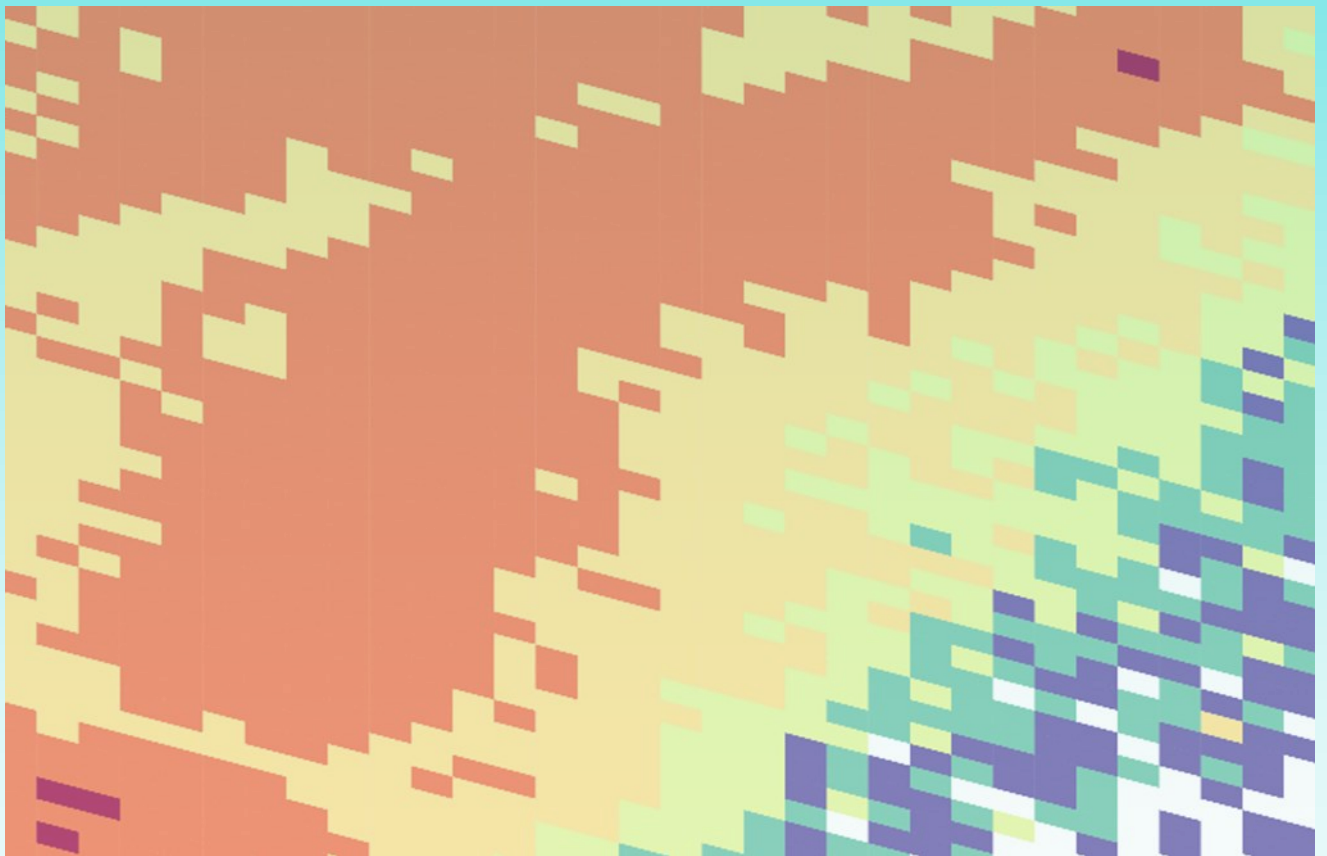


D-Piper (Diagrama Piper de Densidad) Representación de grandes conjuntos de datos mediante el diagrama de Piper Manual de Uso



Luis Moreno Merino
Héctor Aguilera Alonso
Miguel González Jiménez
Elisabeth Díaz Losada
Amalia Romero Prados

D-Piper

(Diagrama Piper de Densidad)

Representación de grandes conjuntos de datos mediante el diagrama de Piper

Manual de Uso

Luis Moreno Merino
Héctor Aguilera Alonso
Miguel González Jiménez
Elisabeth Díaz Losada
Amalia Romero Prados



Serie: HIDROGEOLOGÍA Y AGUAS SUBTERRÁNEAS, Nº 39.

D-Piper (Diagrama Piper de Densidad) Representación de grandes conjuntos de datos mediante el diagrama de Piper. Manual de Uso / Instituto Geológico y Minero de España. Área de Geología Ambiental y Geomatemáticas. Moreno Merino, L., Aguilera Alonso, H., González Jiménez, M., Díaz Losada, E., Romero Prados, A. – Madrid: Instituto Geológico y Minero de España (IGME, CSIC), 2021.

36 pgs; 23 figs; 29 cm
I.S.B.N.: 978-84-9138-116-7

1. Diagrama de Piper. 2. Hidroquímica. 3. Hidrogeología. 4. Leguaje Phytton.
I. Instituto Geológico y Minero de España. II. Moreno Merino, L., Aguilera Alonso, H., González Jiménez, M., Díaz Losada, E., Romero Prados, A.

Ninguna parte de este libro puede ser reproducida o transmitida en cualquier forma o por cualquier medio, electrónico o mecánico, incluido fotografías, grabación o por cualquier sistema de almacenar información sin el previo permiso escrito del autor y editor.

Realización y composición: Instituto Geológico y Minero de España (IGME, CSIC)

IMAGEN DE PORTADA:

Fragmento de un campo de densidades de facies dibujado mediante el programa D-Piper.

© INSTITUTO GEOLÓGICO Y MINERO DE ESPAÑA
C/Ríos Rosas, 23 – 28003 Madrid
Tel: +343495700. Fax: 91 8032200
<http://www.igme.es>

NIPO: 836-21-011-9
ISBN: 978-84-9138-116-7

Índice

INTRODUCCIÓN, ¿QUÉ ES D-PIPER?	3
¿CÓMO UTILIZAR D-PIPER?	11
Función de las diferentes carpetas y archivos	12
Estructura del fichero de datos	13
Opciones disponibles, descripción del fichero D_Piper_v1_options.txt	14
Opciones generales	14
Opciones de grupo	18
Opciones de distribución de puntos y gráficas	21
Histograma de frecuencias	23
Opciones de archivo	25
Opciones de archivos gráficos	26
INTRODUCCIÓN A LOS MÉTODOS DE REPRESENTACIÓN EMPLEADOS EN D-PIPER .	27
NOTAS SOBRE EL CÓDIGO	31
REFERENCIAS	35

INTRODUCCIÓN, ¿QUÉ ES D-PIPER?

D-Piper es un software diseñado para representar grandes conjuntos de análisis químicos de agua en un único gráfico de Piper conservando toda la información relevante. Los grupos de análisis representados pueden tener cualquier tamaño, no hay límite, de hecho cuanto más grande es el grupo mayor es la utilidad del gráfico pues la estructura interna de la información se desvela con más claridad. D-Piper también permite dibujar los diagramas de Piper clásicos e incluye en los mismos, además de opciones como el color y tamaño de los puntos y la posibilidad de superponer el diagrama clásico y el D-Piper, otras menos frecuentes como incluir puntos traslúcidos o modificar el tamaño de las celdas del diagrama. En este documento se explica en detalle la utilidad y funcionamiento de D-Piper, se muestran los principios matemáticos sobre los que se fundamenta su construcción y se incluye información de utilidad para quien desee modificar el código. Se trata de un software gratuito y de libre distribución, escrito en lenguaje Python y distribuido bajo licencia **GPL 3.0**, el código puede ser usado y compartido libremente con la única condición de citar adecuadamente la fuente original no permitiéndose el uso comercial. Los aspectos técnicos y de desarrollo, así como el código fuente, han sido publicados en la revista *Environmental Modelling and Software* ([Moreno et al, 2021](#)) constituyendo este documento un manual detallado de uso.

La representación gráfica es una herramienta muy común para la interpretación de análisis químicos de aguas subterráneas. Los gráficos hidroquímicos y estadísticos, son empleados tanto para comparar la evolución temporal y espacial de distintas muestras de agua, como para estudiar procesos de mezcla, intrusión salina, procesos de ablandamiento o endurecimiento del agua, oxidación de sulfuros y muchas otras reacciones químicas que modifican la proporción de elementos mayoritarios disueltos en el agua. En algunos de estos gráficos, como es el caso del de Piper, en lugar de representar las concentraciones absolutas de aniones y cationes, se muestran sus proporciones relativas y con ello las facies hidroquímicas.

De entre la multitud de gráficos hidroquímicos disponibles ([Figura 1](#)), en la actualidad los más empleados son los propuestos por [Piper \(1944\)](#), [Durov \(1948\)](#), [Stiff \(1951\)](#) y [Schoeller \(1964\)](#) y en menor medida los diagramas de [Chadha \(1999\)](#) y de [Collins \(1923\)](#). Por supuesto que también son de uso común los diagramas de barras y otros de uso general en otros campos de la ciencia y la técnica como los gráficos de tartas o los gráficos de dispersión.

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

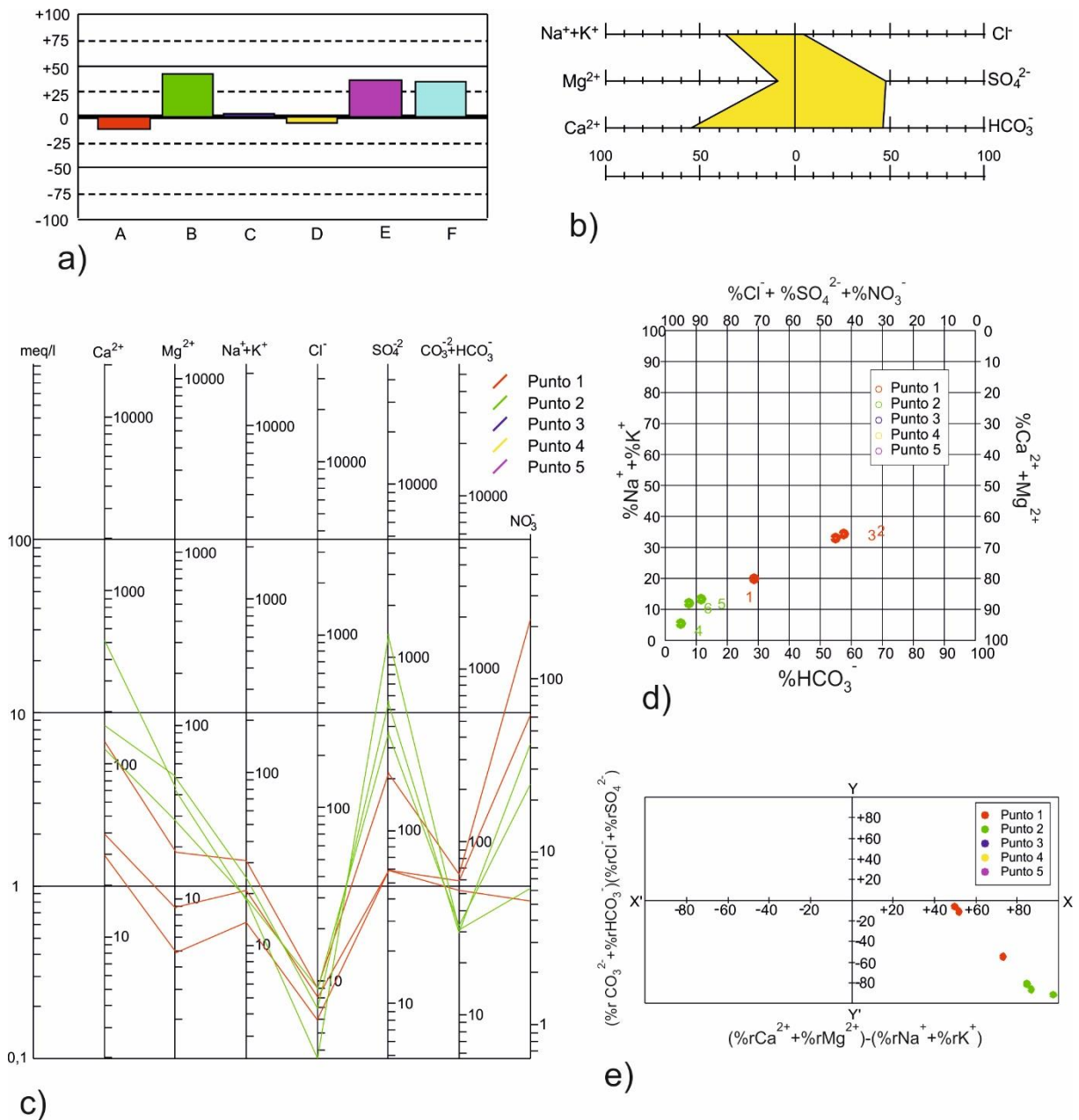


Figura 1.- Principales gráficos empleados en el estudio hidroquímico de las aguas subterráneas. a) Diagrama de Barras representando los índices de D'Amore (1983). b) Gráfico de Stiff (1951). Frecuentemente empleado para representar las facies hidroquímicas. Tiene como principal ventaja su capacidad de presentar en un mismo gráfico facies y concentración de iones mayoritarios. Además, es muy adecuado para la representación de estos parámetros sobre cualquier tipo de cartografía. Su principal inconveniente es que no permite representar más que un único análisis en cada gráfico. c) Diagrama de Schoeller (1964). La representación sobre ejes logarítmicos de las concentraciones iónicas permite mostrar en un único gráfico, de forma simultánea, análisis en un rango de concentración extremadamente amplio. d) Gráfico de Durov (1948). Tiene un uso e interpretación similar al de Piper (1944) aunque se emplea menos debido a que es menos compacto. e) Gráfico de Chadha (1999). Es el menos empleado de todos los presentados en esta figura. Permite representar diversas relaciones iónicas asociadas a matrices minerales específicas. Todos los gráficos presentados en esta figura han sido realizados mediante la Hoja de Cálculo Inaques (Moreno *et al*, 2006) disponible en la sección de publicaciones de la página web del IGME: <http://www.igme.es>

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

La idea de representar análisis de agua mediante diagramas triangulares surgió simultáneamente al principio de la década de 1940 cuando Hill (1940) y Langelier and Ludwig (1942) propusieron de forma independiente el mismo método de representación. Posteriormente Piper (Piper, 1944) diseñó el diagrama que lleva su nombre y que se describirá más adelante. Desde entonces han aparecido algunas propuestas nuevas y numerosas variaciones de este diagrama como el mencionado diagrama de Durov (Durov, 1948); el diagrama de Durov extendido (Burdon and Mazloum, 1958), posteriormente mejorado por Lloyd (1965), o incluso diagramas de concepción totalmente novedosa como el de Chadha (1999).

