AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2015WR018386

Compositional data analysis as a robust tool to delineate hydrochemical facies within and between gas-bearing aquifers

D. Des. R. Owen[1], V. Pawlowsky-Glahn[2], J. J. Egozcue[3], A. Buccianti[4], and J. M. Bradd[5]

[1]School of Earth, Environmental and Biological Sciences, Queensland University of Technology, Brisbane, Queensland, Australia, [2]Department of Computer Science, Applied Mathematics, and Statistics, University of Girona, Girona, Spain, [3]Department of Civil and Environmental Engineering, Technical University of Catalonia, Barcelona, Spain, [4]Department of Earth Sciences, University of Florence, Florence, Italy, [5]School of Earth and Environmental Sciences, University of Wollongong, Wollongong, New South Wales, Australia

Key Points:
- Compositional distinctions within and between coal seam gas groundwater (CSG) are described
- The approach uses isometric log ratios to characterize water types based on the relative proportions of ions in the simplex
- An isometric log ratio plot is presented that can be used to delineate groundwater associated with high gas concentrations

Supporting Information:
- Supporting Information S1

Correspondence to:
D. D. R. Owen,
des.owen@nrra.science

Citation:
Owen, D. D. R., V. Pawlowsky-Glahn, J. J. Egozcue, A. Buccianti, and J. M. Bradd (2016), Compositional data analysis as a robust tool to delineate hydrochemical facies within and between gas-bearing aquifers, *Water Resour. Res.*, *52*, 5771–5793, doi:10.1002/2015WR018386.

Received 17 NOV 2015
Accepted 23 JUN 2016
Accepted article online 27 JUN 2016
Published online 5 AUG 2016

Abstract Isometric log ratios of proportions of major ions, derived from intuitive sequential binary partitions, are used to characterize hydrochemical variability within and between coal seam gas (CSG) and surrounding aquifers in a number of sedimentary basins in the USA and Australia. These isometric log ratios are the coordinates corresponding to an orthonormal basis in the sample space (the simplex). The characteristic proportions of ions, as described by linear models of isometric log ratios, can be used for a mathematical-descriptive classification of water types. This is a more informative and robust method of describing water types than simply classifying a water type based on the dominant ions. The approach allows (a) compositional distinctions between very similar water types to be made and (b) large data sets with a high degree of variability to be rapidly assessed with respect to particular relationships/compositions that are of interest. A major advantage of these techniques is that major and minor ion components can be comprehensively assessed and subtle processes—which may be masked by conventional techniques such as Stiff diagrams, Piper plots, and classic ion ratios—can be highlighted. Results show that while all CSG groundwaters are dominated by Na, HCO_3, and Cl ions, the proportions of other ions indicate they can evolve via different means and the particular proportions of ions within total or subcompositions can be unique to particular basins. Using isometric log ratios, subtle differences in the behavior of Na, K, and Cl between CSG water types and very similar Na-HCO_3 water types in adjacent aquifers are also described. A complementary pair of isometric log ratios, derived from a geochemically-intuitive sequential binary partition that is designed to reflect compositional variability within and between CSG groundwater, is proposed. These isometric log ratios can be used to model a hydrochemical pathway associated with methanogenesis and/or to delineate groundwater associated with high gas concentrations.

1. Introduction

The characterization of water types first proposed by *Chebotarev* [1955] provides information on water-rock interaction and it has become an important, preliminary step in hydrochemical studies. This information is used for water resource management decisions, for example, conceptual hydrogeological models, or monitoring programs.

Traditionally, water types are classified by their major ion components using conventional methods, such as Piper plots, Stiff diagrams, or Schoeller plots. Despite being commonplace, these types of plots can be visually messy, and the concentrations of minor ion components are often masked. Furthermore, plots such as Piper plots are mathematically limited and implicitly invoke spurious correlation caused by scaling because they employ amalgamation which does not preserve distances within what is an already restricted sample space (called the simplex) [*Egozcue and Pawlowsky-Glahn*, 2005; *Bacon-Shone*, 2006; *Pawlowsky-Glahn and Egozcue*, 2006]. This occurs because hydrochemical data are compositional by nature; in particular, they are quantitative data where each component is a proportion of a given total [*Egozcue and Pawlowsky-Glahn*, 2005; *Bacon-Shone*, 2006]. Spurious correlation was initially observed by *Pearson* [1897] and later investigated in a geological context [e.g., *Chayes*, 1960]. The issue was comprehensively examined in the 1980s and additive log ratios (alr) and centered log ratios (clr) transformation were proposed as solutions [*Aitchison*, 1982, 1986]. More recently, the isometric log ratio (ilr) approach was developed which ultimately recognized

that compositions can be represented in orthogonal (Cartesian) coordinates in the simplex [*Egozcue et al*., 2003; *Egozcue and Pawlowsky-Glahn*, 2005]. Collectively, these techniques are termed compositional data analysis.

