Accepted Manuscript

The isometric log-ratio (ilr)-ion plot: A proposed alternative to the Piper diagram

Jenna L. Shelton, Mark A. Engle, Antonella Buccianti, Madalyn S. Blondes

PII: S0375-6742(17)30651-9
DOI: doi:10.1016/j.gexplo.2018.03.003
Reference: GEXPLO 6116

To appear in: *Journal of Geochemical Exploration*

Received date: 22 September 2017
Revised date: 1 February 2018
Accepted date: 8 March 2018

Please cite this article as: Jenna L. Shelton, Mark A. Engle, Antonella Buccianti, Madalyn S. Blondes , The isometric log-ratio (ilr)-ion plot: A proposed alternative to the Piper diagram. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Gexplo(2017), doi:10.1016/j.gexplo.2018.03.003

The isometric log-ratio (ilr)-ion plot: A proposed alternative to the Piper diagram

Jenna L. Shelton[a,*] (jlshelton@usgs.gov), Mark A. Engle[a,b] (engle@usgs.gov), Antonella Buccianti[c,d] (Antonella.buccianti@unifi.it), Madalyn S. Blondes[a] (mblondes@usgs.gov)

[a]Eastern Energy Resources Science Center, U.S. Geological Survey, Reston VA 20192 USA

[b]Department of Geological Sciences, University of Texas at El Paso, El Paso, Texas 79968 USA

[c]Department of Earth Sciences, University of Florence, Via G. La Pira 4, 50121 Firenze, Italy

[d]CNR-IGG (Institute of Geosciences and Earth Resources), Via G. La Pira 4, 50121 Firenze, Italy

*Corresponding Author: jlshelton@usgs.gov; Ph. 1 (703) 648-6489;

Abstract

The Piper diagram has been a staple for the analysis of water chemistry data since its introduction in 1944. It was conceived to be a method for water classification, determination of potential water mixing between end-members, and to aid in the identification of chemical reactions controlling a sample set. This study uses the information gleaned over the years since the release of the Piper diagram and proposes an alternative to it, capturing the strengths of the original diagram while adding new ideas to increase its robustness. The new method uses compositional data analysis to create 4 isometric log-ratio coordinates for the 6 major chemical species analyzed in the Piper diagram and transforms the data to a 4-field bi-plot, the ilr-ion plot. This ilr-ion plot conveys all of the information in the Piper diagram (water mixing, water types, and chemical reactions) while also visualizing additional data, the ability to examine Ca^{2+}/Mg^{2+} versus Cl^-/SO_4^{2-}. The Piper and the ilr-ion plot were also compared using multiple synthetic and real datasets in order to illustrate the caveats and the advantages of using either diagram to analyze water chemistry data. Although there are challenges with using the ilr-ion plot (e.g., missing or zero values zeros in the dataset must be imputed by positive real numbers), it appears that the use of

compositional data analysis coupled with the ilr-ion plot provides a more in-depth and complete analysis of water quality data compared to the original Piper diagram.

Keywords: Compositional Data Analysis; Piper diagram; isometric log-ratio transformation; Dockum Aquifer

1. Introduction

The graphical analysis of surface water and groundwater chemistry data can provide useful information about the hydrology of a system, such as sources of salinity, microbial processes occurring, contaminant transport and fate, mixing between different waters, and groundwater flow or upwelling into river systems. Quick, graphical interpretation of water quality data has traditionally included scatterplots of concentration or ratio of ions (e.g., Br^- vs. Cl^- plots, Cl^-/Br^- vs. Cl^- plots), or multi-element plots, such as the Piper diagram (Piper, 1944) or Stiff plot (Stiff, 1951). The Piper diagram, originally proposed nearly 75 years ago (Piper, 1944) is one of the most popular and useful plots for initial interpretation of water chemistry. It is still used to classify samples based on chemical type (e.g., Jeong, 2001), examine the hydrochemical evolution of natural waters (e.g., Cloutier et al., 2008), and to describe the chemical composition of water in a particular hydrologic setting (e.g., Shekhar and Sarkar, 2013). Despite its common use as a graphical water quality tool, the Piper diagram is prone to a few problems, such as difficulty distinguishing between certain water types and data clustering obscuring interpretations (as discussed in length in section 2). This work suggests an alternative to the Piper diagram, which overcomes some of its issues while retaining its utility and strengths through incorporation of methods of compositional data analysis (CoDA).

This paper is organized as follows: Section 2 provides the basics of the Piper diagram and identifies some potential shortcomings of the technique. Section 3 presents an alternative approach, using CoDA, and compares it to the original Piper diagram. Section 4 presents a comparative case study using

groundwater chemistry data from a portion of the Dockum Aquifer in Texas, USA. Section 5 concludes on key differences between the techniques and makes suggestions about applications of graphical methods in the interpretation of water chemistry data.

2. Creation, applications, and shortcomings of Piper diagrams

The ubiquitous use of the Piper diagram in graphical interpretation of water chemical data is grounded in its elegant simplicity. The diagram consists of two trilinear diagrams, which are linked via a central diamond-shaped region (Fig. 1A). The trilinear diagrams plot relative concentrations of cations (Ca^{2+}, Mg^{2+}, $Na^{+} + K^{+}$) and anions (Cl^{-}, SO_4^{2-}, $HCO_3^{-} + CO_3^{2-}$) commonly identified in natural waters. By convention, the data are converted to units of equivalents per liter then normalized to 100% before plotting. The diagrams are aligned in such a way that the ratio of alkaline earths ($Ca^{2+} + Mg^{2+}$) to alkalis ($Na^{+} + K^{+}$) and the ratio of weak acids ($HCO_3^{-} + CO_3^{2-}$) to strong acids ($Cl^{-} + SO_4^{2-}$) are projected into the central diamond-shaped field.

Piper plots are powerful because they are used for both classification of water quality data and for identification of processes impacting the data, such as end-member mixing of waters, ion exchange, and mineral precipitation and dissolution. In terms of water classification, arguably 9 different types of water can be distinguished in a Piper plot. These types include water dominated by alkaline earths or alkalis (types 1 and 2), strong or weak acids (types 3 and 4), primary or secondary alkalinity (types 5 and 8, respectively), primary or secondary salinity (types 7 and 6, respectively), and water without a single dominating (>50%) cation-anion pair (type 9) (inset to Fig. 1; Piper, 1944).

Beyond broad hydrochemical classification, mixtures between end-member compositions can also be identified using the diamond-shaped field (Piper, 1944). Namely, assuming all species originally identified in the two mixing waters (i.e., end-members) remain in solution while mixing, if a mixture of two end-members would occur, waters sourced from varying fractions of those two end-members

would plot along a straight line between them in the diamond-shaped field. The exact location of each mixed water sample on the line depends on things such as concentration of mixing species in each endmember, fraction of each endmember in the water mixture, and fractions of water mixing (Piper, 1944). Fig. 1A visualizes the original data used by Piper (1944) along with two hypothetical waters (data used found in Table 1), one exhibiting high relative abundance of Na^+ and Cl^- and one exhibiting a low relative abundance of Na^+ and Cl^-, along with a collection of waters composed of different fractions of those two end-members. As expected, the waters composed of mixtures plot along a line between the corresponding end-members on both trilinear diagrams and in the diamond-shaped field. Three-component mixing can also be predicted or identified, by plotting vectors between three samples in the diamond shaped field; the mixture of those three water sources, assuming again that all species remain in solution, would plot within the triangle formed by the vectors drawn between those three points (Piper, 1944; not shown in Fig. 1). Identifying mixtures via the Piper diagram does come with a caveat, addressed by Piper (1944): the Piper diagram alone should not be used for predicting or determining water mixtures, and a mixing model should always be calculated to verify the conclusions suggested by the Piper plot itself.

Similarly, mineral dissolution can be modeled as a two-component mixing model, with the dissolving mineral essentially serving as water with an infinite source of that chemical constituent. For example, if a water was dissolving $CaCl_2$ along its flow path, the resulting evolving waters would be graphically equivalent to those from a mixing scenario with an endmember composed of infinite concentrations of Ca^{2+} and 2 x Cl^-. As a water dissolves a mineral phase, the evolving water at different stages of evolution plots along a straight line on the Piper diagram towards the chemical composition of that pure mineral (100% Ca^{2+} and 100% Cl^- for the example above). Ion exchange is also represented likewise: an exchanging water sample would progress along a straight line away from its original composition, consisting of the ion being exchanged out of solution, towards a hypothetical water,

consisting of the ion being exchanged into solution (see Fig. 6 in Piper (1944) for a visual description of these processes on a Piper plot).

Although there are many obvious benefits of exploring the chemical composition of a group of water samples with the Piper diagram - and it has remained remarkably relevant for nearly 75 years - issues arise in certain circumstances. Despite being sometimes used to discriminate various water types, the Piper diagram makes it difficult to distinguish between waters that are composed of combinations of the ions Mg^{2+}, SO_4^{2-}, Ca^{2+}, and Cl^-. This is because the central diamond-shaped field, the portion of the Piper diagram that compares both cations and anions, cannot differentiate between Mg^{2+} and Ca^{2+} or between SO_4^{2-} and Cl^-. In small datasets or where overlap is minimal, this can be overcome by projecting individual points from central diamond back to the trilinear diagrams, where the relative abundance of cations and anions is clear. However, for larger datasets or where data overlap is more significant, it can become very difficult to follow any individual sample through the plot and determine the relative abundance of Mg^{2+} vs. Ca^{2+} and SO_4^{2-} vs. Cl^-. As an example, Fig. 1B provides a hypothetical dataset where samples are either rich in Mg-SO_4 or Ca-Cl, with two “end-members” (samples with the highest concentrations of Ca^{2+} and Cl^- and Mg^{2+} and SO_4^{2-}, respectively) represented as black triangles. It is apparent that one sample is concentrated in Ca^{2+} and one in Mg^{2+}, but it is difficult to determine if the 2 samples are composed of Ca-Cl and Mg-SO_4, or if they are composed instead of Ca-SO_4 and Mg-Cl, as these four ions are not directly compared to one another. Due to overlapping data in the diamond of the Piper diagram, if special symbols had not been used to identify these points, it would be very difficult to project the data down into the cation and anion triangles and therefore, to determine their composition. This could be overcome by applying an algorithm to automatically classify the data and use symbols to identify the results, but this adds another step in the data analysis process and in many cases specific symbols are used to denote other parameters.