La mayor parte de estos gráficos fueron diseñados hace mucho más de 50 años y aunque son de uso muy extendido ninguno fue pensado para mostrar de manera simultánea más allá de algunas decenas de análisis, nunca cientos ni mucho menos miles.

La utilidad del diagrama de Piper en el campo de la geoquímica y de la hidroquímica no se limita a la clasificación, agrupación e identificación de las aguas por su facies. Mediante este diagrama también se pueden estudiar procesos de mezcla y algunas reacciones químicas. Tanto es así que algunas variantes han sido adaptadas para su uso en otros campos diferentes a la hidroquímica. Por ejemplo, Teng *et al.* (2016) sugirieron emplear el diagrama de Piper como herramienta de visualización para el diseño y optimización de procesos de ingeniería química en la producción de cinnamaldeído; Ray y Mukherjee (2008) propusieron la representación de una combinación de los iones mayoritarios en coordenadas rectangulares, reproduciendo un esquema similar al diagrama de Piper pero mucho más fácil de implementar en una hoja de cálculo; Shelton *et al.* (2018) propusieron una modificación del diagrama de Piper empleando la técnica del análisis composicional. No obstante, este tipo de diagramas no han sido empleados de forma extensiva, pues en realidad no proporcionan información diferente a la que puede obtenerse del diagrama original o su interpretación es muy compleja, mientras que la interpretación del Piper original es notablemente intuitiva y se ha convertido de facto en la forma de representar las facies hidroquímicas.

En el diagrama de Piper (Figura 2) se muestra de forma simultánea la proporción existente entre cationes y aniones mayoritarios expresados en porcentaje de equivalentes por millón (%epm). Para construirlo se dibujan tres paneles: dos triángulos y un rombo. En el triángulo de la izquierda se representan los cationes (Ca^{2+} , Mg^{2+} , $\text{Na}^{+}+\text{K}^{+}$), en el de la derecha los aniones ($\text{HCO}_3^{-}+\text{CO}_3^{2-}$, SO_4^{2-} , Cl^{-}) y en el rombo se proyectan los puntos representados en ambos triángulos.

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

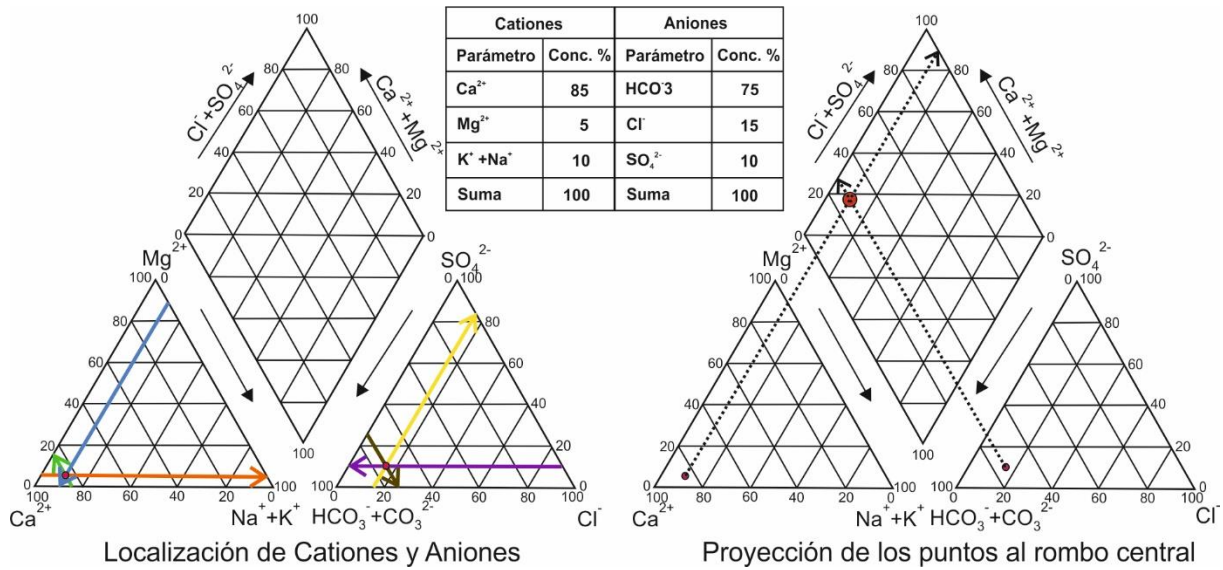


Figura 2. Ejemplo de construcción de un diagrama de Piper. El análisis representado presenta un contenido en cationes mayoritarios del 85% de Ca²⁺, un 5% de Mg²⁺ y un 10% de Na⁺+K⁺. Por su parte, el contenido en aniones mayoritarios es de un 18% de Cl⁻, un 17% de SO₄²⁻ y un 70% de HCO₃+CO₃. La posición de los puntos que representan los análisis sobre los triángulos se obtiene de la intersección de las líneas tal como se observa en la figura. Las líneas punteadas muestran la proyección de los puntos representados en los triángulos sobre el rombo central.

En la **Figura 3**, se presenta un ejemplo de uso del diagrama de Piper en su forma más habitual.

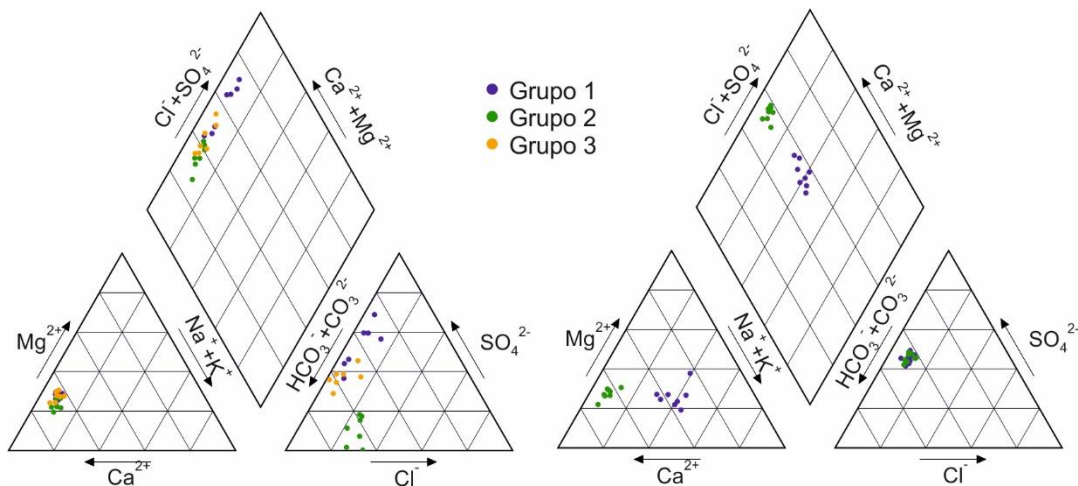


Figura 3. Ejemplo de aplicación del diagrama de Piper en su forma más habitual. El gráfico de la izquierda permite comparar tres grupos de muestras de agua de diferente procedencia y facies. Cada grupo puede ser fácilmente distinguido en función de su facies aniónica, mientras que las facies catiónicas son muy similares y no permiten diferenciarlos. En el gráfico de la derecha se representan dos grupos de agua procedentes del mismo acuífero, donde el grupo azul ha sufrido un proceso de ablandamiento debido al intercambio de Ca²⁺ por Na⁺ por lo que, a pesar de proceder del mismo acuífero su facies se ha trasladado desde posiciones cálcicas o cálcico-magnésicas hacia facies sódicas.

Numerosos problemas hidroquímicos tienen que ver con procesos de mezcla de aguas de diferente procedencia y de diferente facies hidroquímicas. Con frecuencia, este tipo de procesos

pueden ser ilustrados y analizados mediante un diagrama de Piper, aunque se debe cumplir la condición de la no existencia de procesos de precipitación como consecuencia de la mezcla. Como se puede ver en la [Figura 4a](#), bajo estas condiciones, cualquier proporción de mezcla se sitúa sobre una línea que une los puntos que definen las aguas puras sin mezcla. En el caso de tratarse de una mezcla de tres aguas, [Figura 4b](#), y bajo las mismas circunstancias, todas las aguas resultantes de las mezclas se situarían en el interior del triángulo cuyos vértices ocupan las soluciones sin mezcla.

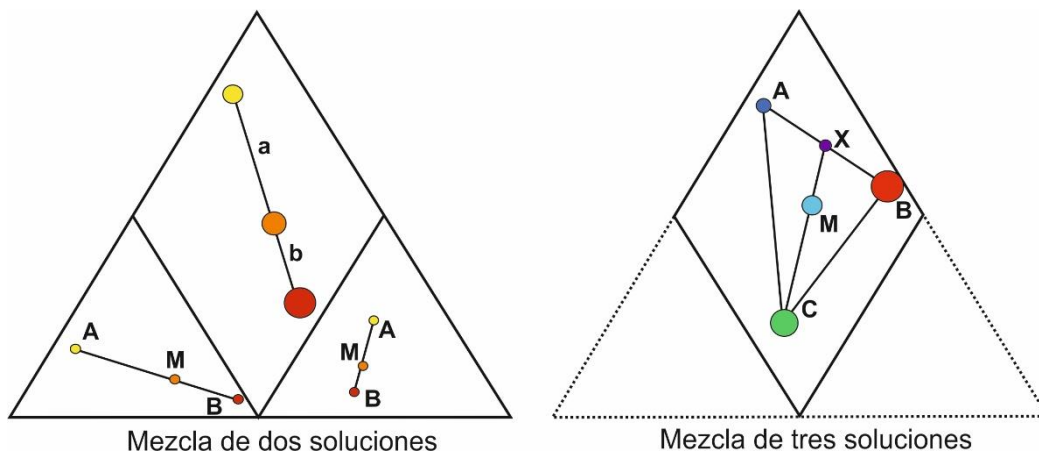


Figura 4. Ejemplo de aplicación del diagrama de Piper que ilustra en la figura de la izquierda aguas procedentes de la mezcla de dos soluciones. En la figura de la derecha se representa la mezcla de tres soluciones. (Modificado del trabajo original de [Piper \(1944\)](#))

En el diagrama de Piper siempre es cierto que la representación del agua resultante de la mezcla de dos distintas cae sobre la línea que une los puntos extremos, correspondientes a cada una de las aguas que han originado la mezcla, mientras que la consideración contraria, es decir, que una muestra alineada con otras dos que hacen de extremo, sea el producto de su mezcla, no tiene por qué ser cierta. La muestra en posición intermedia, que aparenta ser consecuencia de la mezcla, puede tener una mineralización superior o inferior al producto de la mezcla, y, por tanto, no provenir de ella. Recordar que en el diagrama de Piper se representan proporciones y no concentraciones. En la [Figura 5](#) se ilustran las ideas expuestas en este apartado mediante el ejemplo que propuso Piper en su artículo original ([Piper, 1944](#)), para ello:

- V_a = volumen en la mezcla M de agua que tiene la composición A
- V_b = volumen en la mezcla M de agua que tiene la composición B
- E_a = concentración de la muestra A expresada en miliequivalentes
- E_b = concentración en la muestra B expresada en miliequivalentes
- E_m = concentración en la muestra M expresada en miliequivalentes
- a = longitud del segmento entre A y M
- b = longitud del segmento entre B y M

se puede escribir:

$$a/b = (V_b \cdot E_b / V_a \cdot E_a) \quad (\text{eq. 1})$$

$$(V_a/V_b) = (b \cdot E_b / a \cdot E_a) \quad (\text{eq. 2})$$

$$E_m = [E_a \cdot E_b (a+b)] / (a \cdot E_b / a \cdot E_a) \quad (\text{eq. 3})$$

$$V_a = b \cdot E_b / (a \cdot E_a + b \cdot E_b) \quad (\text{eq. 4})$$

$$V_b = a \cdot E_a / (a \cdot E_a + b \cdot E_b) \quad (\text{eq. 5})$$

Con lo que, para cualquier ion, siendo C_a su concentración en el agua A, C_b su concentración en el agua B, y C_m la concentración en la mezcla, resulta que:

$$C_m = (C_a \cdot V_a) + (C_b \cdot V_b) \quad (\text{eq. 6})$$

Siempre que un agua es consecuencia de la mezcla de otras dos, y no se han producido procesos de precipitación, se encontrará que existe concordancia entre las concentraciones medidas y los resultados de la ecuación (3) para la concentración total, y con las ecuaciones (4) y (5) para las concentraciones individuales del ion en estudio.

