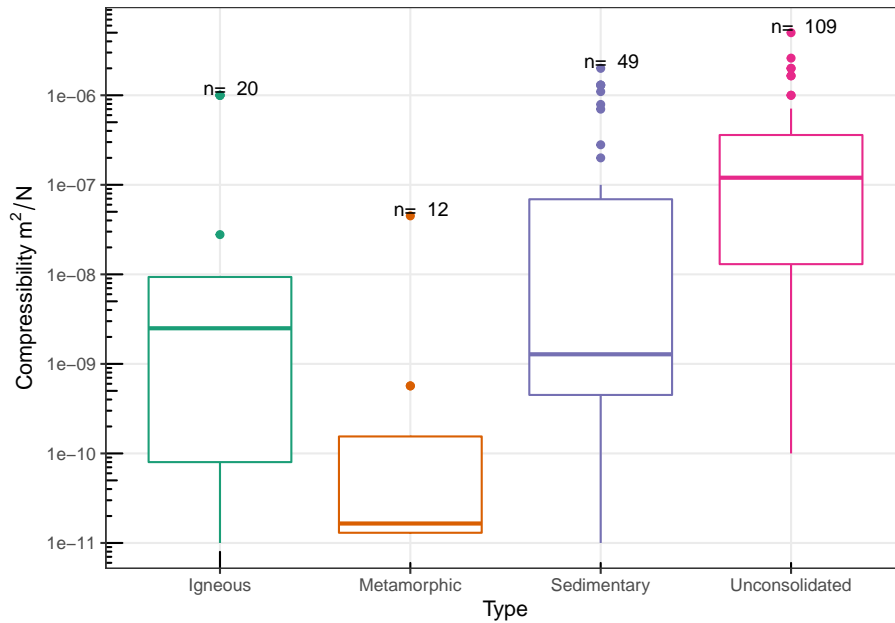


Figura 4.73: Diagramas de caja en escala logarítmica en base 10 de la compresibilidad (m^2/N) para las diferentes clasificaciones litológicas.

Tabla 4.79: Comparación de las medidas de tendencia central de la compresibilidad (m^2/N)

Tipo	Media	Media_geométrica	Mediana
Igneous	2.03e-07	1.96e-09	2.50e-09
Metamorphic	3.86e-09	5.30e-11	1.65e-11
Sedimentary	1.95e-07	4.22e-09	1.28e-09
Unconsolidated	3.30e-07	7.35e-08	1.20e-07

Nota:

Tabla de comparación de las medidas de tendencia central (media, media geométrica y mediana) de la compresibilidad (m^2/N) en los diferentes tipos de litologías.

Tabla 4.80: Medidas de dispersión de la compresibilidad (m^2/N)

Tipo	Rango	IQR	σ	DM	DMd	DMg
Igneous	1.00e-06	9.27e-09	4.09e-07	3.19e-07	2.02e-07	2.02e-07
Metamorphic	4.50e-08	1.42e-10	1.30e-08	6.86e-09	3.84e-09	3.86e-09
Sedimentary	2.00e-06	6.86e-08	4.49e-07	2.94e-07	1.95e-07	1.95e-07
Unconsolidated	5.00e-06	3.47e-07	6.32e-07	3.36e-07	2.95e-07	3.00e-07

Nota:

Tabla de comparación de las medidas de dispersión de la compresibilidad (m^2/N) en los diferentes tipos de litología. Donde se presenta el rango (Rango); rango intercuartílico (IQR); desviación estándar (σ). Así como las desviaciones medias absolutas de un punto central: desviación respecto a la media (DM); desviación respecto a la mediana (DMd); y desviación respecto a la media geométrica (DMg).

de acuerdo con la comparación de las desviaciones estándar que los sedimentos no consolidados son las litologías que presentan una mayor dispersión entre sus datos de compresibilidad, a diferencia de las rocas metamórficas donde la dispersión entre sus datos es menor en comparación con las dispersiones de los otros tipos de litología.

Se puede inferir de acuerdo con análisis de las desviaciones medias absolutas de un punto central que en cada caso la desviación de los datos respecto a la media resulta ser menor que las desviaciones respecto a la media geométrica y la media aritmética. En los casos específicos de rocas ígneas y rocas sedimentarias la desviación respecto a la mediana resulta ser bastante parecida a la desviación respecto a la mediana.

La figura 4.74 presenta los diferentes histogramas para las diferentes clasificaciones litológicas, sin embargo, por la falta de datos resultan poco representativos los casos de rocas ígneas y metamórficas.