Another set of issues arises when data are condensed in the corners and along the edges of the trilinear plots and in the diamond-shaped field (Fig. 1B). As discussed above, this is typically occurs when waters are concentrated in one particular cation and anion, such as Na-HCO_3 type waters. Data plotting along the boundaries are problematic for a few reasons, one being that it can hide the importance of less abundant ions on the overall interpretation of the samples being assessed. For example, the data presented on Fig. 1B are hypothetical waters that are concentrated in either Mg-SO_4 or Ca-Cl, as discussed above. From the trilinear plots, it is apparent that there are minor contributions from the less dominant cation and anion species, $Na^+ + K^+$ and $HCO_3^- + CO_3^{2-}$, respectively. However, when these points are projected into the diamond shaped field, all of the data points condense in the upper corner of the diamond, making it difficult to determine which data points are dominated by which cations and anions, along with the contributions from the minor cation and anion pairs (Fig. 1B). Without the color coding for Mg-SO_4 type waters versus Ca-Cl type waters, it would be impossible to determine which samples are in the diamond-shaped field. When data are either combinations of the ions Mg^{2+}, SO_4^{2-}, Ca^{2+}, and Cl^-, or plot along the diagram boundaries, important information about the sample set can be lost when trying to make interpretations based solely on the Piper diagram. This is particularly true for contributions from minor ions or predicting the dominating species in a water when they are dominated by either SO_4^{2-} or Cl^-, or by Ca^{2+} or Mg^{2+}. The Piper diagram could be scaled (i.e., the trilinear diagram could span from 80% – 100% instead of 0% – 100%) and may eliminate this problem in some cases, but the Piper is not traditionally modified as such and in some cases the data cover enough of the plot that this is not possible.

To overcome these shortcomings of the Piper diagram, we propose a compositional data analysis (CoDA) approach to the traditional Piper diagram. Compositional data analysis, based upon the idea that concentrations of constituents are dependent upon one another and have their own geometry and special data analysis needs, has been increasingly used in the interpretation of water chemistry data

(Blake et al., 2016; Blondes et al., 2016; Buccianti and Pawlowsky-Glahn 2005; Buccianti and Zuo 2016; Engle and Blondes 2014; Engle and Rowan 2013; Otero et al., 2005).

3. A compositional data analysis alternative to the Piper diagram

Using a CoDA approach to transform the Piper diagram is one way to fully capture the usefulness of the diagram while also adding functionality and eliminating its few shortfalls. Water quality data are inherently compositional; the concentrations of all of the identified constituents are relative in nature and depend upon one another. That is, the addition or removal of any constituent in a water sample changes the concentration of every other constituent (e.g., evaporation of water changes the concentration of dissolved ions). The reason for this relationship is that although concentration data are typically thought of as spanning the range of positive real space, in fact a sample set of concentration data with *D* number of constituents is confined to a *D*-1 dimensional hyperplane in real space, called the simplex (Aitchison, 1986). Although the simplex is contained in real space, it has its own algebraic-geometric structure which differs from that of Euclidian geometry in real space (Aitchison, 1986; Pawlowsky-Glahn and Egozcue, 2001). Applications of traditional interpretive methods that rely upon Euclidian geometry in real space (e.g., summary statistics, correlation, etc.) to data on the simplex can provide misleading results (Buccianti 2013; Engle and Rowan, 2013). Previous researchers have shown that interpretation of raw concentration data in aqueous systems can result in incorrect interpretations of natural processes (e.g., Buccianti, 2013; Egozcue and Pawlowsky-Glahn, 2005; Engle and Blondes, 2014; Engle and Rowan, 2013). To avoid such potential problems, various CoDA methods are available to either work with the data in the simplex itself or to convert the data to orthonormal coordinates that are elements of real space (Mateau-Figueras et al., 2011). Here we propose the latter method, where data are converted using the isometric log-ratio (ilr) transformation prior to interpretation (e.g., Egozcue et al., 2003). Although many traditional water quality plots have been transformed into log-ratio coordinate scatterplots (e.g., the Gibbs diagram and the Stiff diagram), the Piper diagram - one of the

most popular water chemistry graphical methods - has yet to be fully captured by CoDA techniques (e.g., Engle et al., 2017).

Creation of isometric log-ratios is rather flexible and multiple techniques to create them exist. However, in this scenario, prior to the creation of ilr coordinates, two amalgamations merging Na^+ and K^+ ($Na^+ + K^+$) and HCO_3^- and CO_3^{2-} ($HCO_3^- + CO_3^{2-}$) were created following identical amalgamations in the Piper diagram (Fig. 1). This is a trivial but important step, because once parts are amalgamated for purposes of CoDA they represent a reduction in the number of parts and the amalgamations cannot be subsequently split apart. Notably, if one were lacking or not interested in one of the components in either amalgamation they could be removed (e.g., $Na^+ + K^+$ could be replaced by just Na^+ or just K^+). After the amalgamations, we used a sequential binary partition (SBP), which allows the user to organize *D* number of parts into *D* - 1 number of non-overlapping groups of parts which can be converted into ilr coordinates (Egozcue and Pawlowsky-Glahn, 2005). Using expert knowledge and the flexibility of the SBP, one can strategically arrange the constituents (Table 2) to maximize possible geochemical interpretations (Buccianti et al., 2006; Engle and Blondes, 2014; Engle and Rowan, 2013). That is, using a set of defined rules (Egozcue and Pawlowsky-Glahn, 2005), the user can decide the numerator (denoted by +) and denominator (denoted by -) of the various parts in the partitions (Table 2). Once the partition has been created, *D* - 1 number of ilr coordinates (z_i) can be calculated from the raw concentration data of each sample ($\mathbf{x}=(x_1,\ldots,x_D)$) via:

$$z_i = \sqrt{\frac{r_i \times s_i}{r_i + s_i}} \ln \frac{\left(\prod x_j\right)^{\frac{1}{r_i}}}{\left(\prod x_l\right)^{\frac{1}{s_i}}}, \text{for } i = 1, \ldots, D-1 \quad (1),$$

where x_j and x_l are constituents coded as + and - and r_i and s_i are the number of constituents coded as + and -, respectively.

Creating a SBP (and thus the ilr coordinates) which mimics the Piper diagram is straightforward (Table 2). For instance, in order to replicate the central diamond of the Piper diagram (the ratio of weak to strong acids and the ratio of alkali to alkaline earth elements), two ilr coordinates were created using the SBP: z_1, the ratio of alkali to alkaline earth elements, and z_3, the ratio of strong acids to weak acids (cf. Fig. 1). Additionally, to mimic the two trilinear diagrams of the Piper, Mg^{2+} and Ca^{2+} are separated in the trilinear diagram of the cations, and SO_4^{2-} and Cl^- are separated in the anion trilinear; identical partitions were created (z_2 and z_4, respectively). As a formality in the SBP creation process, an additional partition separating the anions and cations (z_0) was also created but is not used (Table 2). From the established SBP and Eqn. 1, four ilr coordinates that mimic the Piper diagram are produced:

$$z_1 = \sqrt{\frac{2}{3}} \ln \frac{\sqrt{[Ca^{2+}][Mg^{2+}]}}{[Na^{+}+K^{+}]} \qquad (2),$$

$$z_2 = \frac{1}{\sqrt{2}} \ln \frac{[Ca^{2+}]}{[Mg^{2+}]} \qquad (3),$$

$$z_3 = \sqrt{\frac{2}{3}} \ln \frac{\sqrt{[Cl^{-}][SO_4^{2-}]}}{[HCO_3^{-}+CO_3^{2-}]} \qquad (4), \text{ and}$$

$$z_4 = \frac{1}{\sqrt{2}} \ln \frac{[Cl^{-}]}{[SO_4^{2-}]} \qquad (5),$$

where units of meq/L are used for each ion. These four coordinates convert the data from the simplex to real space, and are represented within as a four-panel scatterplot with each of the 4 ilr coordinates assigned to 2 of the 8 total axes (Fig. 2). This new plot captures all of the data originally presented in a Piper diagram (discussed below), as well as provides comparative data between the sets of cations not provided in the Piper diagram – Ca^{2+}/Mg^{2+} versus Cl^-/SO_4^{2-}.

The use of ilr coordinates is not limited to proposing a CoDA version of the Piper diagram. Another use of ilr coordinates is to associate them to classical statistical tools without bias, thus

enhancing our understanding of the geochemical processes affecting water chemistry. Additional insight is provided when producing frequency histograms of the ilr coordinate values of a geochemical system (e.g., groundwater chemistry). For instance, some indication about the stochastic mechanisms originating randomness can be identified when analyzing the distribution of ilr coordinates of a system (Agterberg, 2014). The presence of normal, log-normal or heavy tail distributions of the ilr coordinate values requires a different dynamic for the geochemical system, that is open and far from the thermodynamic equilibrium of the system (Buccianti and Zuo, 2016). Often, multimodality of the frequency distribution of ilr coordinate values may result from nonlinear responses to other, unobserved drivers or from a multimodality of the distribution of such drivers revealing interesting information about system resilience (Dakos et al., 2015).