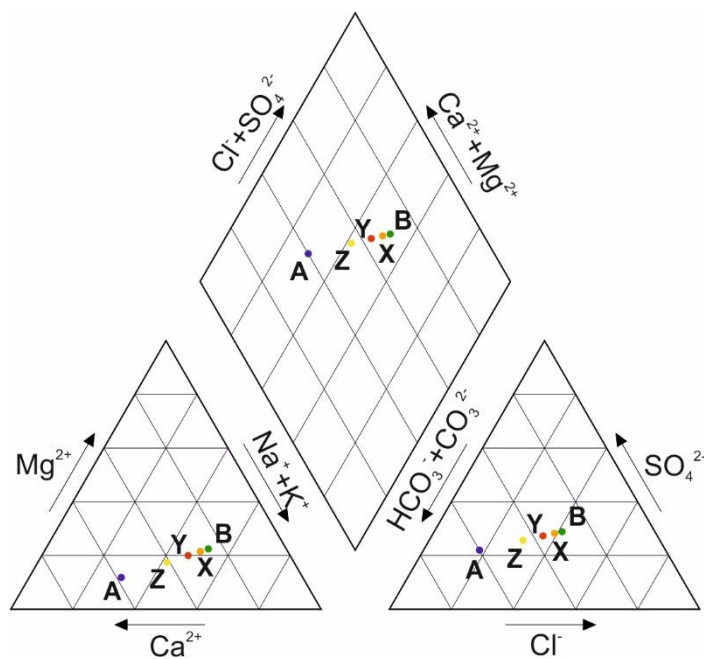


Figura 5. Ejemplo de mezcla de agua. Las muestras A y B son los términos extremos sin mezcla y las muestras X, Y y Z corresponden a una mezcla del 25% de A +75% de B, 50% de A +50% de B y 75% de A +25% de B, respectivamente.

Tabla 1 Composición de las muestras representadas en la [Figura 5](#).

Composición catiónica mg l ⁻¹						Composición aniónica mg l ⁻¹					
	A	B	X	Y	Z		A	B	X	Y	Z
Ca ²⁺	57	81	75	69	63	SO ₄ ²⁻	68	210	169.5	129	88.5
Mg ²⁺	7	44	34.75	25.5	16.25	HCO ₃ ⁻	166	276	248.5	221	193.5
Na ⁺	32	187	148.25	109.5	70.75	CO ₃ ²⁻	0	0	0	0	0
K ⁺	2	12	9.5	7	4.5	Cl ⁻	29	221	173	125	77

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

Un problema inherente a la representación de datos en diagramas triangulares, en los que generalmente se expresa la información como porcentaje, y el diagrama de Piper no es más que una variante de los mismos, es que no proporcionan información acerca del contenido total de los iones representados, sino de las relaciones entre ellos. Con frecuencia esto se intenta paliar representando los análisis no como un punto, sino como un círculo cuyo diámetro es proporcional a la conductividad eléctrica, el residuo seco o el total de meqv l⁻¹ (Figura 6). Cuando la diferencia de mineralización es muy grande, puede representarse, en lugar de los valores citados, sus logaritmos.

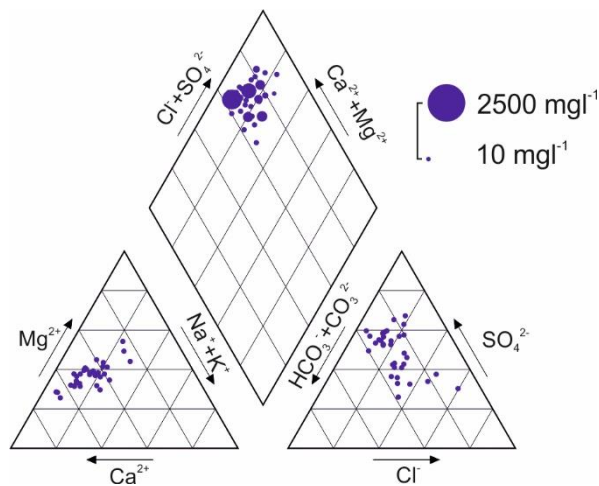


Figura 6. Ejemplo de Gráfico de Piper presentando la mineralización de las muestras en forma de círculos de diámetro proporcional a la misma. Esta forma de representación resulta muy práctica con pocos análisis, pero rápidamente se vuelve confusa cuando los círculos se superponen, y difícil de interpretar cuando se emplean gráficas logarítmicas.

En la actualidad, debido a la notable reducción del costo de los análisis de agua y la facilidad para almacenar y gestionar bases de datos haciéndolas fácilmente accesibles, es frecuente disponer de grandes conjuntos de análisis cuya representación en el diagrama de Piper no tiene mucha utilidad debido al problema de la superposición de puntos.

En la Figura 7a se muestra lo que sucede cuando se intenta representar un grupo de análisis muy numeroso en un único diagrama de Piper, evidenciando su principal limitación: cuando el número de puntos es elevado, estos se superponen y únicamente se observa una mancha de color. En el caso del ejemplo se ve claramente como predomina la facies bicarbonatada-cálcica, pero no se puede observar ninguna otra estructura en los datos representados. Además, resulta imposible tener una idea del número de análisis mostrado en cada una de las figuras.

Una forma de paliar el problema descrito en el apartado precedente es dibujar los puntos que representan los análisis con un cierto grado de transparencia (Figura 7b). De esta manera, en las zonas en las que hay mayor número de análisis se obtiene un color más oscuro, y en las que la densidad de puntos es menor un color más claro. Esta posibilidad no suele estar implementada en los paquetes de software para gráficos químicos (D-Piper sí lo hace) y puede ser una opción aceptable en algunos casos, cuando el número de análisis no es tan elevado como para llevar un elevado porcentaje de la superficie del diagrama al máximo de saturación.

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

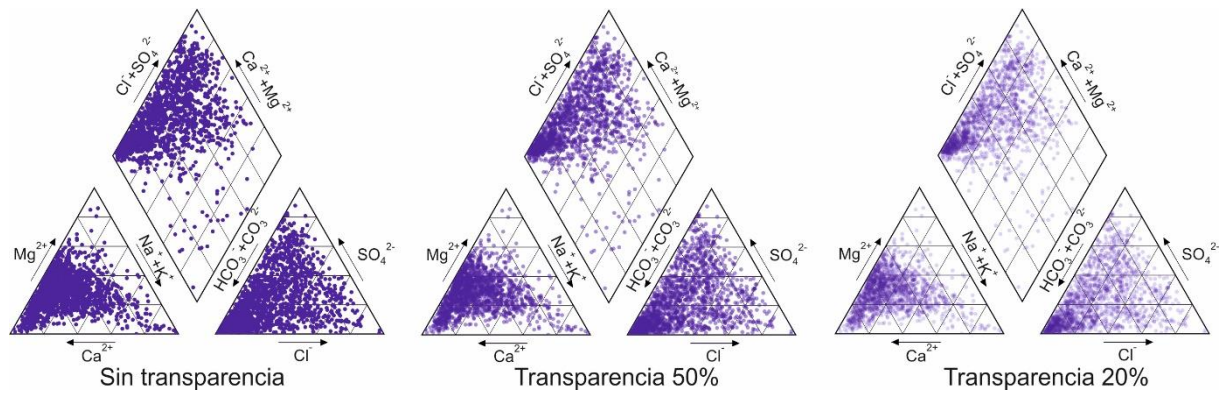


Figura 7. Ilustración de los problemas derivados de representar en el diagrama de Piper un número elevado de análisis, en este caso 1500. Generalmente más de 30 o 40 son suficientes para que aparezcan problemas de visualización.

Recientemente [Russionello \(2020\)](#) propusieron una nueva implementación del diagrama de Piper tradicional llamada PIED Piper. Esta implementación, programada en Matlab, emplea gráficos de calor para representar la distribución de puntos en el interior del Piper y permite agrupar los análisis mediante polígonos convexos. La herramienta propuesta presenta notables limitaciones al no basar la representación en la distribución estadística de los datos, lo que según [Conolly y Lake \(2006\)](#) es un paso esencial para lograr representaciones adecuadas y fiables, pues tal como afirman estos autores, podría obviarse información relevante. Dicho de otra manera, se puede afirmar que, para obtener una visualización óptima de la información, basada en gráficos de densidad, es necesario realizar previamente una selección de intervalos de clase basada en la distribución estadística de la información.

Por su parte, D-Piper es una modificación del diagrama original de Piper basada en representaciones de la distribución espacial de la densidad de puntos, que permite representar cualquier número de muestras manteniendo la mayor cantidad de información sobre la estructura de los datos originales. La solución consiste en representar la distribución de densidad de puntos en cada uno de los paneles que conforman el diagrama, pero empleando previamente técnicas estadísticas con objeto de establecer la distribución de clases óptima. Se ha elegido el lenguaje Python para implementar la aplicación pues existen numerosos intérpretes y entornos de programación de libre distribución y uso gratuito para el mismo.

¿CÓMO UTILIZAR D-PIPER?

Para utilizar el programa D-Piper hay que seguir cinco pasos:

1º Descargar el software de *Github* en forma de fichero comprimido (.zip) de la dirección <https://github.com/chesstor/D-Piper> y descomprimirlo en una carpeta que puede tener cualquier nombre aceptado por el sistema operativo que se esté empleando. Este paso solo será necesario la primera vez que se haga un gráfico. Conviene descargar el programa de esta ubicación, pues es garantía de encontrar la última versión actualizada y libre de software malicioso.

El *script* `D_Piper_v1.py` se proporciona bajo licencia GPL-3.0. Para saber cuáles son las condiciones de uso, distribución o modificación consultar la página Web: <https://www.gnu.org/licenses/gpl-3.0.html>

2º Preparar el fichero de datos analíticos y ubicarlo en una carpeta llamada *Data*.

3º Modificar el fichero de opciones *D_Piper_v1_options.txt* según las necesidades del usuario. No olvidar introducir el nombre del fichero que contiene los datos a representar en la opción *File_Input*.

4º Abrir el *script* `D_Piper_v1.py` en un entorno de desarrollo de Python (en el desarrollo de esta versión se ha empleado Spyder (Python 3.7.6), bajo la distribución de Anaconda; Figura 8).

5º Ejecutar el *script* pulsando el icono “Run”.

En los siguientes apartados se desarrollan estos puntos con más detalle.

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

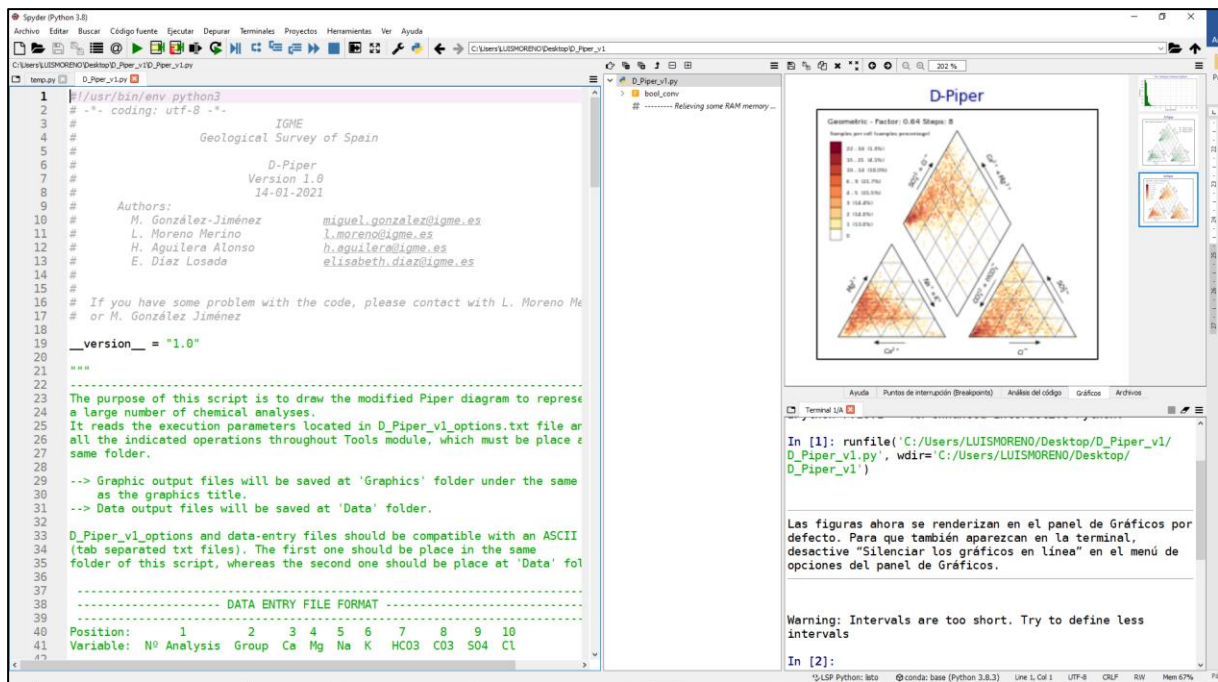


Figura 8. Aspecto del IDE Spyder con el *script* D-Piper abierto en su editor (izquierda) y gráfico de densidad obtenido (derecha) usando un fichero de datos de ejemplo.