El histograma de rocas sedimentarias es unimodal y tiende a tener una ligera simetría, en donde las mayores frecuencias se encuentran en valores cercanos a $1 \times 10^{-9} (m^2/N)$. Para el caso de los sedimentos no consolidados, el histograma es unimodal con un sesgo positivo en el cual las mayores frecuencias se concentran en estimaciones cercanas a $5 \times 10^{-7} (m^2/N)$.

Correlaciones

La figura 4.75 muestra las correlaciones y los modelos de regresión de la compresibilidad en función de cada uno de los parámetros hidrogeológicos (porosidad total, porosidad efectiva, conductividad hidráulica, coeficiente de almacenamiento y rendimiento específico). El análisis muestra que, en las gráficas A, B, C y E de la figura, los modelos de regresión no son lo suficientemente buenos para explicar un porcentaje importante de la proporción de la varianza de la variable dependiente. Por su parte el diagrama

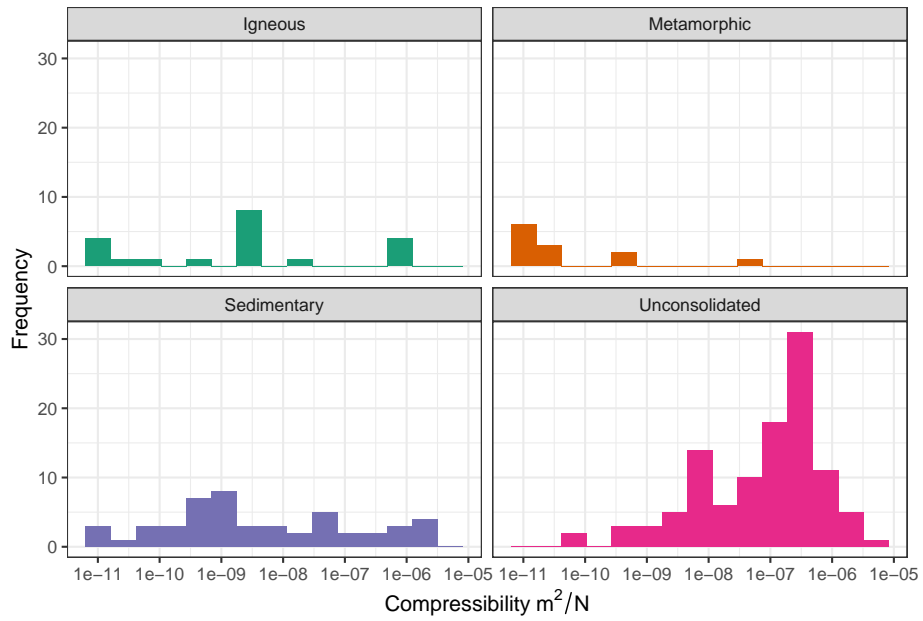


Figura 4.74: Histogramas en escala logarítmica en base 10 de la compresibilidad (m^2/N) para las diferentes clasificaciones litológicas. Histogramas construidos a partir de 15 clases

D (compresibilidad vs Coeficiente de almacenamiento) es el que presenta un mejor ajuste en comparación con los otros casos, con un coeficiente de determinación de 0.48.