3.1 The ilr-ion plot and its relation to the Piper diagram

In the same way that the relative abundance of the major cations and anions are separated out in the trilinear parts of the Piper diagram, similar plots are provided in the lower left (cations) and upper right (anions) panels of the ilr-ion plot through proper arrangement of the ilr coordinates (Fig. 2). Values of zero on any axis indicate that the geometric mean of all x_j components of the numerator is equal to the geometric mean of all x_l components of the denominator (Eqn. 1). As with a Piper diagram, the areas in the cation and anion fields (top right and bottom left) in which one ion exceeds 50% by equivalents of the other ions can be mapped (e.g., Ca^{2+}-type, Na^{+} + K^{+}-type, Mg^{2+}-type waters; cf. Figs. 1 and 2). The plot in the lower right panel includes the projections of z_1 from the cation plot (alkaline earths vs. alkalis) and z_3 from the anions plot (SO_4^{2-} and Cl^{-} vs. HCO_3^{-}+CO_3^{2-}) and mirrors the information included in the diamond-shaped field in the Piper. In this case, solid grey reference lines extend from 0 on the y-axis indicating whether the Na^{+} + K^{+} amalgamation equivalence exceeds that of the geometric mean of Ca^{2+} and Mg^{2+} (z_1); positive values indicate a preference towards the alkaline earths, while negative values for z_1 indicate a relative enrichment in alkali elements. Note that this approach is similar but not identical to

that of a Piper diagram, where what is projected (using cations as the example) is the sum of Ca^{2+} and Mg^{2+} against the total of $Na^{+} + K^{+} + Ca^{2+} + Mg^{2+}$. A similar solid grey line extending from x=0 compares the relative abundance of an amalgamation weak acids ($HCO_3^{-} + CO_3^{2-}$) to the geometric mean of the strong acids (SO_4^{2-} and Cl^{-}). Finally, the upper left panel of the ilr-ion plot contains information not directly presented in the Piper diagram, namely the log-ratio of Cl^{-} vs. SO_4^{2-} vs. the log-ratio of Ca^{2+} and Mg^{2+} (again with lines projected from x=0 and y=0 as reference; Fig. 2). This plot allows for the distinction between waters of Mg^{2+} and Ca^{2+} vs. Cl^{-} and SO_4^{2-}, differentiated by the solid grey lines on the plot. However, this plot does not allow for classification of Mg-Cl-, Ca-Cl-, Mg-SO_4-, or Ca-SO_4-type waters, as no information about $HCO_3^{-} + CO_3^{2-}$ or $Na^{+} + K^{+}$ is provided on the upper left plot (i.e., waters classified as Mg-SO_4 could actually be Mg-($HCO_3 + CO_3$) waters); it simply provides information on the relative dominance of these 4 ions in a given water sample.

In section 2.1, two issues with the Piper diagram were outlined. One possible problem, being that it can be difficult to distinguish between waters that are composed of combinations of the ions Mg^{2+}, SO_4^{2-}, Ca^{2+}, and Cl^{-} in the Piper program (e.g., is a water composed of Mg-SO_4 or Ca-Cl), can be more easily resolved using the ilr-ion plot. While the upper left panel of the ilr-ion plot, like the Piper diagram as a whole, cannot easily classify a water as a distinct water type (e.g., Ca-SO_4-type waters versus Na-Cl-type waters), it can provide more information about contributions of the aforementioned cations and anions. In order to determine the distinct water type (Ca-SO_4-type waters versus Na-Cl-type), the top two panels combined with the bottom left panel must be used: the top right and bottom left panels can classify the major anion and cation, respectively, while projecting the data into the top left panel allows for exact contributions of the 4 individual ions, Ca^{2+}, Mg^{2+}, SO_4^{2-} and Cl^{-}.

The two black triangles in the ilr-ion plot (Fig. 2B), as presented in Fig. 1B and discussed in Section 2, are composed of Ca-Cl and Mg-SO_4, respectively. By using the extra field – the top left field – the two points are Ca-Cl-type and Mg-SO_4-type, as one black triangle has positive x and y values

(indicative of Ca-Cl type water) and the other black triangle has negative x and y values (indicative of Mg-SO_4 type water). That is, the two black triangle points plot in the Ca- and Mg-type cation field and the Cl- and SO_4-type fields, and the corresponding anion-cation pairs can be determined using the upper left panel. Besides the two "end-members," the contributions of each of these cations and anions can also be resolved in the other samples.

A second issue identified with the Piper diagram is that data which fall along the boundaries of the plots are compressed and tend to overlap, potentially inhibiting their interpretation. This issue may be resolved through the application of an ilr transformation, as the range of any variable changes from 0-100% (or the equivalent in meq/L on the Piper diagram) to a range of $-\infty$ to $+\infty$, permissible for ilr coordinates. Of note, distances in log-ratio coordinates are relative rather than absolute (Otero et al., 2005). Data that plot on the edges of a Piper diagram (which may range an order of magnitude or more) will tend to be spread out when translated to the ilr-ion plot, while data in the center of a Piper plot (which perhaps range less than an order of magnitude) may tend to cluster together when translated to the ilr-ion plot. As an example, the data points in Fig. 1B fall along the Mg^{2+} and Ca^{2+} edge of the cation trilinear panel and the Cl^- and SO_4^{2-} edge of the anions trilinear panel. Because the data are compressed along this edge, it is difficult to assess how the relative abundance of $HCO_3^- + CO_3^{2-}$ or $Na^+ + K^+$ change and the relative abundance of the other anions change accordingly. In the ilr-ion plot of the data (Fig. 2B), one sample has an anomalously low $Na^+ + K^+$ content relative to Ca^{2+} and Mg^{2+} (z_1 is approximately equal to 2 in the lower two panels) and a different sample has a particularly low abundance of $HCO_3^- + CO_3^{2-}$ relative to Cl^- and SO_4^{2-} (z_3 is approximately equal to 0 in the upper two panels). The anomalous nature of these two samples is difficult to identify in the corresponding Piper plot as the data are compressed along the plot edges (Fig. 1B).

3.2 Limitations of the ilr-Ion Plot

One of the setbacks of the ilr-ion plot (and compositional data analysis approaches in general) is that, like all techniques involving log-ratios, the plot cannot handle any missing, zero, or "below detection limit" (e.g., censored) concentrations. However, zeros can be either removed or transformed to a non-zero value in a CoDA dataset via different techniques based on the nature of the zeros in the dataset. Zeros in CoDA are categorized as: essential zeros, or zeros assigned to a constituent due to it being truly absent from the sampled water; count zeros, or zeros assigned when sampling limitations (e.g., volume of water collected) caused the absence of a constituent from a dataset; or rounded zeros, zeroes assigned to constituents when concentrations are below or at the detection limit and therefore, undetectable (Aitchison, 1986). These zeroes are typically removed from a water quality dataset by either removing a constituent (e.g., boron) from an entire sample set (where constituents can be easily removed from a dataset due to subcompositional coherence; Aitchison, 1986) if zeros are present for that constituent in some or all of the samples collected, or zeros can be imputed (i.e., converted from 0 to a non-zero value) using various methods, such as the Bayes-Laplace Bayesian Multiplicative replacement method (Martín-Fernández et al., 2015). Further, zeros can be addressed through amalgamations of constituents. In the case of the amalgamations ($Na^+ + K^+$ and $HCO_3^- + CO_3^{2-}$), it may be possible to assume that the abundance of $Na^+ << K^+$ and $HCO_3^- << CO_3^{2-}$ (at neutral and acidic pH), such that missing or zero data for CO_3^{2-} and K^+ can be ignored.

The effect of zeros in a dataset can be observed in Fig. 2A. The Piper (1944) data set, which contains 8 different water sources, has one sample (B2) with a reported concentration of zero for SO_4^{2-}; therefore, the calculated coordinates z_3 and z_4, those which contain SO_4^{2-}, produce 7 different values while the coordinates z_1 and z_2, those that do not contain a SO_4^{2-} component, produce 8 different values. Therefore, Fig. 2A shows 7 points present on each plot including SO_4^{2-} (top 2 plots), and 8 points present on the plots without SO_4^{2-} (bottom 2 plots; these points overlap and make them hard to distinguish on the scale of Fig. 2A). There are two different options to handle the zeros in this dataset: either all of the

SO_4^{2-} concentration data would be removed from the dataset, or the zeros present in the dataset could be imputed so that their values are estimated in between zero and the detection limit.

Another limitation of the ilr-ion plot is the representation of mixing, ion-exchange, and dissolution/mineralization scenarios. On the Piper diagram, these processes are represented as straight lines, where the two end-members serve as endpoints. However, straight lines in Euclidean space are represented as curved lines once transformed to ilr coordinates; these curved lines may be less easy to identify than the straight lines of the Piper plot and more difficult to graphically predict. For example, the mixing/dissolution example from Fig. 1A is represented on the ilr-ion plot in Fig. 2A. These data, which fell along a straight line on the Piper diagram (Fig. 1A), plot on a curvilinear line after ilr transformation, with the curvature being most dramatic near the endpoints (Fig. 2A). Although it would be difficult to plot the two end-members on the ilr-ion plot and easily draw a mixing line between them, as is possible on the Piper plot, the mixing line is still easy to calculate using CoDA techniques. As noted by Piper in the original manuscript, suggestion of mixing from a graphical plot should be verified through manual calculation (Piper, 1944).

Finally, even though the ilr-ion plot has the ability to reduce the impact from plotting along the boundaries of the figure, it is not immune to sample overlap and some localized data condensing. However, if the data are condensed in all 4 fields of the ilr-ion plot between the same group of samples, then that group of samples is geochemically similar. This is not the case in the Piper diagram, as discussed above, where samples with similar total concentrations of Mg^{2+}, Ca^{2+}, SO_4^{2-}, and Cl^- can be condensed together even if individual concentrations of those 4 ions differ in a sample set. Again, this provides support for using the ilr-ion plot, because it graphically presents more data about the water chemistry of a sample set than the Piper diagram. A summary of the advantages and limitations of both the Piper diagram and the ilr-ion plot can be found in Table 3.

4. Case Study: Graphical interpretation of water chemical data from the Dockum Aquifer, west Texas, USA

In order to further demonstrate the application of the ilr-ion plot (particularly with datasets that can be problematic when plotted on a Piper diagram), we provide a brief analysis using water quality data for groundwater samples from a portion of the Dockum Aquifer in west Texas, USA. A study by Reyes (2014) characterized the hydrogeochemistry of waters produced from the portion of the Late Triassic Dockum Group that overlays the eastern half of the Permian Basin (eastern Central Basin Platform and the Midland Basin; Fig. 3), in hopes of determining the suitability of these waters for use in conjunction with hydraulic fracturing operations in the region. The Dockum dataset was formed from the Texas Water Development Board's Brackish Resources Aquifer Characterization Systems (BRACS) database (http://www.twdb.texas.gov/innovativewater/bracs/database.asp). We further subset the data to include only samples with concentrations of all of the constituents and with charge balances better than 10% (n=295).