Compositional data analysis is concerned with understanding the relative proportions of individual measurements within a sample. These mathematical techniques can be applied to any discipline where compositional data exist [*Pawlowsky-Glahn and Egozcue*, 2006; *Lovell et al*., 2015; *Pawlowsky-Glahn et al*., 2015]. Despite being mathematically sound and offering a number of advantages [*Nisi et al*., 2015], compositional data analysis techniques are infrequently employed in hydrochemical studies. Previous research has demonstrated some specific examples of how to describe the evolution of ion proportions during simple processes (e.g., diagenesis of carbonates), or to develop log ratio alternatives to traditional mixing lines (e.g., for Na-Br-Cl brines) [*Bicocchi et al*., 2011; *Engle and Rowan*, 2013; *Engle et al*., 2016].

In addition, compositional data analysis allows researchers to describe distinct relationships between ions, rather than just perform simple classifications based on the dominant ions. However, the application of compositional data techniques in this way has not been explored or demonstrated.

A typical example of the classification of a water type using conventional methods is the Na-HCO_3 or Na-Cl-HCO_3 water type described for coal seam gas (or coal bed methane) groundwater [*Van Voast*, 2003; *Kinnon et al*., 2010; *Hamawand et al*., 2013]. Some variability in the dominance of these ions has been noted and variability in minor ion components, such as Ca and Mg, has also been observed [*Taulis and Milke*, 2007; *Papendick et al*., 2011; *Hamawand et al*., 2013; *Baublys et al*., 2015; *Owen et al*., 2015]. This variability of both major and minor ion components offers important information about the hydrochemistry of this water type that can be explored using compositional data analysis.

In this paper we apply compositional data analysis techniques to describe variability of major ions within and between coal seam gas- (or coal bed methane-) bearing groundwater. We use examples from the Surat/Clarence-Moreton basin (Australia), Gunnedah Basin (Australia), the Powder River Basin (USA), and the Gulf of Mexico Basin (USA). The aims of this paper are to (1) describe fundamental compositional similarities/differences between coal seam gas groundwater; (2) provide a robust and descriptive means of comparing water types; and (3) assess hydrochemical variability within and between large, highly variable data sets with respect to the hydrochemical evolution of coal seam gas or groundwater associated with high gas concentrations. By deriving isometric log ratios from appropriately designed sequential binary partitions, the relative importance of ions within the simplex can be derived. These can be used as a unique identifier of groundwater that is hydrochemically characterized by a particular process, end-member, and/or aquifer.

2. Background

2.1. Geological Overview

Dissolved major ion data from groundwater in the following basins were used in this study: Powder River Basin (USA); Gulf of Mexico Basin (Claiborne, Wilcox, and Comanchean groups—Texas, USA); and Surat and Clarence-Moreton basins (Australia); Gunnedah Basin (Australia). All of these basins are large sedimentary basins where commercial quantities of CSG occur or are likely to occur. Common features between all these basins are (1) the coal seams occur as discontinuous seams throughout the relevant coal-bearing formation(s); (2) the coal is typically low rank, subbituminous, and high in vitrinite; and (3) the gas is typically biogenic [*Grossman et al*., 1989; *Rice*, 1993; *Pratt*, 1998; *Zhang et al*., 1998; *Montgomery*, 1999; *Draper and Boreham*, 2006; *Flores et al*., 2008; *Green et al*., 2008; *Golding et al*., 2013; *Hamilton et al*., 2014; *Quillinan and Frost*, 2014; *Baublys et al*., 2015]. For readability, more information on the hydrogeology of these basins is discussed where relevant throughout the text.