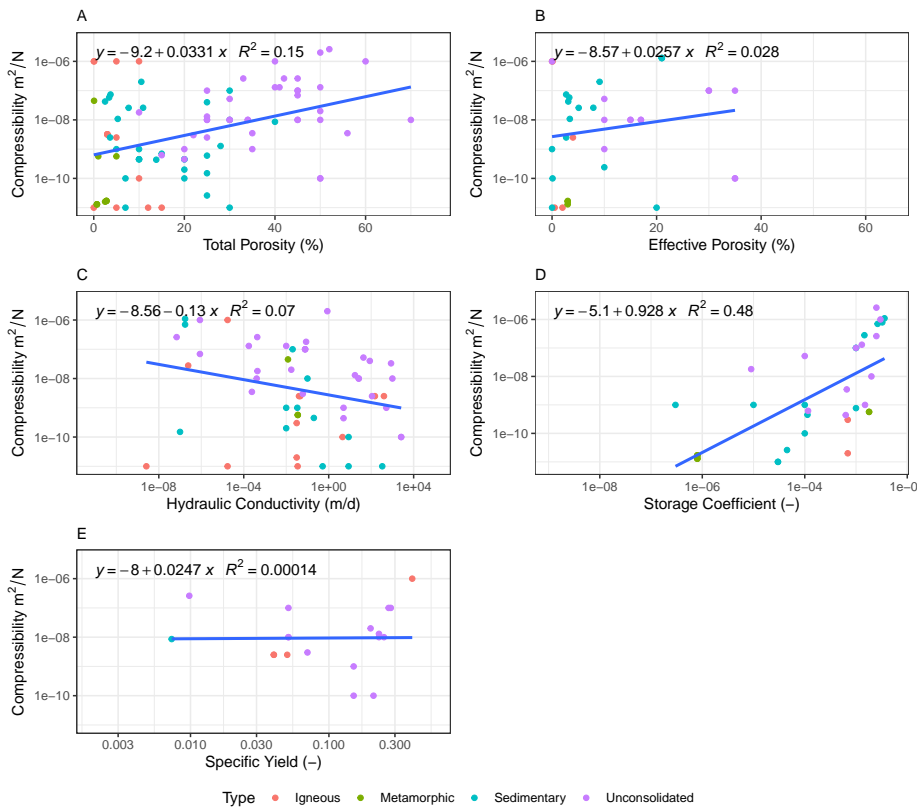


Figura 4.75: Diagramas de correlación de la compresibilidad en función de las diferentes propiedades hidrogeológicas. A: Compresibilidad vs Porosidad Total (transformación logarítmica en base 10 del eje y). B: Compresibilidad vs Porosidad Efectiva (transformación logarítmica en base 10 del eje y). C: Compresibilidad vs Conductividad Hidráulica (transformación logarítmica en base 10 del eje y, y el eje x). D: Compresibilidad vs Coeficiente de Almacenamiento (transformación logarítmica en base 10 del eje y, y el eje x). E: Compresibilidad vs Rendimiento Específico (transformación logarítmica en base 10 del eje y, y eje x)

Capítulo 5

Conclusiones

La base de datos cumple con todas las características necesarias para su máximo aprovechamiento, por una parte, cuenta con una gran cantidad de datos de los diferentes parámetros hidrogeológicos para cada una de las litologías más representativas de una forma ordenada y clara; y por otro lado el acceso a esta información es completamente libre y fácil, por lo que brinda una herramienta poderosa para la comunidad científica y académica.

La base de datos de parámetros hidrogeológicos creada es una herramienta con muchas aplicaciones, entre los principales: almacenar, localizar, seleccionar, coleccionar, procesar, información geológica e hidrogeológica. Una característica muy importante es su función como un medio de consulta que facilita el acceso a información puntual, así como a sus correspondientes referencias, si lo que se busca es un estudio más profundo.

La cantidad de datos que se proporcionan mediante este recurso brinda posibilidades ilimitadas para el análisis y la inferencia estadística. Un ejemplo claro es la capacidad para realizar una estadística descriptiva; la cual se desarrolla en este documento, arrojando información muy completa gracias a la cantidad de información a partir de la cual se desarrolló.

Existen muchas herramientas para el análisis estadístico, sin embargo, el uso del lenguaje de programación R tiene muchas ventajas ya que permite implementar y optimizar procesos, así como brindar libertad en el manejo de datos. R no solo brinda herramientas de computación estadística si no que ofrece una infinidad de paqueterías que permiten la comunicación de los resultados. Una de las ventajas de usar esta herramienta es el soporte de la comunidad de R.

El análisis estadístico fue muy provechoso y permitió localizar y analizar valores anómalos mediante los diagramas de caja dando como resultado un mejor entendimiento del rango de valores en los que varía cada una de las propiedades hidrogeológicas. También fue posible apreciar como las transformaciones y

escalas logarítmicas permiten un mejor manejo y análisis de datos que varían en un gran número de ordenes de magnitud como lo fueron los casos de la conductividad hidráulica y el coeficiente de almacenamiento.

Los métodos estadísticos descriptivos expuestos en este trabajo arrojaron información interesante: como el hecho de que, en la mayoría de las litologías, la mejor medida de tendencia central, cuando existen muchos valores anómalos y cuando se manejan muchos órdenes de magnitud en los datos, es la mediana ya que es la medida de tendencia central que menos es afectada por valores atípicos y extremos. Los histogramas son una herramienta indispensable que permitió identificar diferentes sesgos en la distribución de los datos, así como el comportamiento de las mayores concentraciones de datos en ciertos rangos para cada una de las propiedades hidrogeológicas. En general se puede inferir a partir de los patrones que las diferentes propiedades texturales y estructurales dentro de diferentes litologías del mismo tipo pueden influir en gran medida el valor de las diferentes propiedades hidrogeológicas.