Función de las diferentes carpetas y archivos

Una vez descomprimido el archivo .zip, se creará un sistema de carpetas y archivos, tal como se muestra en la Figura 9, que contiene tanto el código fuente como los ficheros de ejemplo y los gráficos generados. La carpeta `_pycache_` se creará automáticamente la primera vez que se ejecuta el *script*.

Nombre	Fecha de modificación	Tipo	Tamaño
<code>_pycache_</code>	27/01/2021 9:29	Carpeta de archivos	
<code>Data</code>	14/01/2021 9:29	Carpeta de archivos	
<code>Graphics</code>	27/01/2021 9:29	Carpeta de archivos	
<code>D_Piper_v1.py</code>	14/01/2021 12:54	Python File	11 KB
<code>D_Piper_v1_options.txt</code>	29/01/2021 13:13	Documento de te...	6 KB
<code>Tools.py</code>	27/01/2021 9:29	Python File	57 KB

Figura 9 Estructura de archivos que se crea al descomprimir el archivo .zip, es la recomendada para trabajar con el programa D-Piper.

Para un correcto funcionamiento del programa es muy importante respetar la estructura mostrada sin modificar el nombre de las carpetas `Graphics` o `Data`, ni de los archivos `D_Piper_v1.py` o `Tools.py`. A continuación se describe el contenido y función de cada una de estas carpetas y archivos:

D_Piper_v1.py

Script de Python que ha de ser ejecutado para construir los diagramas.

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

Tools.py

Script de Python que contiene un conjunto de funciones sobre las que se apoya el *software*. Actúa como un módulo, que es importado desde el archivo *D_Piper_v1.py*. Se encuentra ampliamente comentado, por lo que puede ser modificado en cualquier momento para adaptarlo a necesidades específicas.

__pycache__

La carpeta *__pycache__* se crea con la primera ejecución del código, y permite una ejecución más eficiente.

D_Piper_v1_options.txt

Archivo en formato ASCII separado por punto y coma (;) en el que se definen los valores de las diferentes opciones para construir los gráficos. Las opciones están agrupadas en seis conjuntos que se explican en detalle más adelante. Para evitar perder las opciones seleccionadas para un determinado gráfico, debe renombrarse el fichero o salvarse en otra carpeta, de forma que, cuando se necesite de nuevo, habrá que volver a cambiar su nombre al original o actualizar el nombre en el fichero *D_Piper_v1.py*

Data

En esta carpeta deben guardarse los ficheros con los datos analíticos. Estos archivos han de tener una extensión *.csv* o *.txt*, siendo recomendable la elección de esta última. Además, en esta carpeta se encuentran los cuatro ficheros de ejemplo que ilustran la publicación de [Moreno et al. \(2021\)](#). Estos fichero pueden ser eliminados si no van a ser utilizados.

Graphics

Carpeta en la que se almacenan los gráficos que se van generando.

Estructura del fichero de datos

Los datos necesarios para hacer el gráfico incluyen la identificación de cada análisis, el número de grupo al que pertenece, y los valores analíticos según el orden que se muestra en la Tabla 2. El fichero se guardará en formato ASCII separado por tabuladores (codificación UTF-8 sin BOM). No debe contener ningún tipo de cabecera. El símbolo de separador de decimales de los datos debe ser el punto (de manera independiente de cuál use el sistema operativo). En la figura 10 se muestra un ejemplo.

El fichero de opciones debe contener el nombre de este archivo con su extensión.

Tabla 2 Estructura del fichero de entrada de datos (valores expresados en mg l^{-1} o ppm)

1	2	3	4	5	6	7	8	9	10
Identificador	Grupo	Ca ²⁺	Mg ²⁺	Na ⁺	K ⁺	HCO ₃ ⁻	CO ₃ ²⁻	SO ₄ ²⁻	Cl ⁻

Identificador: número de identificación de muestra, debe ser diferente para cada una de ellas.

Grupo: cada muestra puede pertenecer a un grupo diferente. Se pueden definir hasta nueve grupos. Si se definen más de nueve grupos no aparecerá ningún mensaje de error, aunque puede haber problemas de superposición de puntos en el gráfico. Diferenciar grupos permite identificar con diferentes colores cada uno de ellos.

Ocho columnas con los resultados analíticos. El contenido iónico de cada muestra se puede expresar en mg l^{-1} (miligramos por litro) o en ppm (partes por millón).

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

Archivo	Edición	Formato	Ver	Ayuda					
1	1	57	22	0	0	128	0	0	3
2	1	48	17	1	0	240	0	0	1
3	1	2	1	0	0	7	0	1	1
4	1	18	4	12	0	75	0	1	10
5	1	20	2	13	0	98	0	1	7
6	1	77	8	27	0	239	0	1	6
7	1	65	3	1	1	188	10	1	5
8	1	57	5	1	1	203	0	1	14
9	1	73	5	1	1	237	0	1	11
10	1	43	16	2	1	221	0	1	4
11	1	44	7	5	1	149	0	1	10
12	1	48	7	5	1	142	13	1	9
13	1	47	10	5	1	174	3	1	6
14	1	55	23	6	1	247	9	1	12
15	1	61	23	6	1	265	18	1	7
16	1	68	14	8	1	198	33	1	13
17	1	64	15	8	1	210	25	1	15
18	1	45	20	8	1	182	0	1	27
19	1	60	13	9	1	219	15	1	16
20	1	101	31	9	1	438	25	1	8
21	1	86	2	10	1	215	13	1	17
22	1	56	3	13	1	161	0	1	23
23	1	40	13	13	1	158	12	1	22
24	1	59	12	15	1	211	8	1	28
25	1	63	20	15	1	274	8	1	21
26	1	101	3	16	1	181	0	1	52

Figura 10. Aspecto de un fichero de datos con la estructura descrita en la Tabla 2.

La primera columna (Identificador) y la segunda (Grupo) han de existir siempre, aunque no se consigne en ellas ningún valor.

Opciones disponibles, descripción del fichero *D_Piper_v1_options.txt*

Todas las opciones del programa, así como los nombres de los ficheros de entrada y salida, se especifican a través del fichero *D_Piper_v1_options.txt*. Es esencial respetar el formato de este fichero (ASCII separado por punto y coma “;”) para evitar errores en tiempo de ejecución. La mejor manera de modificar las opciones es introducirlas de una en una e ir comprobando los resultados y el correcto funcionamiento del *script*.

El fichero de opciones está dividido en seis grandes bloques:

Opciones generales

<p>Title_Piper;</p> <p>Mediante esta opción se define el Título del gráfico de Piper estándar. Aparecerá en la parte superior del mismo. Se recomienda poner títulos de menos de 10 caracteres para que quepan en el recuadro del gráfico.</p> <p>El nombre de los gráficos de salida se obtiene del título de cada uno de ellos. Por esta razón, los títulos de los gráficos han de ser distintos; de lo contrario, surgirían problemas de sobre-escritura.</p>	<pre> ----- GENERAL OPTIONS ----- Title_Piper; ;S-Piper; Title_DPiper; ;D-Piper; Bins_graph; ;50; Point_Size; ;2; Piper_D_Transparency; ;1; Piper_Transparency; ;0.1; Color_Unique; ;black; ColorMap; ;Y1OrRd; ; </pre>
---	--

<p>Title_DPiper;</p> <p>Mediante esta opción se define el Título del gráfico del diagrama de densidad.</p>	<pre> ----- GENERAL OPTIONS ----- Title Piper; ;S-Piper; Title_DPiper; ;D-Piper; Bins_graph; ;50; Point_Size; ;2; Piper_D_Transparency; ;1; Piper_Transparency; ;0.1; Color_Unique; ;black; ColorMap; ;Y10rRd; ; </pre>
<p>Bins_graph;</p> <p>Número de filas y columnas en las que se discretiza cada uno de los paneles (triángulo de cationes, triángulo de aniones y rombo) que componen el diagrama D-Piper. No existe una regla general para seleccionar el número de filas y columnas que permita una mejor visualización. Conviene experimentar y elegir el número de celdas que mejor represente el fenómeno que estamos analizando. Un número superior a 100 celdas puede ralentizar la ejecución del programa de forma apreciable.</p>	<pre> ----- GENERAL OPTIONS ----- Title Piper; ;S-Piper; Title_DPiper; ;D-Piper; Bins_graph; ;50; Point_Size; ;2; Piper_D_Transparency; ;1; Piper_Transparency; ;0.1; Color_Unique; ;black; ColorMap; ;Y10rRd; ; </pre>
<p>Point_Size;</p> <p>Esta opción permite especificar el tamaño de los puntos que representan los análisis en el gráfico. En general tamaños pequeños (entre 2 y 4) proporcionan una imagen más realista de la distribución de los análisis en el diagrama. La mejor opción se consigue cuando el punto se visualiza cómodamente sin que invada el resto de las áreas vacías.</p>	<pre> ----- GENERAL OPTIONS ----- Title Piper; ;S-Piper; Title_DPiper; ;D-Piper; Bins_graph; ;50; Point_Size; ;2; Piper_D_Transparency; ;1; Piper_Transparency; ;0.1; Color_Unique; ;black; ColorMap; ;Y10rRd; ; </pre>
<p>Piper_D_Transparency;</p> <p>Establece el grado de transparencia de las celdas en el diagrama de densidad (D-Piper). Oscila entre 0 y 1, recomendándose valores próximos a 1. Se ha de tener en cuenta que, cuanto menor sea este valor, más tenues serán los colores de las celdas, lo que puede dificultar la visualización e identificación de los intervalos.</p>	<pre> ----- GENERAL OPTIONS ----- Title Piper; ;S-Piper; Title_DPiper; ;D-Piper; Bins_graph; ;50; Point_Size; ;2; Piper_D_Transparency; ;1; Piper_Transparency; ;0.1; Color_Unique; ;black; ColorMap; ;Y10rRd; ; </pre>

<p>Piper_Transparency;</p> <p>Establece el Grado de transparencia de los puntos del gráfico de Piper. Valores entre 0.5 y 0.8 suelen dar buenos resultados, aunque es necesario experimentar pues dependiendo de la densidad de puntos puede ser conveniente un valor elevado, si hay pocos puntos, o bajo si se superpone un gran número de ellos.</p>	<pre> ----- GENERAL OPTIONS ----- Title_Piper; ;S-Piper; Title_DPiper; ;D-Piper; Bins_graph; ;50; Point_Size; ;2; Piper_D_Transparency; ;1; Piper_Transparency; ;0.1; Color_Unique; ;black; ColorMap; ;YlOrRd; ; </pre>
--	--

<p>Color_Unique;</p> <p>Permite especificar el color de los puntos que representan los análisis cuando no se diferencian grupos. Se puede indicar de dos maneras posibles:</p> <p>a) A través de la nomenclatura de la librería Matplotlib (https://matplotlib.org/stable/gallery/color/named_color_s.html). Ejemplo: <i>blue</i></p> <p>b) Según la nomenclatura HEX. Ejemplo: #BB2EB3 (figura 11)</p>	<pre> ----- GENERAL OPTIONS ----- Title_Piper; ;S-Piper; Title_DPiper; ;D-Piper; Bins_graph; ;50; Point_Size; ;2; Piper_D_Transparency; ;1; Piper_Transparency; ;0.1; Color_Unique; ;black; ColorMap; ;YlOrRd; ; </pre>
---	--

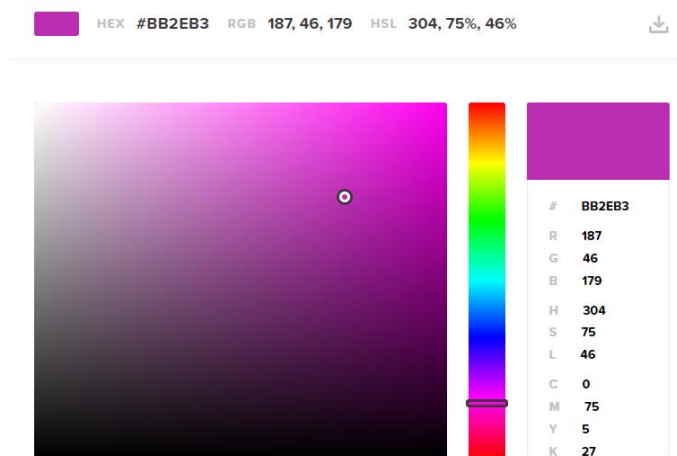


Figura 11. Ejemplo de selección de color mediante la nomenclatura HEX. <https://htmlcolorcodes.com/es/>

<p>ColorMap;</p> <p>Esta opción permite especificar la paleta de colores utilizada en el diagrama de densidad. Se puede elegir cualquiera de las paletas que proporciona la librería Matplotlib (figura 12) (https://matplotlib.org)</p> <p>Al estar conformado el diagrama de densidad por intervalos siempre crecientes, las paletas secuenciales son las que mejor resultados proporcionan. También se</p>	<pre> ----- GENERAL OPTIONS ----- Title_Piper; ;S-Piper; Title_DPiper; ;D-Piper; Bins_graph; ;50; Point_Size; ;2; Piper_D_Transparency; ;1; Piper_Transparency; ;0.1; Color_Unique; ;black; ColorMap; ;YlOrRd; ; </pre>
---	--

pueden elegir de tipo divergentes o categóricas. Cuando se elijan éstas últimas, el número de colores ha de ser mayor o igual al número de intervalos fijado.