Water in the Dockum Aquifer is meteoric and of Holocene- and Pleistocene-age that recharged the formation in eastern New Mexico and subsequently flowed down gradient to the east and southeast, with local recharge from overlying aquifers around the basin margins (Bradley and Kalaswad, 2001; Dutton, 1995; Dutton and Simpkins, 1986). The Dockum Aquifer represents the lowest portion of the regional Upper Aquifer System and is underlain by the gypsum- and halite-rich evaporite confining aquifer system of late Permian age (Basset and Bentley, 1982). It is overlain in the northern portion of the study area by the Ogallala Aquifer, in the southern half of the study area by the Edwards-Trinity Aquifer and in the southwest region of the study area by the Pecos Valley Alluvial Aquifer (Fig. 4) and communication with overlying aquifers locally appears to be an important source of recharge to the Dockum Aquifer (Dutton and Simpkins, 1986). Reyes (2014) determined that there are likely two major types of meteoric water found in aquifer. The groups are distinguished by their relative abundance of

Na^+ versus Ca^{2+} and their salinity, based on a simple 4-part (Ca^{2+}, Cl^-, Mg^{2+}, and Na^+) principal component analysis (Engle et al., 2017). Group 1 waters are more Na^+-rich and generally exhibit higher salinity while Group 2 waters contain relatively more Ca^{2+} and generally exhibit lower salinity (Reyes, 2014). The relative locations of the collected waters also vary within the subsurface: Group 1 waters are found in the western and center parts of Midland Basin and Group 2 waters exist the eastern and southern margins of the Midland Basin. The Na^+-rich, saline waters of Group 1 were hypothesized to be meteoric recharge from eastern New Mexico, acquiring its salinity and mineral content via reactions with silicates, clays, carbonates and evaporate minerals as the groundwater migrates eastward into the basin (Reyes, 2014). The low salinity Group 2 waters are likely older waters that acquired their geochemical make-up via water-rock interactions with overlying aquifers and interactions with the Dockum Group itself.

When plotting the data on a Piper diagram, the groupings assigned by Reyes (2014) and Engle et al. (2017) appear to be valid. The Group 1 data (yellow squares on Fig. 5) are generally more concentrated in $Na^+ + K^+$ (ca. 80 – 100% of the cations in each samples) than the Group 2 samples (gray squares on Fig. 5A). The anion trilinear diagram shows that the Group 1 waters also appear to be mostly, on average, composed of SO_4^{2-} and Cl^-, as suggested by Reyes (2014), while Group 2 waters tend to be scattered across the trilinear diagram, suggesting mixed anion waters. The Group 1 samples appear to have few, if any, geochemical outliers when observing where the data plot on the two trilinear graphs and the diamond-shaped field of the Piper, whereas specific outliers are difficult to observe in the Group 2 dataset, as the data do not seem to follow any specific geochemical trends in the diamond-shaped field and the anion trilinear diagram (Fig. 5A).

Although there is clear separation between the two previously defined groups of samples (particularly apparent in the cation trilinear diagram and the diamond-shaped field), the two trilinear plots suggest less variability in the cation and anion concentrations of Group 1 samples compared to Group 2 samples. Specifically, Group 1 data plot in the $Na^+ + K^+$ corner of the cation trilinear plot,

suggesting little variability in the relative concentrations of the three anions, Ca^{2+}, Mg^{2+}, and $Na^{+} + K^{+}$ (waters are dominated by Na^{+}) and more specifically, fairly constant values of Ca^{2+}/Mg^{2+} (due to data seemingly plotting on a straight line). When projected into the diamond-shaped field of the Piper diagram, the Group 1 data again plot along the edges and in the right corner of the diamond, hindering visual interpretation. The Group 2 waters seems to be much more geochemically diverse, based on the Piper diagram, particularly when considering their anion composition (Fig. 5A). However, they tend to be SO_4^{2-} rich and exhibit higher Ca^{2+} and Mg^{2+} to $Na^{+} + K^{+}$ ratios, with much more variability in the Ca^{2+}/Mg^{2+} ratio than the Group 1 samples.

Inspection of the Piper Diagram provides some insight into the controlling reactions affecting the samples. The Group I waters transition along a J-shaped trajectory in the middle diamond, from Na-Cl-SO_4 to Na-Cl to Na-Ca-Cl. Examination of TDS for the same data showing increasing salinity across this trend suggesting reactions that increase the relative abundance of Ca^{2+} and reduce the relative abundance of SO_4^{2-} are dominant. These findings would suggest dissolution of halite and/or input of reducing Na-Ca-Cl basinal brines from the underlying Evaporite Confining Aquifer System (Dutton and Simpkins, 1986), although more recent isotope data suggest the latter is not likely to be occurring (Reyes, 2014). The relatively low abundance of Mg^{2+} and $HCO_3^{-} + CO_3^{2-}$ make it difficult to discern any patterns of the data. Results for the Type II waters show little coherent pattern suggesting multiple local sources and reactions. This has been interpreted as local input for overlying aquifer units along the study areas (Bradley and Kalalwad, 2001).

When ilr-transforming the Dockum aquifer data and plotting the resulting coordinates on the ilr-ion plot, new conclusions can be made about the original data (Fig. 5B). Crowding and overlap of the data present in the cation trilinear plot of the Piper diagram is removed in the ilr-ion plot (bottom left panel), allowing for better observations of data trends. As expected, the Group 1 data all have a negative y coordinate values and plot in the Na + K-type water region, meaning that relative

concentration of $Na^+ + K^+$ is greater than $Ca^{2+} + Mg^{2+}$. The Group 2 data plot either in the Na + K-type region or in the mixed ion region, as also concluded by the trilinear plot of the Piper diagram. Where the ilr-ion plot really expands on Piper's interpretations is when focusing on determining the relative ratio of Ca^{2+} to Mg^{2+} between the two groups. Due to distortion along the edges of the Piper diagram, the Ca^{2+}/Mg^{2+} ratio of the Group 2 data appeared larger than for the Group 1 samples. With this distortion removed in the ilr-ion plot, we can see that the Ca^{2+}/Mg^{2+} ratios for the samples in Group 1 and Group 2 are roughly equivalent. This is particularly apparent when viewing the histogram of values across all samples for the $[Ca^{2+}|Mg^{2+}]$ ilr coordinate (Fig. 5C), which presents all values for the ilr transformed Ca^{2+}/Mg^{2+} ratio. Moreover, the symmetric histograms of the ilr coordinates for cations (namely $[Ca^{2+}, Mg^{2+}|Na^+ + K^+]$ and $[Ca^{2+}|Mg^{2+}]$) indicate that the ratio between the chemical species is approximately governed by a log-normal distribution. Thus, for this particular study area the product of independent environmental components seems to affect mono and bivalent cations in their transfer from rocks and soils to water (Buccianti and Zuo, 2016).

Transforming the data to ilr coordinates also allows for the identification of outliers; some samples exhibit unusual relative abundances of cations, specifically, two points in the Group 1 data set plot on the far opposite ends of the x axis (annotated on Fig. 5B). These outliers could not be observed in the Piper diagram due to contraction of distances near the plot edges. Furthermore, the ilr-ion plot also features the apparent lack of connection between the alkaline earth metals (Ca^{2+}, Mg^{2+}) and the alkalis ($Na^+ + K^+$). As the ratio of Ca^{2+} and Mg^{2+} to $Na^+ + K^+$ changes (moving up and down the y axis), there is no apparent response in the Ca^{2+}/Mg^{2+} ratio, suggesting that no processes in-situ are affecting the concentrations of both alkaline earths and alkalis or that the Ca^{2+}/Mg^{2+} ratio is heavily buffered agreeing with previous interpretations (Dutton and Simpkins, 1986; Reyes, 2014). This observation is not clear in the Piper plot.

In terms of anions, it is difficult to distinguish the two groups in the Piper diagram due to overall apparent variability in the datasets; this generally appears to also be the case in right panel of the ilr-ion plot. However, the Group 1 waters tend to be positively correlated (r = 0.45), suggesting that as the Cl^- /SO_4^{2-} ratio increases, the ratio of Cl^- and SO_4^{2-} versus the carbonate species also increases. One could interpret this to suggest that if carbonate species are relatively constant (due to equilibrium with carbonate-bearing minerals; Dutton and Simpkins, 1986), then SO_4^{2-} and more so Cl^- are increasing in concentration along the flow path. This agrees with recent interpretation, suggesting input from dissolution of anhydrite and halite (Reyes, 2014). In terms of the distributions of the anion ilr coordinates, both appear to follow non-normal distributions (Fig. 5C). This indicates that various chemical processes can be occurring in the subsurface, impacting cations independently of anions, vice-versa, and arguably processes may be occurring that impact the anions more than the cations (e.g., sulfate reduction, CH_4 oxidation/CO_2 reduction), as is common for higher salinity waters in the Permian Basin (Engle and Blondes, 2014).

The bottom right panel of the ilr-ion plot mimics the diamond-shaped field of the Piper (Fig. 5B). Due to the infinite scale of ilr coordinates, the Group 1 data are no longer are condensed in the corner of the plot, and variability within the cations and anions are easier to distinguish. Both plots show a clear differentiation between the two sample groups, likely owing to the difference in alkali to alkaline earth element ratios. In general, from the Group 1 data, as the relative abundance of Cl^- and lesser so, SO_4^{2-}, increase in the anions, the ratio of alkaline earth to alkali elements increases. The Group 2 samples still exhibit the same randomness that is observed in the Piper, uncorrelated to any change in individual cation or anion concentration.