2.2. Evolution of Coal Seam Gas Groundwater Hydrochemistry

In order to extract coal seam gas (or coal bed methane) resources, water is extracted from the coal-bearing aquifer. As a result, both effective gas exploration and management of coal seam gas (CSG) extraction requires an understanding of the water resource that accompanies the gas resource in the coal-bearing aquifer. The typical Na-HCO_3 water type of CSG waters is, by now, well known. In an important seminal paper, *Van Voast* [2003] proposed that the evolution of CSG groundwater in a number of

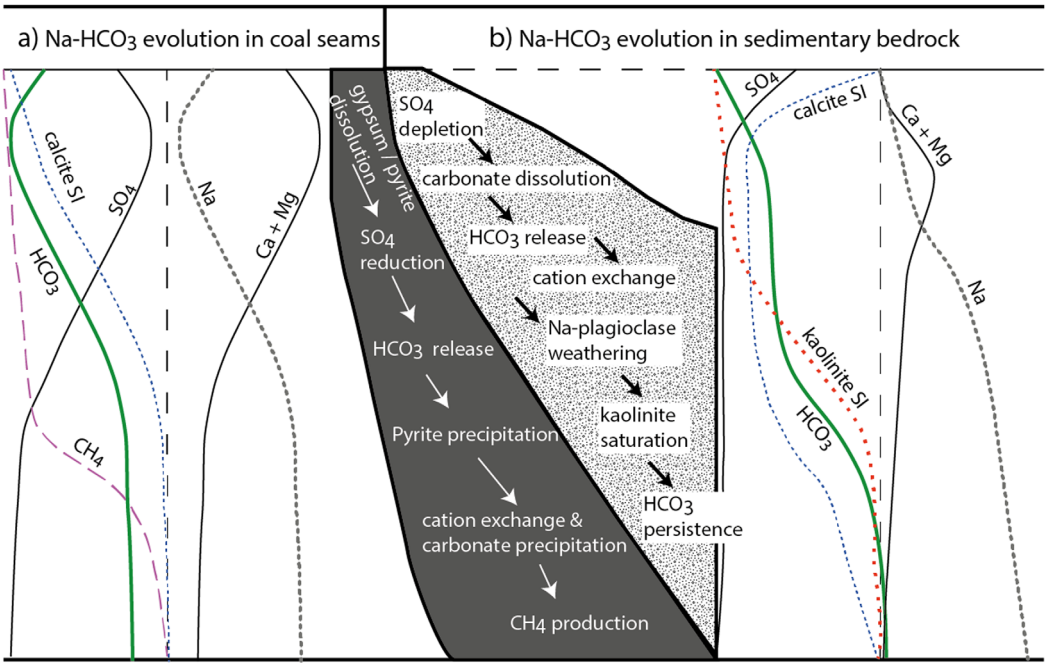

Figure 1. Conceptual model for the evolution of Na-HCO_3 water types in (a) coal-seam-gas bearing aquifers, as first proposed by *Van Voast* [2003], and (b) in non-coal-bearing aquifers [*Venturelli et al.*, 2003; *Chae et al.*, 2006]. The conceptual cross section (center) shows the sequential steps in hydrochemical evolution, and the associated hydrogeological processes that are proposed as universal processes contributing to the evolution of these water types. Examples of the changes in abundance of individual anions and cations for (a) (CSG groundwater) and (b) (other Na-HCO_3 water types) are shown to the left and right, respectively. For CSG groundwater, (a) SO_4 reduction and the dissolution of coal and methanogenesis drive HCO_3 enrichment, leading to the depletion of SO_4, Ca, and Mg depletion at peak CH_4 concentrations [*Van Voast*, 2003; *Taulis and Milke*, 2007]. In the absence of coal and methanogenesis (b), the depletion of Ca and Mg can also proceed via cation exchange and sodium-plagioclase weathering. However, for Na-HCO_3 waters to persist under these conditions, a saturation point for certain weathering products, e.g., kaolinite, must be reached in order to maintain enriched HCO_3 values as weathering proceeds [*Venturelli et al.*, 2003].

basins in the USA followed a pathway where the bacterial reduction of sulfate was the primary HCO_3 production mechanism (equation (1)), which subsequently drove depletion of Ca and Mg, via carbonate precipitation.

$$2CH_2O + SO_4^{2-} \rightarrow 2HCO_3^- + H_2S^-. \quad (1)$$

In addition, cation exchange can also drive Ca and Mg depletion and Na enrichment (Figure 1a). Prior to *Van Voast*'s [2003] review, a case study suggested that the Na-HCO_3 water type for gas-bearing waters in the Gulf of Mexico Basin may also be evolving earlier on in the hydrochemical pathway via processes such as cation exchange [*Grossman et al.*, 1989].

Ultimately, the resulting Na-HCO_3 or Na-HCO_3-Cl water type of CSG groundwater is a common hydrochemical end-member that has been derived from a complex set of processes, some of which are not known or well understood [*Gluskoter and Ruch*, 1971; *Huggins and Huffman*, 1995; *Kimura*, 1998; *Conrad*, 1999; *Hjelm et al.*, 1999; *Conrad*, 2005; *Öberg et al.*, 2005; *Öberg and Sandén*, 2005; *Yudovich and Ketris*, 2006] such as:

1. The extent to which biological processes are consuming or producing bicarbonate.
2. The influence of clays and organic matter on the retention of Cl and associated ions such as Na.
3. The role of biogeochemical cycling of chlorine.
4. The nature of the coal cleat and continuity of coal seams.