Reverse_ColorMap;

Algunos mapas de color representan mejor los datos si se invierte el orden de los colores de la paleta original. Si este es el caso poner esta variable en (Y), si se quiere mantener el mapa original ponerla en (N).

```

---- GENERAL OPTIONS ----
Title_Piper;           ;S-Piper;
Title_DPiper;          ;D-Piper;
Bins_graph;           ;50;
Point_Size;           ;2;
Piper_D_Transparency; ;1;
Piper_Transparency;  ;0.1;
Color_Unique;         ;black;
ColorMap;             ;YlOrRd;
;
;
;
;
;
Reverse_ColorMap;     ;Y;
;
    
```

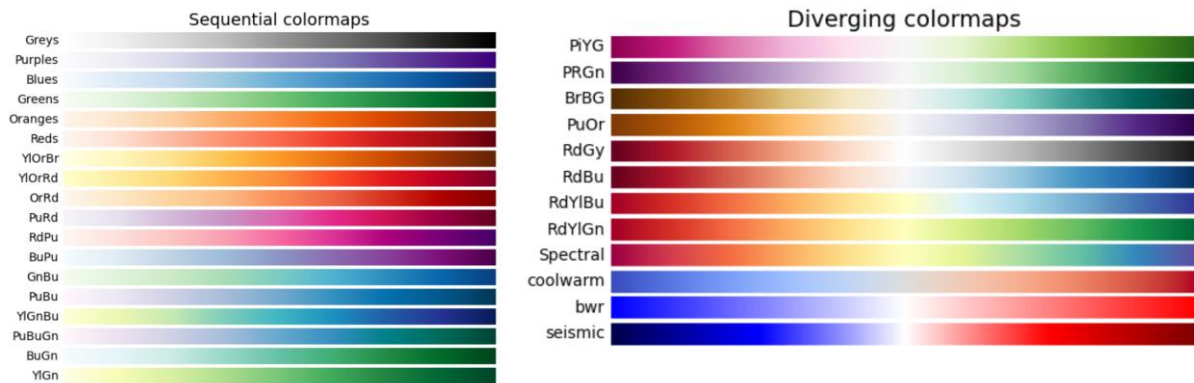


Figura 12. Ejemplo de paletas de tipo secuencial (izquierda) y divergentes (derecha) disponibles en la librería Matplotlib

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

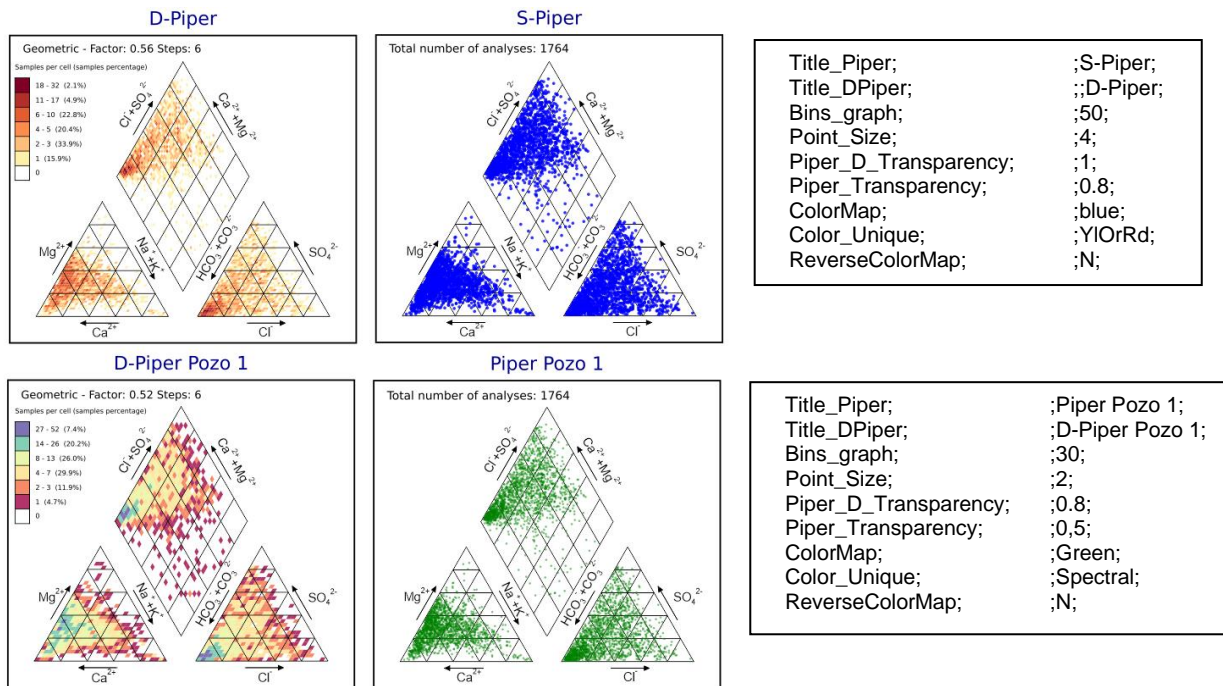


Figura 13 Ejemplo de aplicación de las Opciones Generales y su efecto sobre el aspecto de los gráficos D-Piper.

Opciones de grupo

Separate_Groups;

Esta opción permite distinguir con diferente color cada grupo de análisis (Y). Cuando no se quieran separar (N), el color de los puntos vendrá dado por la variable Color_Unique

```

---- GROUP OPTIONS ----
Separate_Groups; ;Y;
WhatGroupToDraw; ;1,2,3;
;
GroupsNames; ;G 1,G 2,G 3;
;
SetColors; ;blue,green,red;
Identify; ;N;
LabelColor; ;green;
Combine; ;N;
Combination_title; ;Com Diagram;
;
    
```

WhatGroupToDraw;

Mediante esta variable se especifican, separados por comas, los grupos que han de ser diferenciados. Se genera una advertencia cuando se incluye un grupo inexistente.

```

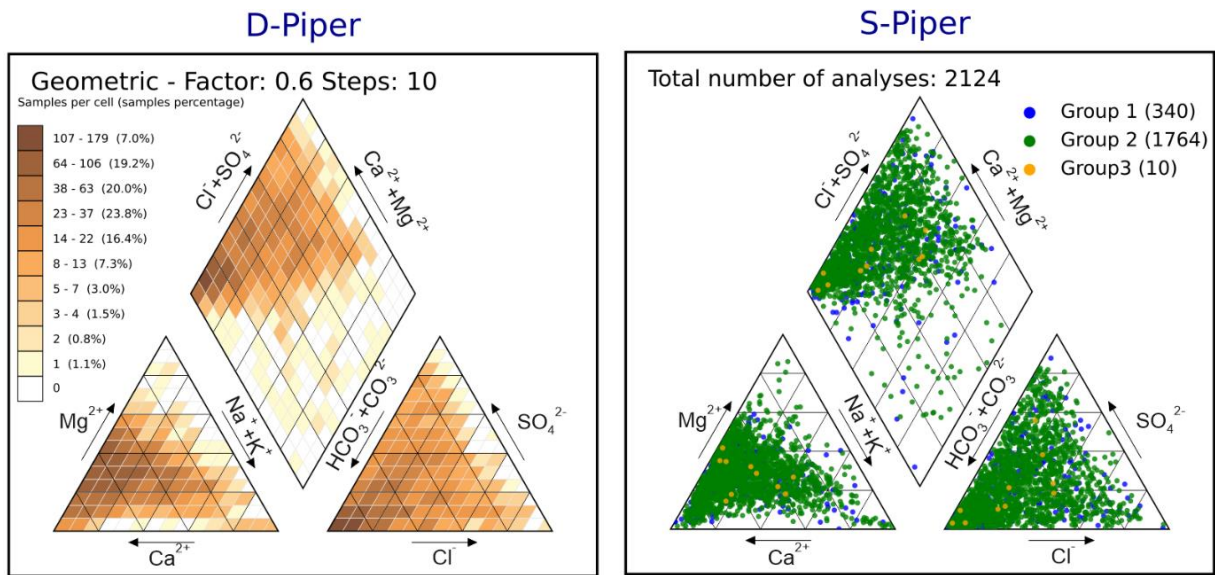
---- GROUP OPTIONS ----
Separate_Groups; ;Y;
WhatGroupToDraw; ;1,2,3;
;
GroupsNames; ;G 1,G 2,G 3;
;
SetColors; ;blue,green,red;
Identify; ;N;
LabelColor; ;green;
Combine; ;N;
Combination_title; ;Com Diagram;
;
    
```

<p>GroupsNames;</p> <p>En esta variable se indican los nombres de cada uno de los grupos que se han diferenciado en la variable anterior 'WhatGroupToDraw'. Se han de introducir tantos nombres como grupos de desee diferenciar. De lo contrario habrá grupos que se queden sin representar.</p>	<pre> ----- GROUP OPTIONS ----- Separate_Groups; ;Y; WhatGroupToDraw; ;1,2,3; ; GroupsNames; ;G 1,G 2,G 3; ; SetColors; ;blue,green,red; Identify; ;N; LabelColor; ;green; Combine; ;N; Combination_title; ;Com Diagram; ; </pre>
<p>SetColors;</p> <p>Mediante esta variable se asigna un color específico a cada uno de los grupos de datos diferenciados. Al igual que en la variable anterior, es importante introducir tantos colores como grupos se deseen diferenciar para evitar grupos con el mismo color. El color de cada uno de los grupos se indica tal y como se explica en la variable 'Color_Unique'.</p>	<pre> ----- GROUP OPTIONS ----- Separate_Groups; ;Y; WhatGroupToDraw; ;1,2,3; ; GroupsNames; ;G 1,G 2,G 3; ; SetColors; ;blue,green,red; Identify; ;N; LabelColor; ;green; Combine; ;N; Combination_title; ;Com Diagram; ; </pre>
<p>Identify;</p> <p>Cuando se activa esta opción (Y) se anota el número de identificación de los análisis en cada uno de los puntos del gráfico de Piper estándar.</p>	<pre> ----- GROUP OPTIONS ----- Separate_Groups; ;Y; WhatGroupToDraw; ;1,2,3; ; GroupsNames; ;G 1,G 2,G 3; ; SetColors; ;blue,green,red; Identify; ;N; LabelColor; ;green; Combine; ;N; Combination_title; ;Com Diagram; ; </pre>
<p>LabelColor;</p> <p>Esta opción permite definir el color de las etiquetas que identifican los análisis en el gráfico cuando la variable anterior (Identify) se encuentra activada (Y).</p>	<pre> ----- GROUP OPTIONS ----- Separate_Groups; ;Y; WhatGroupToDraw; ;1,2,3; ; GroupsNames; ;G 1,G 2,G 3; ; SetColors; ;blue,green,red; Identify; ;N; LabelColor; ;green; Combine; ;N; Combination_title; ;Com Diagram; ; </pre>

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

<p>Combine;</p> <p>Cuando está activada esta opción (Y) se genera un tercer gráfico resultado de la superposición del diagrama de Piper estándar sobre el de densidad. Si no se desea generar el tercer gráfico basta poner la opción (N).</p>	<pre> ----- GROUP OPTIONS ----- Separate_Groups; ;Y; WhatGroupToDraw; ;1,2,3; ; GroupsNames; ;G 1,G 2,G 3; ; SetColors; ;blue,green,red; Identify; ;N; LabelColor; ;green; Combine; ;N; Combination_title; ;Com Diagram; ; </pre>
---	---

<p>Combination_title;</p> <p>Título que se pone al diagrama combinado cuando la variable anterior (Combine) está activada (Y).</p>	<pre> ----- GROUP OPTIONS ----- Separate_Groups; ;Y; WhatGroupToDraw; ;1,2,3; ; GroupsNames; ;G 1,G 2,G 3; ; SetColors; ;blue,green,red; Identify; ;N; LabelColor; ;green; Combine; ;N; Combination_title; ;Com Diagram; ; </pre>
---	---



Separate_Groups;	;Y;
WhatGroupToDraw;	;1,2,3;
GroupsNames;	;Group 1,Group 2,Group 3;
SetColors;	;blue,green,orange;
Identify	;N;
LabelColor;	;black;
Combine;	;Y;
Combinations_title	;Combination Diagram;

Figura 14. Ejemplo de aplicación de las Opciones de Grupo y su efecto sobre el aspecto de los gráficos D-Piper.