The top left panel (Fig. 5B), which has no equivalent Piper plot part, allows for the parsing out the relative abundances of Ca^{2+} and Mg^{2+} (x-axis) versus Cl- and SO_4^{2-} (y-axis). The Group 1 data tend to have positive x coordinates, meaning that the Ca^{2+} concentration is generally greater than the Mg^{2+}

concentration, and also tend to have positive y coordinates, as expected, indicating that the samples are more concentrated in Cl^- than SO_4^{2-}. The Group 2 data do not exhibit any real trends; the data are scattered fairly evenly around the origin, spreading into all quadrants of the top left field. These results indicate that as Cl^-/SO_4^{2-} ratios change, there is no corresponding change in the Ca^{2+}/Mg^{2+} ratios. Such findings suggest that if the dominant source of SO_4^{2-} is $CaSO_4$ or gypsum, then there is a buffering reaction occurring which prevents a change in the Ca^{2+}/Mg^{2+} ratio (such as Na^+ for Ca^{2+} ion exchange or precipitation of calcite) (Dutton and Simpkins, 1986; Dutton 1995). Again, no inference can be made from the original Piper diagram, as the Piper does not directly compare Cl^- and SO_4^{2-} to Ca^{2+} and Mg^{2+}. This exercise serves to show different interpretations that can be made between the same set of data when analyzed on both the Piper diagram and the ilr-ion plot.

Either the bottom left or upper right panels can be made to classify waters based on their compositional type (e.g., Na-HCO_3 type waters), similar to how it can be performed in a Piper plot. That is, the position of a water that plots in the top left or bottom right panels panel can be projected, using straight lines, down and across to the bottom left and top right panels, respectively, in order to determine the dominant cation and anion for each sample based upon which field they plot in (see the dotted lines connected to three points on Fig. 6). However, much like the Piper plot, if the data tend to overlap significantly, water type can be difficult to determine for individual points. Alternatively, the water type could be determined prior to plotting and a unique symbol could be assigned for each sample based on water type, as applied to the data set plotted in Fig. 6.

5. Conclusions

Compositional data analysis has become a useful tool in analyzing the geochemical composition of groundwater over the past few decades. Here, we illustrate the usefulness of applying CoDA methods to transform the Piper diagram into a seemingly more data-rich and useful plot for the classification and

analysis of water types and geochemical processes occurring in natural waters. The benefits of using the ilr-ion plot are most apparent when a dataset lie near the edges and corners of the trilinear plots and the diamond-shaped field of the Piper diagram, and when mixed cation and anion compounds (Mg^{2+}, Ca^{2+}, SO_4^{2-}, and Cl^-) occur in larger sample set at varying concentrations, making it difficult to decipher which compounds exist in which samples due to sample overlap preventing projection back into the triangular anion and cation panels. Moreover, analysis of the distribution of the ilr balances used to create the ilr-ion plot also provides information about geochemical processes. However, there are some limitations to the ilr-ion plot. For instance, zeros and missing data must be eliminated or estimated, and easy-to-interpret straight mixing lines are transformed to curves. A relative rather than absolute nature of distances in ilr coordinates can also have the effect compressing together data found in the center of the Piper plot. However, it is apparent that the benefits of using this plot out-weigh the possible challenges of using the Piper plot, as more in-depth geochemical data interpretations can be performed when using the ilr-ion plot, as shown by the multiple datasets used as examples here.

6. Acknowledgements

This work was supported by the USGS Energy Resources Program (Walter Guidroz, Program Coordinator). We thank C. Özgen Karacan, Tina Roberts-Ashby, and two anonymous reviewers for thorough reviews of this manuscript, which greatly increased the quality of this work. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

7. Supplemental Files

The subset data from Reyes (2014) and R code used to create Fig. 5B are provided to allow others to modify the code for their own use.

8. Citations

Agterberg, F. (2014). Geomathematics: Theoretical Foundations, Applications and Future Developments (Quantitative Geology and Geostatistics Series, vol. 18). Springer, 553 pp.

Aitchison, J. (1986). The statistical analysis of compositional data. Chapman and Hall Ltd., London, (Reprinted in 2003 with additional material by The Blackburn Press), 416p.

Bassett, R. L., & Bentley, M. E. (1982). Geochemistry and hydrodynamics of deep formation brines in the Palo Duro and Dalhart basins, Texas, U.S.A. *Journal of Hydrology*, 59(3–4), 331–372.

Blake, S., Henry, T., Murray, J., Flood, R., Muller, M. R., Jones, A. G., & Rath, V. (2016). Compositional multivariate statistical analysis of thermal groundwater provenance: A hydrogeochemical case study from Ireland. *Applied Geochemistry*, *75*, 171-188. DOI: 10.1016/j.apgeochem.2016.05.008

Blondes, M. S., Engle, M. A., & Geboy, N. J. (2016). A Practical Guide to the Use of Major Elements, Trace Elements, and Isotopes in Compositional Data Analysis: Applications for Deep Formation Brine Geochemistry. In Martín-Fernández, J.A and Thió-Henestrosa, S., eds., Compositional Data Analysis: CoDaWork, L'Escala, Spain, *Springer Proceedings in Mathematics & Statistics*, v.187, p. 13-29. ISBN: 978-3-319-44810-7.

Bradley, R. G., & Kalaswad, S. (2001). The Dockum Aquifer in West Texas. In R. E. Mace, W. F. Mullican III, & E. S. Angle (Eds.), Aquifers of West Texas, Texas Water Development Board (pp. 167–174).

Buccianti, A., & Pawlowsky-Glahn, V. (2005). New perspectives on water chemistry and compositional data analysis. Mathematical Geology, 37(7), 703-727. DOI: 10.1007/s11004-005-7376-6.

Buccianti, A., Mateu-Figueras, G., & Pawlowsky-Glahn, V. (Eds.). (2006). Compositional data analysis in the geosciences: from theory to practice. Geological Society of London.

Buccianti, A. (2013). Is compositional data analysis a way to see beyond the illusion? Computers & Geosciences, 50, 165-173. DOI: 10.1016/j.cageo.2012.06.012.

Buccianti, A., & Zuo, R. (2016). Weathering reactions and isometric log-ratio coordinates: Do they speak to each other? Applied Geochemistry, 75, 189-199. DOI: 10.1016/j.apgeochem.2016.08.007.

Cloutier, V., Lefebvre, R., Therrien, R., & Savard, M. M. (2008). Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. Journal of Hydrology, 353(3), 294-313. DOI: 10.1016/j.jhydrol.2008.02.015.

Coleman, J. L., Jr., & Cahan, S. M. (2012). Preliminary catalog of the sedimentary basins of the United States. U.S. Geological Survey Open-File Report 2012-1111, 27 p.

Dakos, V., Carpenter S.R., van Nes, E.H., & Scheffer, M. (2015). Resilience indicators: prospects and limitations for early warnings of regime shift. Philosophical Transactions B, 370, 20130263. DOI: 10.1098/rstb.2013.0263.

Dutton, A. R. (1995). Groundwater isotopic evidence for paleorecharge in US High Plains aquifers. Quaternary Research, 43(2), 221-231. DOI: 10.1006/qres.1995.1022.

Dutton, A. R., & Simpkins, W. W. (1986). Hydrogeochemistry and water resources of the Triassic lower Dockum Group in the Texas Panhandle and eastern New Mexico. Texas Bureau of Economic Geology Report of Investigations 61, 51 p.

Egozcue, J. J., & Pawlowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. Mathematical Geology, 37(7), 795-828. DOI: 10.1007/s11004-005-7381-9

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., & Barcelo-Vidal, C. (2003). Isometric log ratio transformations for compositional data analysis. Mathematical Geology, 35(3), 279-300. DOI: 10.1023/A:1023818214614.

Engle, M. A., & Blondes, M. S. (2014). Linking compositional data analysis with thermodynamic geochemical modeling: oilfield brines from the Permian Basin, USA. Journal of Geochemical Exploration, 141, 61-70. DOI: 10.1016/j.gexplo.2014.02.025.

Engle, M. A., & Rowan, E. L. (2013). Interpretation of Na–Cl–Br systematics in sedimentary basin brines: comparison of concentration, element ratio, and isometric log-ratio approaches. Mathematical Geosciences, 45(1), 87-101. DOI: 10.1007/s11004-012-9436-z.

Engle, M.A., Buccianti, A., Olea, R.A., & Blondes, M.S. (2017). Merging key concepts in the chemistry of natural waters with compositional data analysis: Updates to basic water quality plots. In Compositional Data Analysis: CoDaWork, Hron, K. and Tolosana Delgado, R. (eds.), Abbadia San Salvatore, Italy, June 2017 (pp. 35-46).

Engle, M. A., Reyes, F. R., Varonka, M. S., Orem, W. H., Ma, L., Ianno, A. J., Schell, T.M., Xu, P., & Carroll, K.C. (2016). Geochemistry of formation waters from the Wolfcamp and "Cline" shales: Insights into brine origin, reservoir connectivity, and fluid flow in the Permian Basin, USA. Chemical Geology, 425, 76–92. http://doi.org/10.1016/j.chemgeo.2016.01.025

Jeong, C. H. (2001). Effect of land use and urbanization on hydrochemistry and contamination of groundwater from Taejon area, Korea. Journal of Hydrology, 253(1), 194-210. DOI: 10.1016/S0022-1694(01)00481-4.

Martín-Fernández, J. A., Hron, K., Templ, M., Filzmoser, P., & Palarea-Albaladejo, J. (2015). Bayesian-multiplicative treatment of count zeros in compositional data sets. Statistical Modelling, 15(2), 134-158. DOI: 10.1177/1471082X14535524.

Mateu-Figueras, G., Pawlowsky-Glahn, V., & Egozcue, J.J. (2011). The principle of working on coordinates. In Compositional Data Analysis: Theory and Applications, Pawlosky-Glahn, V. & Buccianti, A. (eds), John Wiley & Sons, 31-42.

Otero, N., Tolosana-Delgado, R., Soler, A., Pawlowsky-Glahn, V., & Canals, A. (2005). Relative vs. absolute statistical analysis of compositions: a comparative study of surface waters of a Mediterranean river. Water research, 39(7), 1404-1414. DOI: 10.1016/j.watres.2005.01.012.