Una de las partes más importantes en un trabajo de investigación es la capacidad de comunicar resultados, es por eso por lo que la página de consulta creada en este trabajo resulta en una herramienta poderosa con material didáctico e interactivo que puede ser usado por la comunidad académica y científica.

Este trabajo también resultó en una guía práctica paso por paso para un análisis estadística creando gráficos y tablas con “R”, así como para realizar y comunicar resultados con las diferentes paqueterías de R. Muestra de ello son los scripts y sus respectivas descripciones que pueden encontrarse a lo largo de la metodología y que muestran el procedimiento que se llevó a cabo para realizar gráficos, tablas y una página de consulta de la información creada a partir de la base de datos.

Capítulo 6

Referencias citadas en el texto

ADES (2008). *ADES groundwater national portal*.

Allen, D., Brewerton, L., Coleby, L., Gibbs, B., Lewis, M., MacDonald, A., Wagstaff, S., and Williams, A. (1997). The physical properties of major aquifers in England and Wales.

Anderson, M. P., Aiken, J., Webb, E., and Mickelson, D. (1999). Sedimentology and hydrogeology of two braided stream deposits. *Sedimentary Geology*, 129(3-4):187–199.

Bañuelos, A. and Manzanares, N. (2016). *Probabilidad y Estadística. Tema 1 análisis estadístico de datos muestrales (Apuntes de clase)*. Universidad Nacional Autónoma de México.

Bolster, C. H., Genereux, D. P., and Saiers, J. E. (2001). Determination of specific yield for the Biscayne Aquifer with a canal-drawdown test. *Ground Water*, 39(5):768–777.

Brassington (2007). *Field Hydrogeology*. John Wiley and Sons, U.K.

Byers, R. A., Yáñez, L. H., Urrea, M. F. H., and Sánchez, A. V. (1987). *Introducción a las bases de datos con dBase III Plus*. Number 005.7565 B94 1987. McGraw-Hill.

Carrera-Hernández, J. and Gaskin, S. (2007). The Basin of Mexico aquifer system: regional groundwater level dynamics and database development. *Hydrogeology Journal*, 15(8):1577–1590.

CGDI (2007). *The Canadian geospatial data infrastructure*.

- Cheong, J.-Y., Hamm, S.-Y., Kim, H.-S., Ko, E.-J., Yang, K., and Lee, J.-H. (2008). Estimating hydraulic conductivity using grain-size analyses, aquifer tests, and numerical modeling in a riverside alluvial system in South Korea. *Hydrogeology Journal*, 16(6):1129.
- Chinnasamy, P., Maheshwari, B., Dillon, P., Purohit, R., Dashora, Y., Soni, P., and Dashora, R. (2018). Estimation of specific yield using water table fluctuations and cropped area in a hardrock aquifer system of Rajasthan, India. *Agricultural Water Management*, 202:146–155.
- Comunian, A. and Renard, P. (2009). Introducing wwhypda: a world-wide collaborative hydrogeological parameters database. *Hydrogeology Journal*, 17(2):481–489.
- Cordero Sánchez, J. (2013). Bases de datos en R : análisis gráfico y estadístico de valores atípicos y ausentes.
- Dagan, G. (2002). An overview of stochastic modeling of groundwater flow and transport: From theory to applications. *Eos, Transactions American Geophysical Union*, 83(53):621–625.
- Dalgaard, P. (2005). ISwR: introductory statistics with R. *R package version*, pages 1–6.
- Darcy, H. (1856). Les fontaines publiques de la ville de Dijon, Dalmont. 1956.
- Davis, S. N. and DeWiest, R. J. (1966). Hydrogeology. Technical report.
- De Dreuzy, J.-R., Bodin, J., Le Grand, H., Davy, P., Boulanger, D., Battais, A., Bour, O., Gouze, P., and Porel, G. (2006). General database for ground water site information. *Groundwater*, 44(5):743–748.
- De Marsily, G., Delay, F., Gonçalves, J., Renard, P., Teles, V., and Violette, S. (2005). Dealing with spatial heterogeneity. *Hydrogeology Journal*, 13(1):161–183.
- Delleur L, J. W. (2006). Geological Occurrence of Groundwater. In *The Handbook of Groundwater Engineering*, pages 63–122. CRC Press.
- Devore, J. L. (2008). *Probabilidad y estadística para ingenierías y ciencias*. Cengage Learning Editores.
- Domenico, P. A., Schwartz, F. W., et al. (1998). *Physical and chemical hydrogeology*, volume 506. Wiley New York.
- Einstein, H. H. and Baecher, G. B. (1983). Probabilistic and statistical methods in engineering geology. *Rock mechanics and rock engineering*, 16(1):39–72.
- Field, A. P., Miles, J., and Field, Z. (2012). *Discovering statistics using R/Andy Field, Jeremy Miles, Zoë Field*. London; Thousand Oaks, Calif.