Furthermore, while CSG groundwater can range from fresh to brackish, many coals associated with CSG formed in freshwater, lacustrine, and environments [*Rice et al.*, 2008; *Cook and Draper*, 2013; *Hamawand et al.*, 2013; *Jell et al.*, 2013]. Therefore, unless there has been brine migration or marine incursions, other known hydrochemical tools, such as Na/Br, Cl/Br ratios/mixing lines cannot be used as hydrochemical indicators of coal or methanogenesis.

In addition, Na-HCO_3 water types may also evolve independently of SO_4 reduction via other processes, such as clay mineral weathering under kaolinite saturation and cation exchange [*Venturelli et al*., 2003; *Chae et al*., 2006] (Figure 1b). A challenge in CSG studies, therefore, is being able to use major ions to effectively describe and compare hydrochemical variability within and between CSG aquifers, particularly in large data sets with a high degree of variability.

2.3. The Need for More Informative Characterization of CSG Hydrochemistry

A first step when attempting to understand water resources associated with CSG is often the characterization of the hydrochemistry within and between aquifers. In many cases, large data sets are compiled using historical major ion data, such as government monitoring data, exploration data and data from CSG production wells to make these assessments [e.g., *Bartos and Ogle*, 2002; *WorleyParsons*, 2012; *Feitz et al*., 2014; *GeoScience Australia*, 2016]. In these cases, conventional techniques are often applied to characterize CSG groundwater and groundwater in surrounding aquifers [*Bartos and Ogle*, 2002; *Van Voast*, 2003; *Taulis and Milke*, 2007; *Kinnon et al*., 2010; *Papendick et al*., 2011; *WorleyParsons*, 2012; *Taulis and Milke*, 2013; *Atkins et al*., 2015].

Considering that we know that Na-HCO_3 water types can evolve via other means not related to coal, we have to ask how informative an approach that simply highlights the dominant ions actually is. During the preliminary stage of characterizing water types in CSG-related studies a researcher is presented with a dilemma: How can a CSG groundwater end-member be characterized effectively? Processes, such as SO_4 reduction (equation (1)), could help inform the hydrochemical evolution of these water types, but only if this individual process has a dominant influence on the ion composition. The process of SO_4 reduction and cation exchange are not endemic to CSG water types.

While we know that there must be other complex processes that influence the CSG hydrochemistry, how can we begin to make use of this complexity to understand hydrochemical variability if these processes are not well understood and/or the hydro-biogeochemical dynamics of the system are not yet described? Actually, this complexity is to the advantage of the researcher because complex sets of processes may result in distinct hydrochemical responses that can be used to better characterize water types. Compositional data analysis provides a means of characterizing water types based on the real relative abundance of ions in the sample. As a result, the hydrochemical characterization of water types within and between aquifers can be made in a much more informative way.

3. Methods

3.1. Data and Data Preparation

A number of data sets pertaining to geographical locations were used in this study, as follows: Powder River basin—publicly available data from the Montana Bureau of Mines and Geology database (http://mbmggwic.mtech.edu/) and production water data from *Rice et al*. [2000]; Gulf of Mexico Basin—published data from *Grossman et al*. [1989] and *Zhang et al*. [1998] (raw data provided by Prof. Grossman E.L., Texas A&M University); Surat and Clarence-Moreton basins—publicly available data from the QLD government groundwater database (data provided by QLD Department of Natural Resources and Mines: database manually screened as per *Owen and Cox* [2015]) in combination with published data from *Duvert et al*. [2015] and unpublished data from J. Li (data provided by Queensland University of Technology), and CSG data provided by Arrow Energy; Gunnedah Basin—unpublished data provided by J. Bradd, University of Wollongong.

New data for the Surat and Clarence-Moreton basins are also presented. These were collected at CSG production wells or at domestic (landholder) wells screened in the coal measures. Samples for CH_4 were collected in glass vials with rubber butyl stoppers and a sulfuric acid preservative. All samples were placed on ice until analysis. Major ions were analyzed using the following methods: APHA 3120B; APHA 4500-Cl-G; and APHA 4500-SO4-E at NATA accredited ALS Laboratories, Brisbane, QLD. CH_4 data were analyzed via headspace gas chromatography, also at ALS Laboratories, Brisbane, QLD.