Pawlowsky-Glahn, V., & Egozcue, J. J. (2001). Geometric approach to statistical analysis on the simplex. Stochastic Environmental Research and Risk Assessment, 15(5), 384–398. DOI: 10.1007/s004770100077.

Piper, A. M. (1944). A graphic procedure in the geochemical interpretation of water-analyses. Eos, Transactions American Geophysical Union, 25(6), 914-928. DOI: 10.1029/TR025i006p00914.

Reyes, F R. 2014. “Exploring the Hydrogeologic Controls on Brackish Water and Its Suitability for Use in Hydraulic Fracturing: the Dockum Aquifer, Midland Basin, Texas.” M.S. Thesis: University of Texas at El Paso, 48 p.

Shekhar, S., & Sarkar, A. (2013). Hydrogeological characterization and assessment of groundwater quality in shallow aquifers in vicinity of Najafgarh drain of NCT Delhi. Journal of Earth System Science, 122(1), 43-54. DOI: 10.1007/s12040-012-0256-9.

Stiff Jr, H. A. (1951). The interpretation of chemical water analysis by means of patterns. Journal of Petroleum Technology, 3(10), 15-3. DOI: 10.2118/951376-G.

9. Tables

Table 1. Data used for Figures 1A and 2A.

Sample	Ca^{2+} (mg/L)	Mg^{2+} (mg/L)	$Na^{+} + K^{+}$ (mg/L)	$HCO_3^{-} + CO_3^{2-}$ (mg/L)	SO_4^{2-} (mg/L)	Cl^{-} (mg/L)
Mixing Example 1*	0.0001	0.0001	390	0.0001	0.0001	608
Mixing Example 2*	39	10	47	204	24	16
A_1^	39	10	47	204	24	16
B_1^	40	10	52	207	21	32
b_1^	39	11	56	204	26	32
A_2^	102	19	57.6	203	6.7	199
a_2^	42	22	152	203	49	199
B_2^	466	77	255	166	0	1346
b_2^	65	98	808	199	207	1346
C^	393	1228	10573	139	2560	18360
$CaCl_1$#	40	2	10	10	10	71
CaCl2#	45	22	10	10	75	80
CaCl3#	38	5	10	10	22	75
CaCl4#	33	2	10	20	7	33
CaCl5#	22	12	10	24	45	60
CaCl6#	65	30	10	20	65	85
$MgSO_41$#	4	29	10	10	90	10
$MgSO_42$#	20	36	10	10	60	35
$MgSO_43$#	15	30	10	10	32	24

$MgSO_4$4[#]	8	12	10	10	30	18
$MgSO_4$5[#]	13	24	10	10	66	20
$MgSO_4$6[#]	10	26	10	18	94	10
Endmember1[#]	85	4.9	10	10	19.2	160
Endmember2[#]	8.2	55	10	10	200	14.2

*Synthetic data used in Figures 1A and 2A

^ Data from Piper (1944), used in Figures 1A and 2A

[#]Synthetic data used in Figures 1B and 2B

Table 2. Sequential binary partitions of ions in the Piper plot, used to convert concentration data to isometric log-ratios.

Table 2 - Sequential Binary Partition for Piper Transformation						
	Na + K	Ca	Mg	Cl	$HCO_3 + CO_3$	SO_4
z_0	+	+	+	-	-	-
z_1	-	+	+			
z_2		+	-			
z_3				-	+	-
z_4				+		-

*z_0 is not used because it is not represented in the Piper plot

Table 3. Summary of the pros and cons of the Piper Diagram and ilr-ion plot

	Piper Diagram	**ilr-Ion Plot**
Distinguish Water Types		X
Compares 3 Cations	X	X
Compares 3 Anions	X	X
Compares Mg + Ca to Na + K	X	X
Compares Cl + SO_4 to HCO_3 + CO_3	X	X
Can distinguish between waters with combinations of Ca, Mg, Cl and SO_4		X
Reduces Data Clustering		X
Easily Handles Zeros in the Dataset	X	
Easy Visual identification of Water Processes (e.g., Dissolution)	X	
Easy visualization of less abundant ions		X

10. Figure Captions

Fig. 1. Real (A; Piper, 1944) and hypothetical (B) waters represented on a Piper diagram. Water types, as defined in the diamond-shaped field (Piper, 1944), are presented in the small diamond at the top of the figure. Type 1 water would plot within the un-shaded zones, Type 2 water would plot within the shaded zones, Type 3 water would plot in areas colored yellow, and Type 4 waters plot in areas colored red-orange. All other water types are labeled accordingly. The three different grey dashed lines present on all four trilinear diagrams create the triangular zones which, if data plotted within that triangle, a particular anion (pair) or cation (pair) would dominate (>50% of the total) compared to the others. A: The data from Piper (1944) are presented as red triangles; the blue triangles represent two hypothetical end-members, one with a high Na-Cl concentration and one with a low Na-Cl concentration. The blue line represents locations that a mixture of these two hypothetical end-members would fall, dependent on the initial chemical composition of the endmember and the proportion of each used to create the mixture. B: Hypothetical waters with various concentrations of Mg-SO_4 and Ca-Cl; waters with concentrations of Mg-SO_4 > Ca-Cl (orange triangles), waters with concentrations of Ca-Cl > Mg-SO_4 (green triangles); the 2 black triangles on each plot represent the 2 waters with the highest concentration of Mg-SO_4 and Ca-Cl.

Fig. 2. Isometric log-ratio (ilr)-ion plot. The ilr-ion plot's four fields mimic the Piper diagram; the upper-right panel mimics the anion trilinear diagram of the Piper, the lower-left panel mimics the cation trilinear diagram of the Piper, the lower-right panel mimics the diamond-shaped field of the Piper, while the upper-left panel presents data not specifically presented in the Piper. The three grey dashed lines on both figures equate to the same grey dashed lines present on Fig. 1 in the trilinear diagrams. The solid grey lines denote where y and x equal 0, which indicate a shift in dominance from the denominator of

the ilr coordinate ($x < 0$, $y < 0$) to the numerator of the plotted ilr coordinate ($x > 0$, $y > 0$). The top left panel of each ilr plot explicitly states which cation or anion dominates on either side of the y=0 and x=0 lines, respectively. To simplify labeling of the axes, the formulas for the ilr coordinates are reduced to the ions shown in the numerator and the denominator, separated by a vertical line. That is, the equation for z_2 is reduced from its full form provided above to [Ca|Mg]. 2A: Transformation of Fig 1A, where the original data from Piper (1944) is represented as red triangles while a hypothetical mixing/dissolution example is represented as blue triangles connected by a blue mixing line in each of the four panels of the plot. 2B: Transformation of Fig 1B, where the green triangles are hypothetical waters with concentrations of Ca and Cl > Mg and SO_4, the orange triangles are hypothetical waters with concentrations of Mg and SO_4 > Ca and Cl, and the black triangles are 2 hypothetical waters with drastically different Ca-Mg-Cl-SO_4 ion chemistry. The orange and green dashed lines, correlate to the $MgSO_4$ concentrated black triangle and $CaCl_2$ concentrated black triangle, respectively. The lines indicate how the triangles project into the upper left field from their positions on the two fields (bottom left and top right) that mimic the cation and anion plots, and how the exact composition of those black triangles can be determined.

Fig. 3. Map showing the extent of the Dockum Aquifer in Texas (Bradley and Kalaswad, 2003) and the two subbasins of the Permian Basin, the Delaware and Midland Basins (Coleman and Cahan, 2012).

Fig. 4. Stratigraphic chart of the Midland Basin, Texas showing major aquifer systems. Modified from Bassett and Bentley (1982), Bradley and Kalaswad (2003), Engle et al. (2016), and references therein.

Fig. 5. Piper diagram (A), ilr-ion plot (B), and histograms of ilr coordinates (C) for data from the Dockum aquifer as compiled by Reyes (2014). The three grey dashed lines on the ilr-plot correspond to the same dashed lines on the Piper plot. These zones distinguish where one cation or anion (or pair) dominates (> 50%) compared to the other two. Group 1 data is labeled with yellow points and bars, while Group 2

data is labeled with gray points and bars on all plots. The solid grey lines on the ilr-plot indicate where y and x equal 0, which indicate a shift in dominance from the denominator of the ilr coordinate ($x < 0$, $y < 0$) to the numerator of the plotted ilr coordinate ($x > 0$, $y > 0$), explicitly defined in the top left panel. The R code used to create 3B is provided in the supplementary material.

Fig. 6. ilr-ion plot of groundwater chemical data from the Dockum aquifer (data compiled in Reyes, 2014), classified as either Group 1 (circles) or Group 2 (triangles), and color coded by water type, as determined using the raw water chemistry data. The dotted and solid grey lines are the same as those present on Fig. 3B. The dotted red, teal, and green lines link one sample (a red, teal and green triangle) across three different panels: the upper left, the upper right, and the bottom left. This allows for distinct water classification; the red triangle is Na + K-HCO_3 + CO_3 type, the teal triangle is Ca-Cl type, and the yellow triangle is Cl type.

Figure 1.