- Foster, S. S., Ellis, A. T., Losilla-Penon, M., and Rodriguez-Estrada, H. V. (1985). Role of Volcanic Tuffs in Ground-Water Regime of Valle Central, Costa Rica. *Groundwater*, 23(6):795–801.
- Freeze and RA, C. (1979). *Groundwater. Englewood Cliff.*
- Freeze, R. A., Massmann, J., Smith, L., Sperling, T., and James, B. (1990). Hydrogeological decision analysis: 1. A framework. *Groundwater*, 28(5):738–766.
- Fretwell, B., Burgess, W., Barker, J., and Jefferies, N. (2005). Redistribution of contaminants by a fluctuating water table in a micro-porous, double-porosity aquifer: Field observations and model simulations. *Journal of contaminant hydrology*, 78(1-2):27–52.
- Gilli, ., Mangan, C., and Mudry, J. (2012). *Hydrogeology: objectives, methods, applications.* CRC Press.
- Golian, M., Teshnizi, E. S., and Nakhaei, M. (2018). Prediction of water inflow to mechanized tunnels during tunnel-boring-machine advance using numerical simulation. *Hydrogeology Journal*, 26(8):2827–2851.
- Gupta, A. S. and Rao, S. K. (2001). Weathering indices and their applicability for crystalline rocks. *Bulletin of Engineering Geology and the Environment*, 60(3):201–221.
- Hahn, J., Lee, Y., Kim, N., Hahn, C., and Lee, S. (1997). The groundwater resources and sustainable yield of Cheju volcanic island, Korea. *Environmental Geology*, 33(1):43–53.
- Heath, R. C. (2004). *Basic ground-water hydrology.* US Geological Survey Reston, VA.
- Hiscock, K. M. (2009). *Hydrogeology: principles and practice.* John Wiley and Sons.
- Huang, Y., Scanlon, B. R., Nicot, J.-P., Reedy, R. C., Dutton, A. R., Kelley, V. A., and Deeds, N. E. (2012). Sources of groundwater pumpage in a layered aquifer system in the Upper Gulf Coastal Plain, USA. *Hydrogeology Journal*, 20(4):783–796.
- Jeong, J., Park, E., Han, W. S., Kim, K., Choung, S., and Chung, I. M. (2017). Identifying outliers of non-Gaussian groundwater state data based on ensemble estimation for long-term trends. *Journal of Hydrology*, 548:135–144.
- John, J. and Draper, N. R. (1980). An alternative family of transformations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(2):190–197.
- Jones, H., Morris, B., Cheney, C., Brewerton, L., Merrin, P., Lewis, M.,