Opciones de distribución de puntos y gráficas

<p>Transform_Log;</p> <p>Cuando se activa esta función, los métodos de representación se realizan a partir de los valores logarítmicos de las celdas (log-10) y no sobre sus valores originales.</p>	<pre> ---- POINT DISTRIBUTION OPTIONS AND Transform_Log; ;N; DivisionMethod; ;Geometric; Intervals; ;8; -- Manual Options: Manual_Steps_Def; ;1,89,150,600; ; -- Quantiles Options: InterpolationMet; ;linear; ; </pre>
<p>DivisionMethod;</p> <p>Esta opción permite seleccionar el método para separar los grupos de análisis. Las opciones disponibles son:</p> <ul style="list-style-type: none"> Manual Equidistant Quantiles Geometric Jenks DStandard 	<pre> ---- POINT DISTRIBUTION OPTIONS AND Transform Log; ;N; DivisionMethod; ;Geometric; Intervals; ;8; -- Manual Options: Manual_Steps_Def; ;1,89,150,600; ; -- Quantiles Options: InterpolationMet; ;linear; ; </pre>
<p>Intervals;</p> <p>Número de intervalos discretos en el que dividir el rango de valores. Este valor se ignora cuando el método de división de intervalos seleccionado es Manual, pues en este caso hay que especificar el inicio y fin de cada intervalo. Tenga en cuenta que los extremos de los intervalos siempre son cerrados ([]).</p>	<pre> ---- POINT DISTRIBUTION OPTIONS AND Transform_Log; ;N; DivisionMethod; ;Geometric; Intervals; ;8; -- Manual Options: Manual_Steps_Def; ;1,89,150,600; ; -- Quantiles Options: InterpolationMet; ;linear; ; </pre>
<p>Manual Options: Manual_Steps_Def;</p> <p>En este apartado se indican los límites de cada intervalo. El número de intervalos será igual al seleccionado en 'Manual_Steps_Def' menos uno. Los valores se introducen empleando como separador la coma (,).</p>	<pre> ---- POINT DISTRIBUTION OPTIONS AND Transform_Log; ;N; DivisionMethod; ;Geometric; Intervals; ;8; -- Manual Options: Manual_Steps_Def; ;1,89,150,600; ; -- Quantiles Options: InterpolationMet; ;linear; ; </pre>

<p>Quantiles Options: InterpolationMet;</p> <p>Esta variable permite especificar la forma en que se aplica la interpolación mediante cuantiles. Se debe especificar el método de interpolación a usar cuando el cuantil deseado se sitúa entre dos puntos $i < j$: $i * \text{fracción}$, donde fracción es la parte decimal del índice acotado por i y j. lower i higher j midpoint el punto medio $(i+j)/2$ nearest i o j, el que se encuentre más próximo</p> <p>En la dirección https://numpy.org puede encontrarse una explicación detallada de la función de NumPy empleada para interpolar en el método de cuantiles.</p>	<pre> ---- POINT DISTRIBUTION OPTIONS AND Transform_Log; ;N; DivisionMethod; ;Geometric; Intervals; ;8; -- Manual Options: Manual_Steps_Def; ;1,89,150,600; ; -- Quantiles Options: InterpolationMet; ;linear; ; </pre>
<p>Std D Options: DSfactor;</p> <p>Número de desviaciones estándar respecto a la media para construir el intervalo (p.e. 1, 0.5, 0.333, etc). La mitad se sitúan por encima del valor medio y la otra mitad por debajo. Surge una advertencia cuando los intervalos exceden los límites del rango de valores, estableciéndose menos intervalos de los indicados.</p>	<pre> -- Quantiles Options: InterpolationMet; ;linear; ; ; ; -- Standard Deviation Options: DSfactor; ;0.25; -- Graphic Option: Grid; ;N; </pre>
<p>Graphic Option: Grid;</p> <p>Esta variable permite mostrar una rejilla ajustada al número de celdas (Y) o no mostrarla (N).</p>	<pre> -- Quantiles Options: InterpolationMet; ;linear; ; ; ; -- Standard Deviation Options: DSfactor; ;0.25; -- Graphic Option: Grid; ;N; ; </pre>

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

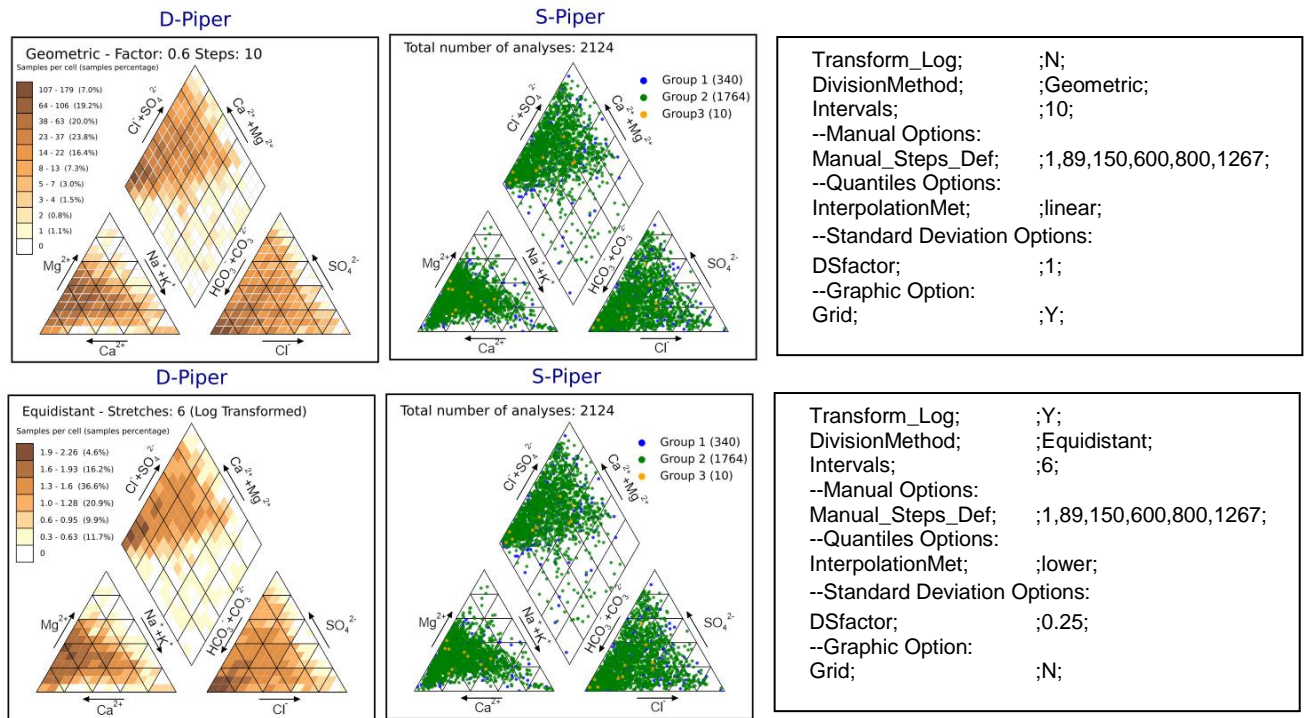


Figura 15. Ejemplo de aplicación de las Opciones de Puntos y Gráficas, y su efecto sobre el aspecto de los gráficos D-Piper.

Histograma de frecuencias

<p>ShowHisto;</p> <p>Esta variable permite dibujar el histograma de frecuencias (Y/N)</p>	<pre> ---- FREQUENCY HISTOGRAM ---- ShowHisto; ;Y; ColorHisto; ;green; BinsHisto; ;50; RangeHisto; ;0;100; LogHisto; ;N; ; </pre>
<p>ColorHisto;</p> <p>Mediante esta opción se establece el color de las barras del histograma de frecuencias.</p>	<pre> ---- FREQUENCY HISTOGRAM ---- ShowHisto; ;Y; ColorHisto; ;green; BinsHisto; ;50; RangeHisto; ;0;100; LogHisto; ;N; ; </pre>

<p>BinsHisto;</p> <p>Mediante esta opción se define el número de barras del histograma (debe ser un número entero), o selección automática (auto).</p>	<pre> ---- FREQUENCY HISTOGRAM ----- ShowHisto; ;Y; ColorHisto; ;green; BinsHisto; ;50; RangeHisto; ;0;100; LogHisto; ;N; ; </pre>
<p>RangeHisto;</p> <p>Esta variable permite indicar, separado por punto y coma (;), los valores mínimo y máximo del rango entre el cual se dibujará el histograma.</p>	<pre> ---- FREQUENCY HISTOGRAM ----- ShowHisto; ;Y; ColorHisto; ;green; BinsHisto; ;50; RangeHisto; ;0;100; LogHisto; ;N; ; </pre>
<p>LogHisto;</p> <p>Seleccionando esta variable se puede dibujar el eje Y del histograma de frecuencias en unidades logarítmicas (Y) o en unidades lineales (N)</p>	<pre> ---- FREQUENCY HISTOGRAM ----- ShowHisto; ;Y; ColorHisto; ;green; BinsHisto; ;50; RangeHisto; ;0;100; LogHisto; ;N; ; </pre>

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

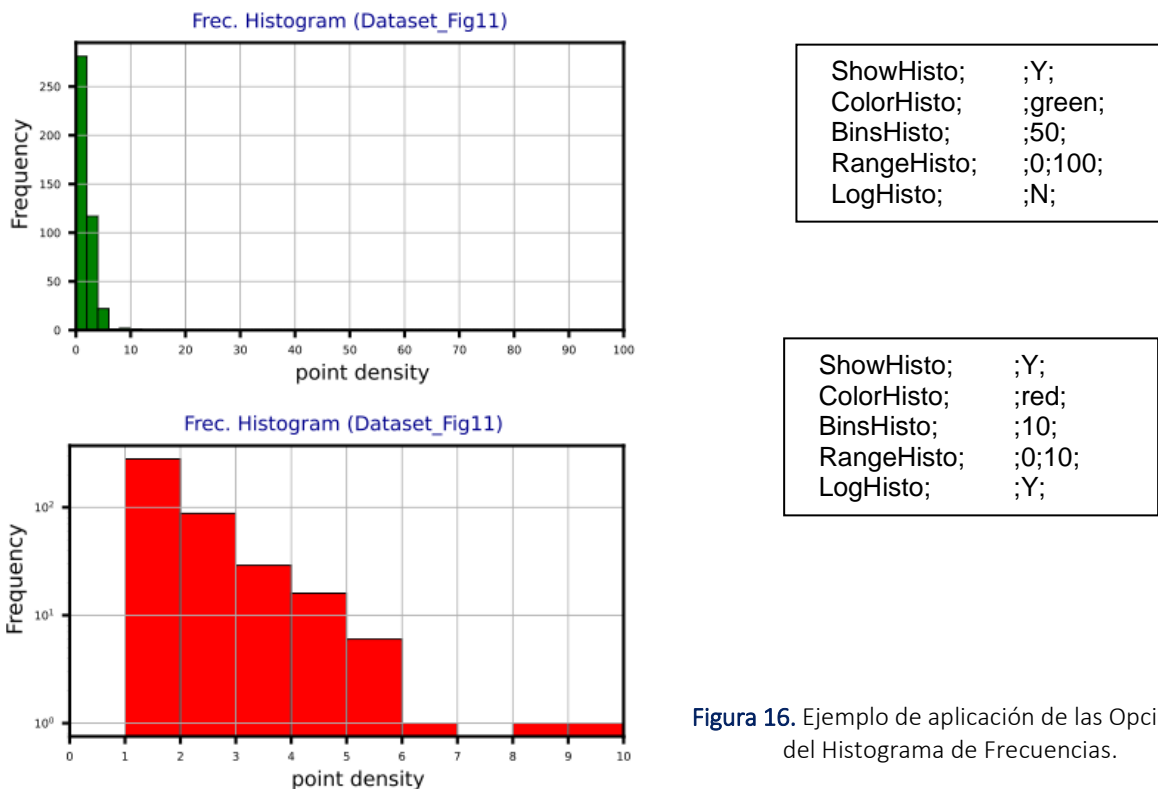


Figura 16. Ejemplo de aplicación de las Opciones del Histograma de Frecuencias.

Opciones de archivo

File_Input;

Nombre del fichero de entrada de datos. Seguir las convenciones del Sistema Operativo sobre el que está instalado el intérprete de Python utilizado.

```
----- FILE OPTIONS -----
File_Input; ;Fig2b&3.txt;
Save_Files; ;N;
;
```

Save_Files;

En caso de querer guardar en un fichero la tabla de información que se genera al ejecutar el programa, indicar la opción (Y), si no se desea guardar, indicar (N). Este archivo se almacenará en la carpeta *Data*.

El fichero creado tendrá el mismo nombre que el de datos precedido del texto "Info_Table_Dataset_" por ejemplo, si el fichero de datos se llama Data.txt el nuevo fichero se llamará Info_Table_Dataset_Data.txt.