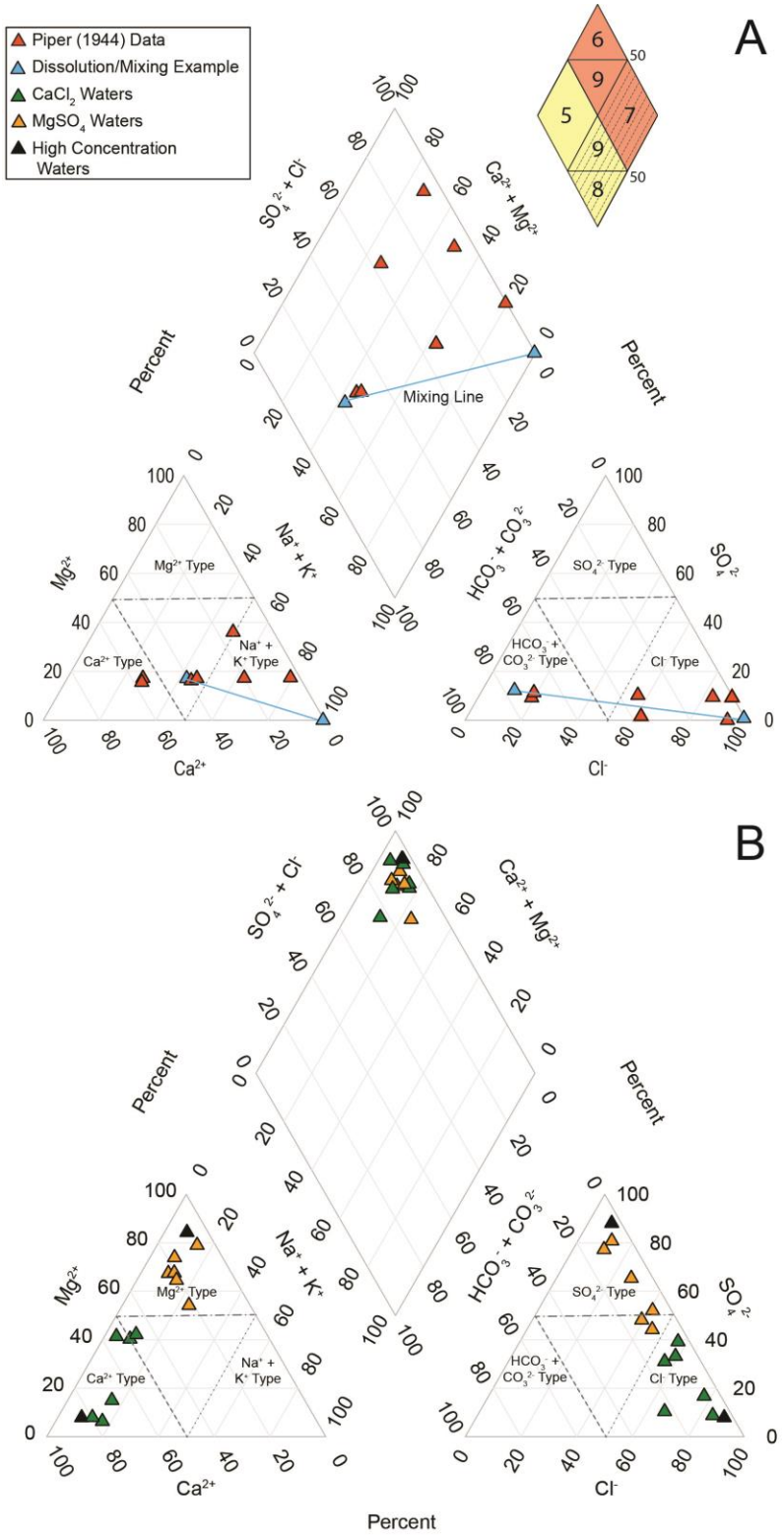


Figure 2.

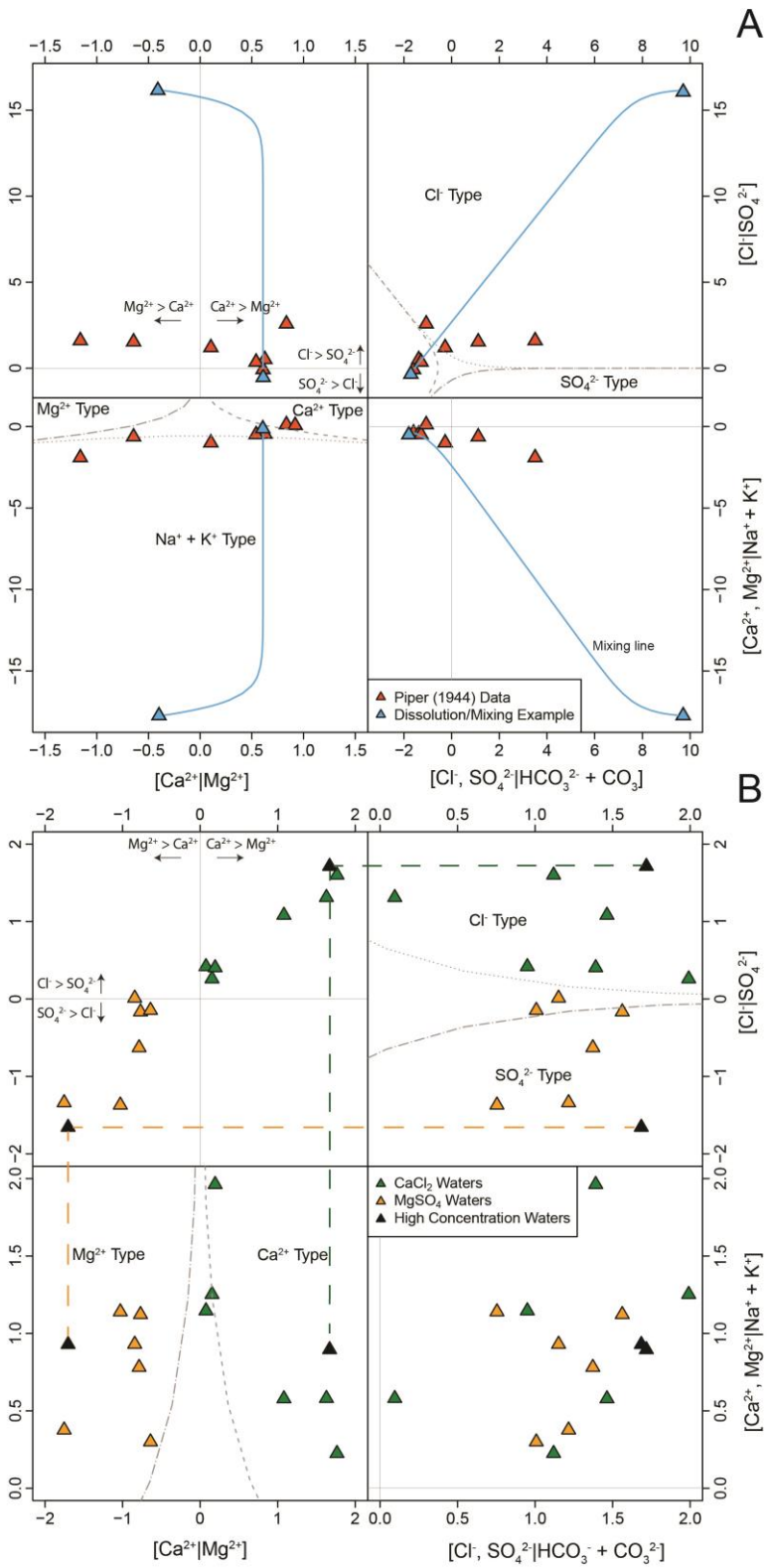


Figure 3.

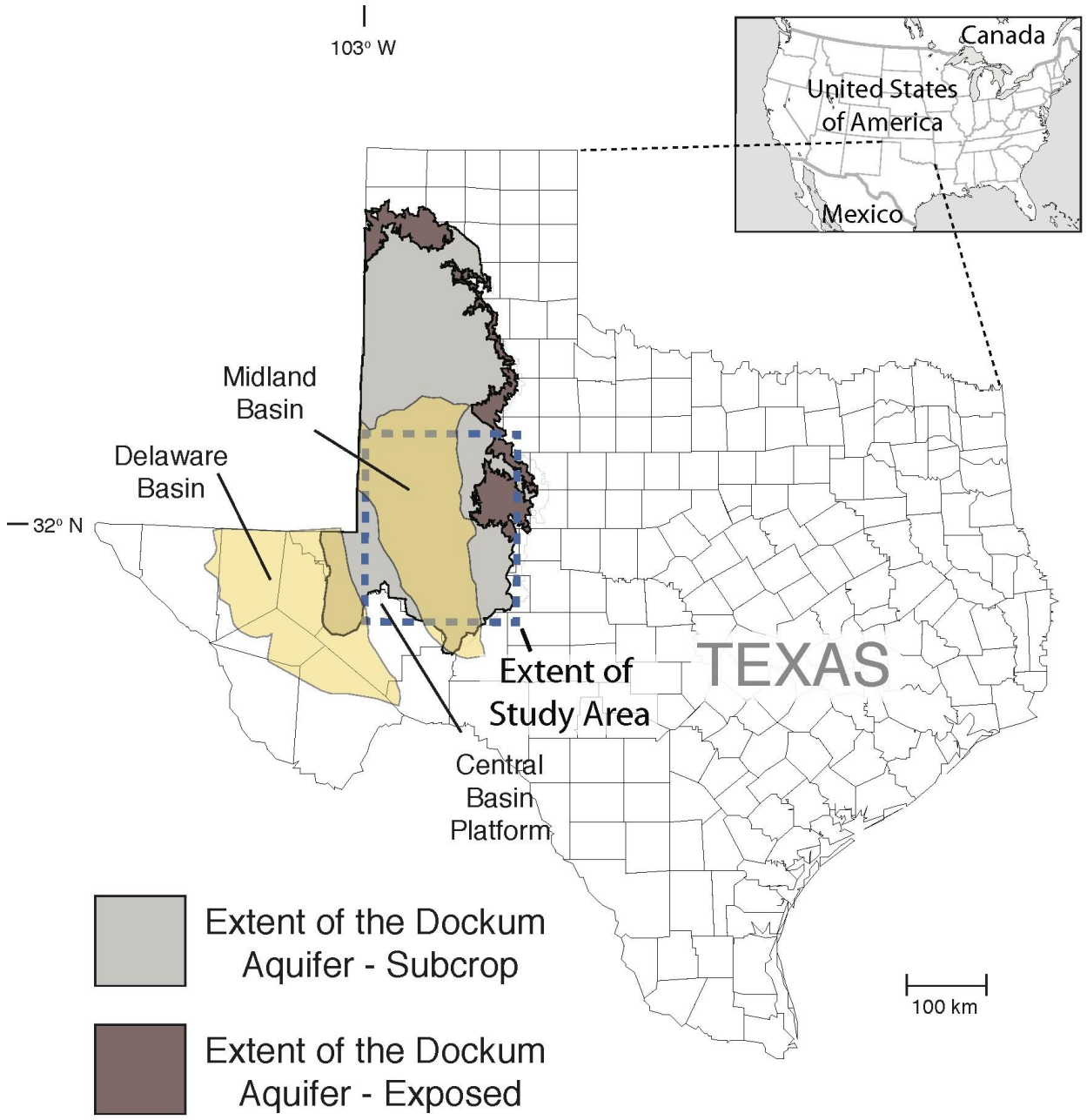


Figure 4.