- MacDonald, A., Coleby, L., Talbot, J., McKenzie, A., et al. (2000). The physical properties of minor aquifers in England and Wales.
- Kruseman, G. P., De Ridder, N. A., and Verweij, J. M. (1970). *Analysis and evaluation of pumping test data*, volume 11. International institute for land reclamation and improvement The Netherlands.
- Kumar, T. J. R., Balasubramanian, A., Kumar, R., Dushiyanthan, C., Thiruneelakandan, B., Suresh, R., Karthikeyan, K., and Davidraju, D. (2016). Assessment of groundwater potential based on aquifer properties of hard rock terrain in the Chittar–Uppodai watershed, Tamil Nadu, India. *Applied Water Science*, 6(2):179–186.
- MacDonald, A. M. and Allen, D. J. (2001). Aquifer properties of the Chalk of England. *Quarterly Journal of Engineering Geology and Hydrogeology*, 34(4):371–384.
- Mendenhall, W., Scheaffer, R. L., Wackerly, D. D., De la Fuente Pantoja, A., and Verbeeck, D. V. (1986). *Estadística matemática con aplicaciones*. Number 519.5 M4E8 QA276 M4E8. Grupo Editorial Iberoamericana California.
- Morandat, F., Hill, B., Osvald, L., and Vitek, J. (2012). Evaluating the design of the R language. In *European Conference on Object-Oriented Programming*, pages 104–131. Springer.
- Motyka, J. (1998). A conceptual model of hydraulic networks in carbonate rocks, illustrated by examples from Poland. *Hydrogeology Journal*, 6(4):469–482.
- Muñoz, D. R. (2000). *Manual de estadística*. Juan Carlos Martínez Coll.
- Newell, C. J., Hopkins, L. P., and Bedient, P. B. (1990). A hydrogeologic database for ground-water modeling. *Groundwater*, 28(5):703–714.
- Olea, R. A. (2009). *A practical primer on geostatistics*. US Geological Survey.
- Olivier, J., Johnson, W. D., and Marshall, G. D. (2008). The logarithmic transformation and the geometric mean in reporting experimental IgE results: what are they and when and why to use them? *Annals of Allergy, Asthma & Immunology*, 100(4):333–337.
- Paradis, E. (2002). *R para Principiantes*. Institut des Sciences de l'Évolution. Universit Montpellier.
- Poehls, D. and Smith, G. J. (2011). *Encyclopedic dictionary of hydrogeology*. Academic press.
- Rivera, M. d. C. G. (1994). La base de datos. Importancia y aplicación en educación. *Perfiles educativos*, (65).
- Rotzoll, K., El-Kadi, A. I., and Gingerich, S. B. (2007). Estimating Hydraulic Properties of Volcanic Aquifers Using Constant-Rate and Variable-Rate Aquifer Tests 1. *JAWRA Journal of the American Water Resources Association*, 43(2):334–345.

- Seo, S. (2006). *A review and comparison of methods for detecting outliers in univariate data sets*. PhD thesis, University of Pittsburgh.
- Sharma, K. and Kumar, S. (2008). Hydrogeological research in India in managing water resources. In *Groundwater dynamics in hard rock aquifers*, pages 1–19. Springer.
- Singhal, B. B. S. and Gupta, R. P. (2010). *Applied hydrogeology of fractured rocks*. Springer Science & Business Media.
- Spiegel, M. R. and Stephens, L. J. (2009). *Estadística. Mac Graw Hill*, 4.
- Stuart, A., Ord, J. K., and Kendall, M. G. (1991). *Kendall's Advanced Theory of Statistics: Classical Inference and Relationship*. Edward Arnold.
- Tartakovsky, Daniel M, W. C. and Larrabee (2008). Uncertain future of hydrogeology. *Journal of Hydrologic Engineering*, 13(1):37–39.
- Taylor, G. and Eggleton, R. A. (2001). *Regolith geology and geomorphology*. John Wiley & Sons.
- Team, R. C. (2000). R language definition. *Vienna, Austria: R foundation for statistical computing*.
- Team, R. C. (2020). Shiny-Rstudio. *Boston, E.U.*
- Terzaghi (1925). *Erdbaumechanik auf bodenphysikalischer Grundlage*.
- Tukey, J. W. (1977). *Exploratory data analysis*, volume 2. Reading, MA.
- USGS (2008). *Ground water atlas of the Unites States*. United States Geological Survey, Reston, VA.
- Vittecoq, B., Reninger, P.-A., Violette, S., Martelet, G., Dewandel, B., and Audru, J.-C. (2015). Heterogeneity of hydrodynamic properties and groundwater circulation of a coastal andesitic volcanic aquifer controlled by tectonic induced faults and rock fracturing–Martinique island (Lesser Antilles–FWI). *Journal of Hydrology*, 529:1041–1059.
- White, W. B. (2016). Groundwater flow in karstic aquifers. In *The handbook of groundwater engineering*, pages 581–614. CRC Press.
- Wickham, H. and Golemund, G. (2016). *R for data science: import, tidy, transform, visualize, and mdel data*. O'Reilly Media, Inc.
- Won, J.-H., Lee, J.-Y., Kim, J.-W., and Koh, G.-W. (2006). Groundwater occurrence on Jeju Island, korea. *Hydrogeology Journal*, 14(4):532–547.
- Wösten, J., Lilly, A., Nemes, A., and Le Bas, C. (1999). Development and use of a database of hydraulic properties of European soils. *Geoderma*, 90(3-4):169–185.