Este nuevo fichero estará en formato ASCII separado por tabuladores y contendrá 10 columnas:

- ID_Analysis: N^o de identificación del análisis.
- Group: N^o de grupo
- CAT_sum Suma de cationes expresada en mg/l
- ANI_sum Suma de aniones expresada en mg/l

```
----- FILE OPTIONS -----
File Input; ;Fig2b&3.txt;
Save_Files; ;N;
;
```

Ca_epm	Contenido en Ca expresado en epm
Mg_epm	Contenido en Mg expresado en epm
NaK_epm	Suma de Na y K expresado en em
HCO3CO3_epm	Suma de HCO ₃ y CO ₃ expresado en epm
SO4_epm	Contenido en sulfato expresado en epm
CL_epm	Contenido en Cloruro expresado en epm

Opciones de archivos gráficos

<p>Store_Graph;</p> <p>En esta variable se especifica si se desea guardar los gráficos generados en el disco duro del ordenador indicando la opción (Y), y si no se quieren guardar se debe indicar (N). Los gráficos se guardarán en la carpeta <i>Graphics</i></p>	<pre>----- OUTPUT GRAPHIC FILE OPTIONS----- Store_Graph; ;Y; Extension_Graph; ;png; Graphic_Res; ;300;</pre>
---	--

ID_Analysis	Group	CAT_sum	ANI_sum	Ca_epm	Mg_epm	NaK_epm	HCO3CO3_epm	SO4_epm	CL_epm
57	2	5.233	4.806	47.678	40.875	11.447	78.423	8.656	12.921
58	2	6.85	5.737	64.832	21.62	13.547	71.989	15.234	12.777
59	2	0.375	0.334	13.333	21.867	64.8	19.76	12.575	67.665
60	2	9.013	7.941	47.065	21.003	31.932	30.613	36.708	32.679
61	2	6.179	6.241	16.961	2.67	80.369	64.076	7.002	28.922

Figura 17. Ejemplo de fichero de salida de resultados.

<p>Extension_Graph;</p> <p>La extensión del archivo, se indica al programa el formato de gráfico a guardar. Las opciones son: jpg, png, svg, pdf, ps</p>	<pre>----- OUTPUT GRAPHIC FILE OPTIONS----- Store Graph; ;Y; Extension_Graph; ;png; Graphic_Res; ;300;</pre>
---	--

<p>Graphic_Res;</p> <p>En esta variable se determina la resolución (en ppp, puntos por pulgada) a la que guardan los gráficos.</p>	<pre>----- OUTPUT GRAPHIC FILE OPTIONS----- Store_Graph; ;Y; Extension Graph; ;png; Graphic_Res; ;300;</pre>
---	--

INTRODUCCIÓN A LOS MÉTODOS DE REPRESENTACIÓN EMPLEADOS EN D-PIPER

D-Piper representa mediante un código de colores la densidad de puntos (análisis de agua) que caen en el dominio de cada una de las celdas en las que se discretizan los triángulos de cationes, de aniones y el rombo. Para obtener una representación visual adecuada se puede seleccionar el método de clasificación de las muestras en diferentes intervalos de clase. Es importante tener una cierta idea de la distribución de los datos para poder elegir el mejor método para definir los intervalos (Conolly and Lake 2006), para facilitar esta elección, se ha implementado la posibilidad de dibujar un histograma de distribución de densidades.

A continuación, se describen los seis métodos de clasificación implementados. En la aplicación se puede seleccionar tanto el método como el número de clases:

- a) *Intervalo definido por el usuario*: el usuario especifica directamente los puntos de corte para cada intervalo de clase.
- b) *Intervalos iguales*: este método es útil cuando el histograma tiene forma rectangular (la distribución es uniforme). El rango de datos se divide en el número indicado de clases (k) hasta un máximo de ocho. Los límites se calculan como sigue:

$$\begin{aligned}
 &L + (H - L)/k \\
 &L + 2(H - L)/k \\
 &\dots \\
 &L + (k - 1)(H - L)/k \quad (\text{eq. 7})
 \end{aligned}$$

donde H y L son los valores máximo y mínimo de los datos respectivamente.

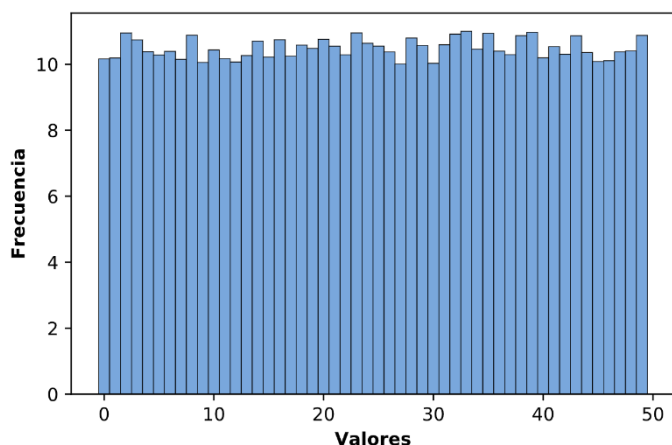


Figura 18. Ejemplo de Histograma de Frecuencias, en el que todas las clases tienen una frecuencia de aparición muy similar, que hace recomendable emplear el método de *Intervalos Regulares* para la definición de clases a representar en el Diagrama D-Piper.

- c) *Intervalo geométrico*: este método de clasificación se emplea para visualizar datos continuos cuya distribución en el histograma es asimétrica, está sesgada y muestran tasas de cambio muy pronunciadas (Conolly and Lake 2006). El programa resuelve la igualdad mostrada en la ecuación 8 para calcular la razón de la progresión geométrica.

$$Hr^k = L \quad (\text{eq. 8})$$

$$r = (L/H)^{1/k} \quad (\text{eq. 9})$$

donde H y L son los valores máximo y mínimo de los datos, respectivamente, r es la razón de la progresión geométrica, y k es el número de clases especificada por el usuario.

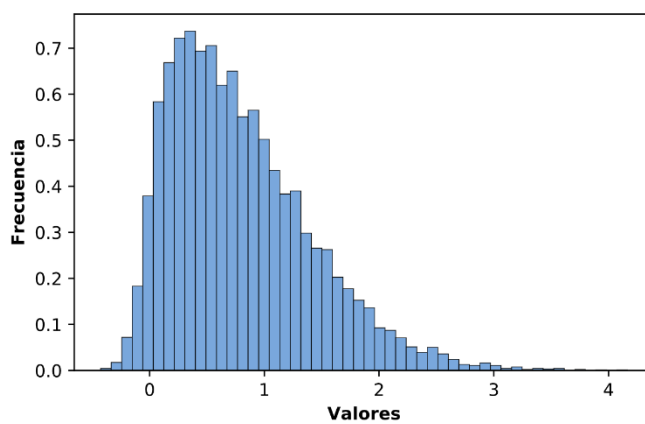


Figura 19. Ejemplo de Histograma de Frecuencias, en el que la distribución de clases muestra una notable asimetría, que hace recomendable emplear el método de *Intervalo Geométrico* para la definición de clases a representar en el Diagrama D-Piper.

Entonces, el límite de intervalo I_n se define según la expresión:

$$I_n = H * r^{n-1}, n = 1, \dots, k - 1 \quad (\text{eq. 10})$$

- d) *Cuantiles*: este método es adecuado cuando los datos están distribuidos de forma lineal. Produce intervalos irregulares que aseguran un número igual de celdas en cada clase y minimizan la importancia de los límites de clase (Dent 1999). Los valores de densidad de puntos se ordenan de forma ascendente, calculándose los puntos de corte de cada clase (K) dividiendo el número total de observaciones por el número de clases (cuantiles) especificado por el usuario. A partir del valor más bajo, K valores se incluyen en la primera clase, K en la siguiente clase, etc. La función para calcular cuantiles tiene un parámetro que especifica el método de interpolación que se usará en caso de que el límite del intervalo coincida con un cambio de valor.

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

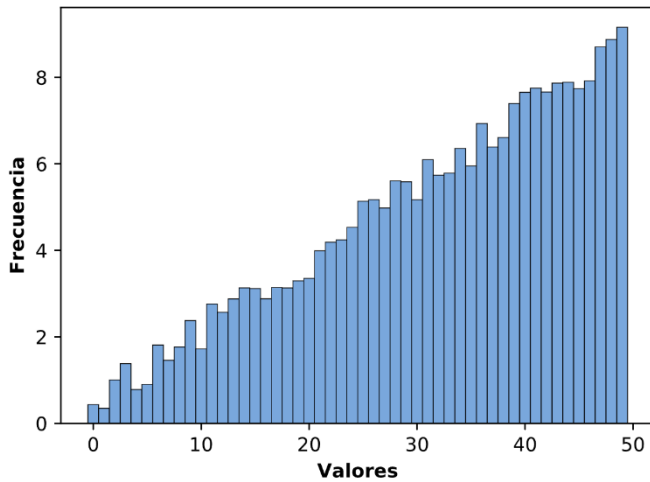


Figura 20. Ejemplo de Histograma de Frecuencias, en el que la distribución de clases muestra una clara distribución lineal que hace recomendable emplear el método de *Cuantiles* para la definición de clases a representar en el diagrama D-Piper.

- e) *Desviación estándar*: Este es el método de elección cuando los datos siguen una distribución normal o lognormal (Conolly and Lake 2006; Dent 1999). Este procedimiento muestra cuanto varía la densidad de puntos respecto a la densidad media. El algoritmo determina el valor medio de las observaciones y coloca los límites de clase, por encima y por debajo, cada $1/4$, $1/3$, $1/2$, 1, o el número de desviaciones estándar que se especifique, hasta que todos los valores de los datos estén contenidos dentro de las clases.

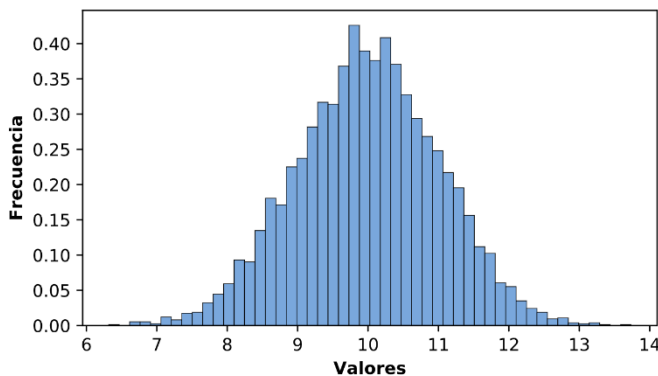


Figura 21. Ejemplo de Histograma de Frecuencias, en el que la distribución de clases muestra una distribución claramente normal, esto hace recomendable emplear el método de *Intervalo Desviación Standard* para la definición de clases a representar en el Diagrama D-Piper.

- f) *Método de cortes naturales de Jenks* (Jenks 1967): este algoritmo de agrupación de datos optimiza la disposición de un conjunto de valores en clases "naturales" tratando de minimizar la desviación respecto a la media en cada clase mientras maximiza la desviación respecto a las medias de los otros grupos. Se trata de un proceso iterativo que prueba repetidamente diferentes intervalos en el conjunto de datos hasta determinar la configuración óptima. El método es adecuado cuando los datos se distribuyen aleatoriamente o aparecen agrupados en torno a ciertos valores, existiendo cortes claros en el histograma (discontinuidades). La principal desventaja de este método es que es difícil de replicar, pues modificaciones pequeñas en los datos de entrada pueden cambiar el esquema de clasificación. Además, los datos anómalos (*outliers*) pueden recibir demasiada importancia porque generan clases únicas. Más detalles sobre el método pueden encontrarse en Dent (1999).

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

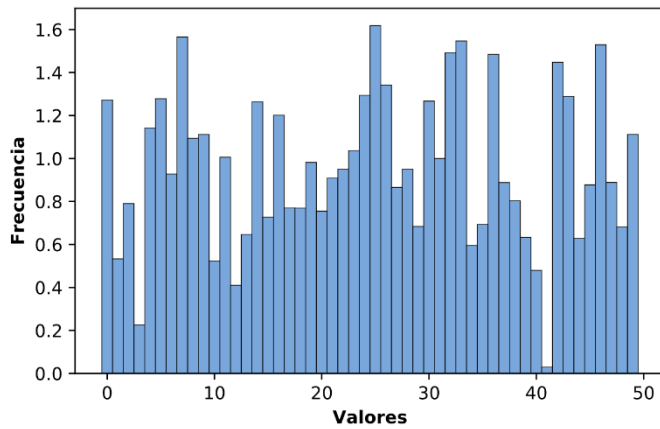


Figura 22. Ejemplo de Histograma de Frecuencias, en el que la distribución de clases muestra una estructura aleatoria y no se detecta ninguno de los patrones presentados en apartados anteriores. Este tipo de distribución hace recomendable emplear el método de *Intervalo Geométrico* para la definición de clases a representar en el Diagrama D-Piper.

Todos los métodos descritos pueden ser aplicados tanto sobre los datos sin procesar como sobre la transformación logarítmica de los mismos. Se aplica una transformación $\log_{10}(\text{densidad de puntos} + 1)$ para evitar la pérdida de información en celdas con una densidad de 1.