Era	System	Series	Group	Formation	Aquifer	Hydrogeologic Unit
Cenozoic	Quaternary			Pecos Valley Alluvium	Pecos Valley	Upper Aquifer System (UAS)
	Tertiary	Late Miocene to Pliocene		Ogallala	Ogallala	
Mesozoic	Cretaceous		Fredericksburg	Kiamichi	Edwards-Trinity (High Plains)	
				Edwards		
				Comanche Peak		
				Walnut		
			Trinity	Antlers		
	Jurassic		Not present	Not present	Not present	
	Triassic		Dockum	Trujillo Sandstone	Upper Dockum	
				Tecovas Formation	Lower Dockum	
				Santa Rosa Formation		
Paleozoic	Permian	Ochoan		Dewey Lake, Salado, Castile	Non-potable water and hydrocarbons	Evaporite Confining System (ECS)
		Guadalupian		Multiple		
		Leonardian		Sprayberry, Dean		Deep Basin Brine Aquifer System (DBBAS)
		Wolfcampian		Wolfcamp		
	Penn. to Upper Devonion			Cisco, Canyon, Strawn, Atoka, Barnett, Mississippian, Woodford		
	Lower Devonion to Cambrian			Thirtyone, Wristen Group, Fusselman, Montoya, Simpson Group, Ellenberger, Cambrian		Deep Basin Meteoric Aquifer System (DBMAS)
Precambrian						Basement Aquitard (BA)

Figure 5.

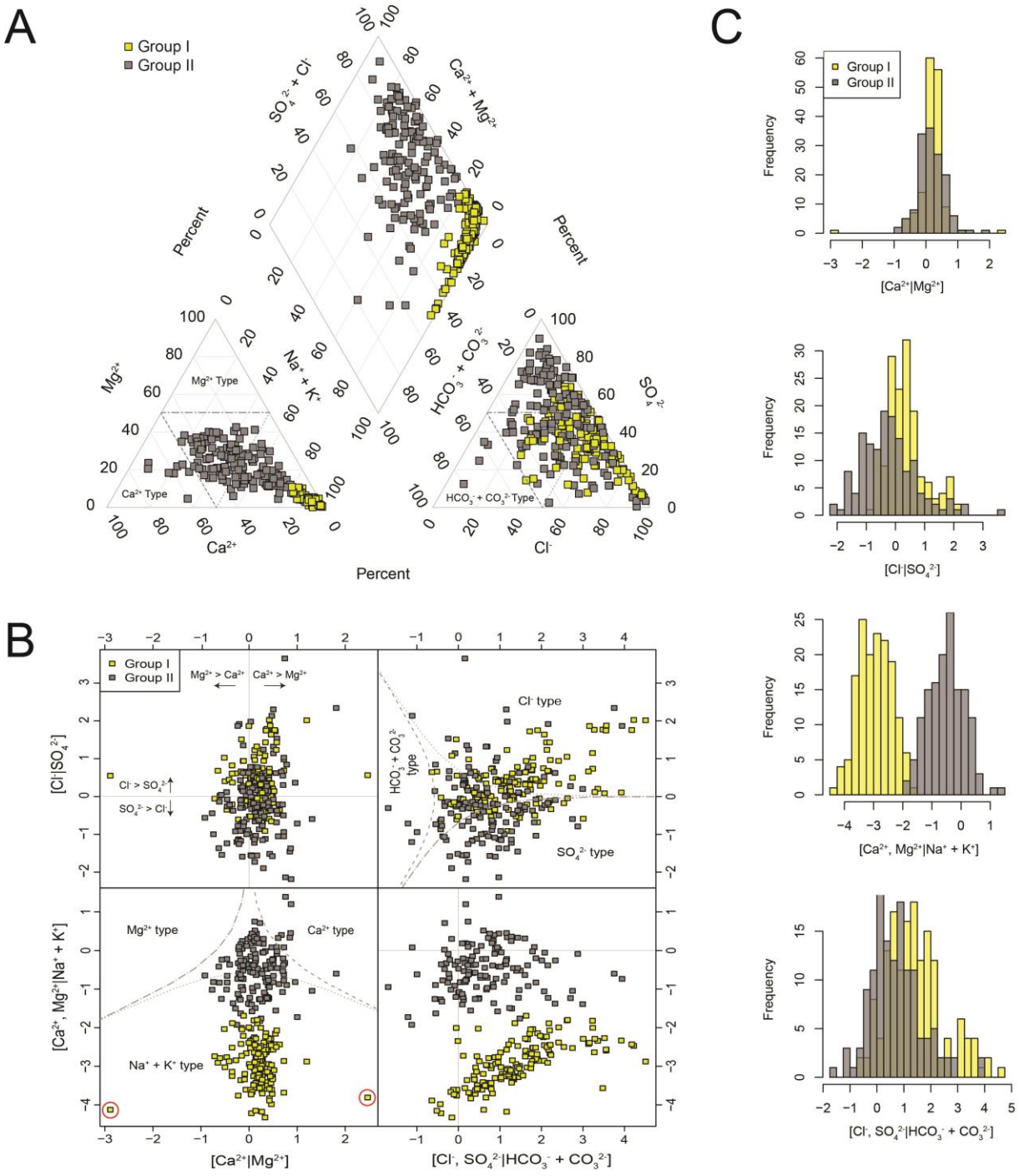
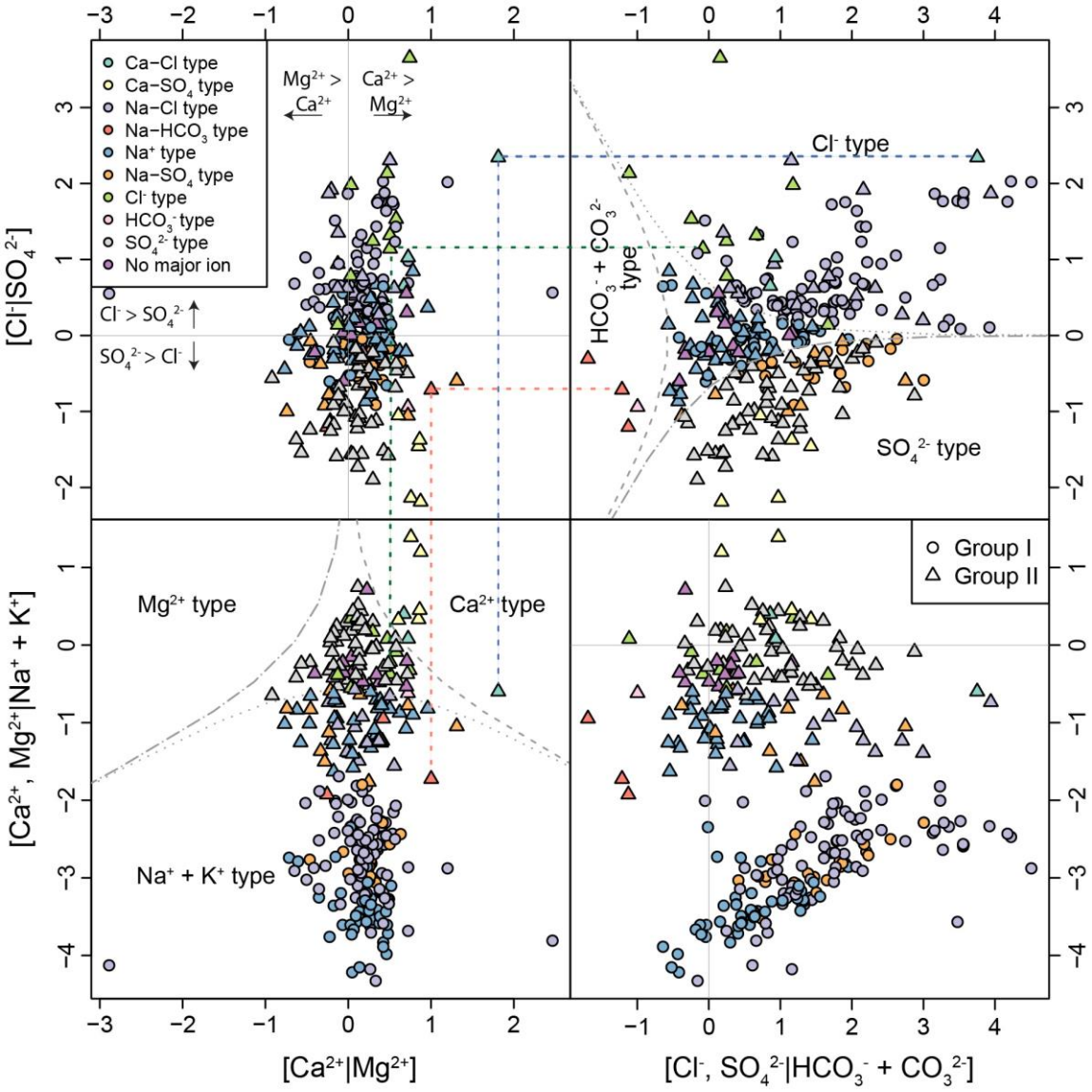


Figure 6.



Highlights

- Piper diagram is a staple of graphical water chemistry interpretation
- The Piper loses detail along plot edges and some ions cannot be differentiated
- Geochemical species in Piper diagram converted to isometric log-ratio coordinates
- New plot made from coordinates has advantages of Piper and fixes its shortfalls