NOTAS SOBRE EL CÓDIGO

D-Piper (Moreno et al, 2021) ha sido implementado en lenguaje Python 3.7.6 empleando librerías standard: NumPy 1.17.4 (Harris et al., 2020), Pandas 1.1.3 (McKinney, 2010) y Matplotlib 3.2.2 (Hunter, 2007).

La librería *jenksy* 0.2.0 se ha empleado para calcular los intervalos cuando se utiliza el método de Jenks (Jenks, 1967) y puede ser descargada de la web <https://github.com/mthh/jenksy> o a través de alguna de las opciones que se indica en la Tabla 3.

Tabla 3. Opciones de instalación de la librería Jenksy según la plataforma de distribución utilizada y el espacio de ejecución.

Plataforma de distribución	Desde el entorno de desarrollo	Desde Window
Anaconda	<code>conda install jenksy</code>	<code>conda install -c conda-forge jenksy</code>
Winpy	<code>pip install jenksy</code>	<code>py -m pip install jenksy</code>

En el caso de querer compilar la librería a partir del código fuente será necesario disponer además de un compilador de código C.

El programa ha sido testeado con 32.732 análisis procedentes de la base de datos Aguas del CN Instituto Geológico y Minero de España (<http://info.igme.es/BDAGuas/>). Además, para comprobar la capacidad del código para representar conjuntos de datos muy grandes, se han generado muestras sintéticas, representando cada una de las 14 facies identificadas en la base de datos como combinación de las facies catiónica y aniónica individuales. La definición de las facies se ha basado en el criterio del 50%, según el cual se considera que un agua presenta una determinada facies cuando uno de los iones supera el 50% del total y tendrá facies mixta si ninguno de los iones supera dicha cifra. Por ejemplo, si el HCO_3^- supera el 50% para los aniones y el Ca^{2+} supera el 50% para los cationes, la facies resultante será bicarbonatada cálcica.

Las nuevas muestras para cada facies se han creado empleando el algoritmo SMOTE (*Synthetic Minority Over-sampling Technique*, Chawla et al., 2002). SMOTE genera nuevos datos para cada facies definida formando combinaciones lineales convexas entre muestras próximas de una misma facies, y muestreando aleatoriamente a lo largo de las líneas. Mediante este método se conservan las facies hidroquímicas originales en las nuevas muestras. El número de datos sintéticos para cada una de las facies es aproximadamente igual al del número de datos de la facies dominante (bicarbonatada cálcica) en el set original ($n=10.291$), lo que ha dado lugar a un conjunto extendido de 143.327 muestras con el que hacer las pruebas.

De forma resumida, el código ejecuta las siguientes operaciones:

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

1. Lee el fichero *D_Piper_v1_options.txt*, que contiene el valor de todas las variables y opciones necesarias, así como los nombres de los ficheros de entrada y salida, para generar los gráficos.
2. Lee el fichero de datos de la carpeta *Data*, que contiene los análisis químicos que se quieren representar. Debe estar en formato ASCII (con extensión *.txt* o *.csv*), y contener los análisis expresados en mg l^{-1} o ppm.
3. Transforma las unidades analíticas de mg l^{-1} a porcentaje de miliequivalentes.
4. Dibuja el diagrama de Piper estándar de todos los puntos o grupos elegidos. Se pueden seleccionar diversas opciones como, por ejemplo, el nivel de transparencia.
5. Según la selección del usuario se discretizan los paneles del diagrama de Piper, tanto los dos triángulos como el rombo se dividen siguiendo el mismo patrón.
6. Se cuenta el número de puntos que caen en cada una de las celdas definidas en el paso anterior.
7. Se establece un rango entre el valor de la celda más poblada y el valor de la celda menos poblada. (≥ 1). Este rango se divide en un número discreto de intervalos de acuerdo con el método de representación seleccionado, tal como se describe en apartados anteriores.
8. Se colorean los intervalos según la paleta de colores elegida. Además, se presenta en la leyenda el porcentaje de muestras que contiene cada uno de los intervalos representados.

Como se puede observar, en el paso 7 se excluyen del análisis las celdas que no contienen ningún punto en su interior (celdas con un valor igual a cero), las cuales se representan en blanco en el diagrama D-Piper. Esto se hace por dos razones: a) las celdas sin dato son muy abundantes en la mayoría de los diagramas y su inclusión en el cálculo de los intervalos a representar puede fácilmente ocultar los patrones de densidad en el rango inferior de densidades; b) las celdas en blanco facilitan la visualización e interpretación de las áreas con baja densidad.

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

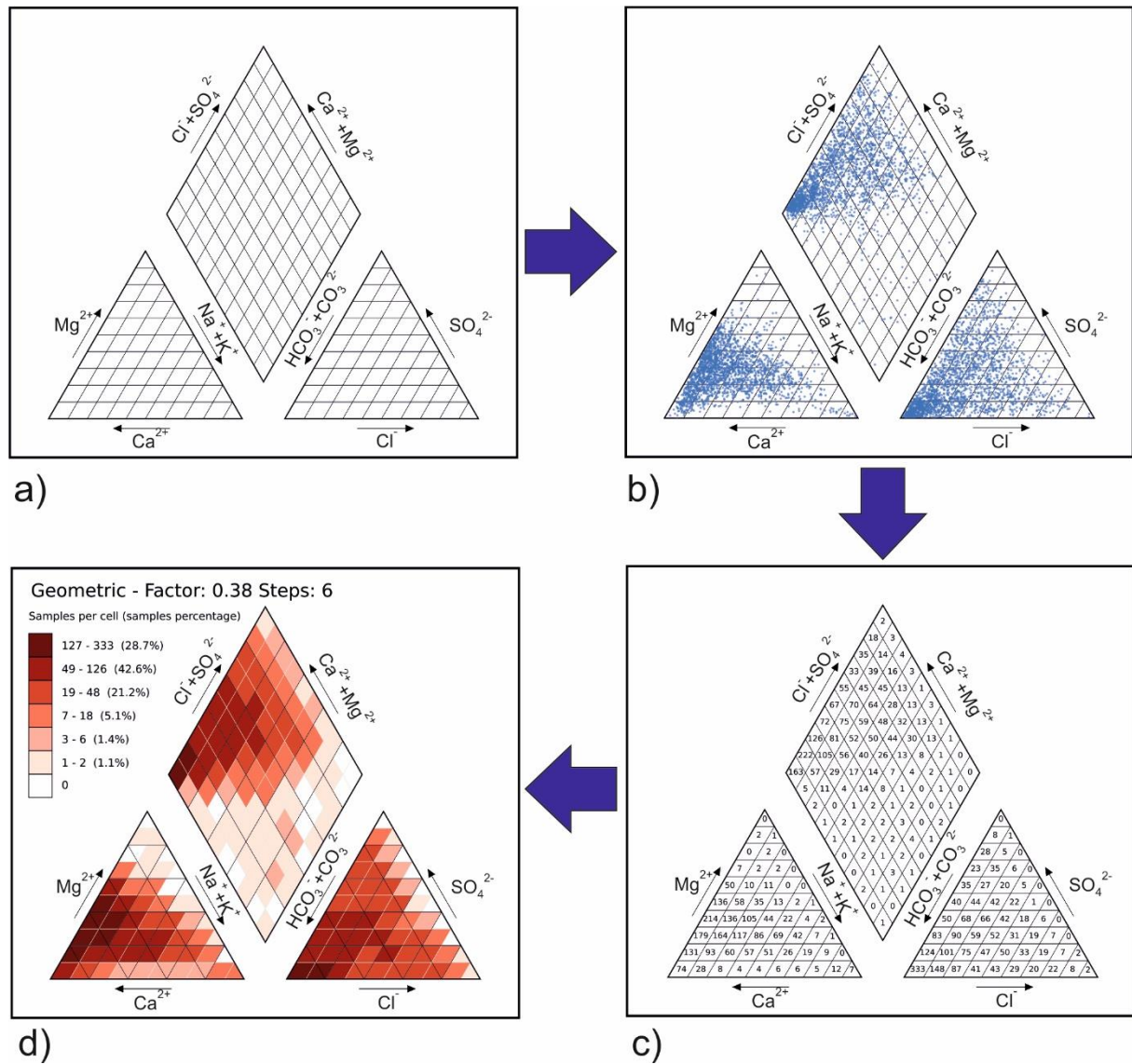


Figura 23 Esquema del procedimiento de construcción del gráfico de D-Piper. **a)** resultado de proceso de discretización de los triángulos y del rombo. **b)** presentación de los puntos sobre el rombo como en un diagrama de Piper clásico. **c)** estimación de los puntos caen en cada uno de las celdas. **d)** Una vez seleccionado el método de división en clases, asignación de un color a cada una de ellas y coloreado de las celdas.

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

Esta aplicación ha sido desarrollada en el marco del proyecto CLIGRO, que cuenta con la financiación del Plan Estatal de Investigación Científica y Técnica (CGL2016-77473-C3-1-R) y del proyecto IGME HDROCAMBIO Ref 2616. Asimismo, es parte de las actividades subvencionadas dentro del Sistema Nacional de Garantía Juvenil (PEJ2018-002477), financiado por la Fundación Tripartita para la Formación en el Empleo, la iniciativa YEI (Youth Employment Initiative) y el Fondo Social Europeo (FSE).

REFERENCIAS

- Burdon, D.J. and Mazloum, S. 1958. Some chemical types of groundwater from Syria. Proceedings of the UNESCO Symposium, Tehran UNESCO, Paris, 73-90.
- Chadha, D.K. 1999. A proposed new diagram for geochemical classification of natural waters and interpretation of chemical data. *Hydrogeology Journal*, 7, 431-439.
- Chawla, N.V, Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Collins, W.D. 1923. Graphic representation of water analyses. *Journal of Industrial and Engineering Chemistry*, 15, 4, 394.
- Conolly, J. and Lake, M. 2006. *Geographical Information Systems in Archaeology* Cambridge: Cambridge University Press. Cambridge 360 pp
- D'Amore, F., Scandiffio, G. and Panichi, C. 1983. Some observations on the chemical classification of ground waters. *Geothermics*, 12, 141-148.
- Dent, B.D. 1999. *Cartography. Thematic Map Design*. Fifth Edition. McGraw-Hill Science, London.
- Durov, S.A. 1948. Natural waters and graphic representation of their composition. *Doklady Akademii Nauk SSSR*, 59, 87-90.
- Harris, C.R., Millman, K.J., van der Walt, S.J. et al. 2020. Array programming with NumPy. *Nature*, 585, 357–362.
- Hill, R.A. 1940. Geochemical patterns in the Coachella Valley, California. *Transactions of the American Geophysical Union*, 21, 46-49.
- Hunter, J.D. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90-95.
- Jenks, G.F. 1967. The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography*, 7, 186-190.
- Langelier, W.F. and Ludwig, H.F. 1942. Graphical methods for indicating the Mineral character of Natural Waters. *American Water Works Association Journal*, 34, 335-352.
- Lloyd, J.W. 1965. The hydrochemistry of the aquifers of Northern Jordan. *Journal of Hydrology*, 3, 319-330.
- McKinney, W. 2010. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman (eds.), *Proceedings of the 9th Python in Science Conference*. Austin Texas. 51–56.
- Moreno Merino, L., De la Losa Román, A., Romero Prados, A., Carro Martín, S. 2006. Interpretación de análisis químicos de aguas subterráneas mediante la hoja INAQUAS. Instituto Geológico y Minero de España, Madrid, 82 pp.

D-Piper: Diagrama de Piper de densidad. Representación de grandes conjuntos de datos mediante el diagrama de Piper.

Moreno Merino, L.; Aguilera, H.; González Jiménez, M.; Díaz Losada, E. (2021) D-Piper, a modified Piper diagram to represent big sets of hydrochemical analyses. *Environmental Modelling and Software*, 138.

Piper, A.M. 1944. A graphic procedure in the geochemical interpretation of water-analyses. *EOS Transactions American Geophysical Union*, 25(6), 914–928.

Ray, R.K. and Mukherjee, R. 2008. Reproducing the piper trilinear diagram in rectangular coordinates. *Groundwater* 46(6), 893-896.

Russoniello, C.J. and Lautz, L.K. 2020. Pay the PIED Piper: Guidelines to visualize large geochemical datasets on Piper Diagrams. *Groundwater*, 58(3), 464-469.

Schoeller, H. 1964. La classification géochimique des eaux. I.A. S.H. Publication no. 64, Gen. Assembly of Berkeley 1963. 4, 16-24.

Shelton, J.L., Engle, M.A., Buccianti, A. and Blondes, M.S. 2018. The isometric log-ratio (ilr)-ion plot: A proposed alternative to Piper diagram. *Journal of Geochemical Exploration*, 190, 130-141.

Stiff, H.A. Jr. 1951. The interpretation of chemical water analysis by means of patterns. *Journal of Petrology Technology*, 3(10), 15-17.

Teng, W.C., Fong, K.L., Shenkar, D., Wilson, J.A. and Foo, D.C.Y. 2016. Piper diagram-a novel visualisation tool for process design. *Chemical Engineering Research and Design*, 112, 132-145.