The following major ions were included in analyses: Na, K, Ca, Mg, HCO_3, Cl, SO_4, and, where available, F. Data below detection limit (typically SO_4 for CSG water types) were imputed using the R package *zCompositions* via the log ratio Data Augmentation function (lrDA): a function that is based on the log ratio Markov Chain Monte Carlo Data Augmentation (DA) algorithm [*Palarea-Albaladejo and Martín-Fernández*, 2015]. This

approach differs from conventional methods of replacing left-censored data with arbitrary values below detection limit. It provides a means of estimating values below the detection limit in a manner that preserves the relative structure of the data. For the Powder River Basin data set, multiple detection limits were dealt with by first imputing (estimating) data from the sub–data set with the highest detection limit and then appending this imputed data to the data set with the next highest detection limit; this process was completed until all data below detection limit were imputed.

3.2. Compositional Data Analysis

Compositional data analysis essentially applies three main principles [*Egozcue and Pawlowsky-Glahn*, 2011]: (1) scale invariance—the information is equivalent regardless of the units used; (2) subcompositional coherence—analysis of a subset must not contradict analysis of the entire composition; (3) permutation invariance—the order of parts should be irrelevant, although this third principle is in general trivially satisfied by any given data set.

Here we apply two log ratios when analyzing the data set: (1) the centered log ratio (clr), a log ratio of an individual component and the geometric mean of all (equation (2)); and (2) the isometric log ratio (ilr), a log ratio of geometric means of groups of parts or components (equation (5)). Both transformations have different properties which should be taken into account when interpreting data.

3.2.1. The Centered Log Ratio Transformation (clr)

The clr transformation is obtained by dividing components by the geometric mean of the parts (components, ions) and then taking the log ratios [*Aitchison*, 1982, 1986]. The result of a clr transformation for a *D* matrix is a *D* matrix number of clr values.

$$\mathrm{clr}(\mathrm{x}) = \left[\ln\frac{x_1}{g(\mathrm{x})};\ldots;\ln\frac{x_D}{g(\mathrm{x})}\right], \tag{2}$$

where *g(x)* is the geometric mean as defined by

$$g(\mathrm{x}) = \sqrt[D]{x_1 \cdots x_D}. \tag{3}$$

The clr is a transformation which treats the parts of the composition symmetrically. It produces coordinates which are useful for examining properties of the whole clr-vector, i.e., the behavior of ions with respect to all ions in the composition. The clr transformed data are particularly useful for constructing biplots to represent an approximation of the variability within the data set. However, because the clr-transformed coordinates ultimately sum to zero, they are constrained to a subspace of real space and are represented on a hyperplane in D-dimensional real space; consequently, the covariance and correlation matrices are singular [*Egozcue et al*., 2003; *Pawlowsky-Glahn and Egozcue*, 2006]. As a result, some care is needed when interpreting clr transformed data because the clr-coordinates represent the proportion in the numerator with respect to the geometric mean of all the parts in the composition. Here we use clr transformed data to construct biplots to describe variability and to then assist in developing meaningful sequential binary partitions that are used to compute ilr-coordinates.

3.2.2. The Isometric Log Ratio Transformation (ilr)

The ilr transformation differs from the clr transformation in that it uses a sequential binary partition (Table 1) to describe the orthonormal basis to which correspond *D*-1 Cartesian coordinates (ilr-coordinates): these orthonormal coordinates, called balances, can then be modeled [*Egozcue et al*., 2003].

Table 1. Sequential Binary Partition (SBP) of a Seven-Part Composition ($x_1, x_2 \ldots x_7$) Deriving Six Orthonormal Coordinates ($z_1, z_2 \ldots z_6$) for ilr Calculation

	Partition of Parts						
Balance	x_1	x_2	x_3	x_4	x_5	x_6	x_7
z_1	1	1	1	1	−1	−1	−1
z_2	1	1	−1	−1			
z_3	1	−1					
z_4			1	−1			
z_5					1	−1	−1
z_6						1	−1

Each step in the sequential binary partition divides the composition into separate parts (x_i and x_j): an important step in this process is the first partition, which defines the ratio for the first balance. Any balance is orthogonal to another balance as long as it is derived from the same sequential binary partition. The partition process applies some objectivity. Here we show a number of examples of how to develop suitable sequential binary partitions for use in CSG groundwater studies, based on geochemical intuition and the use of ilr dendrograms to

assess the effectiveness of particular ilr-coordinates. It does not matter what units the hydrochemical data are in [*Buccianti and Pawlowsky-Glahn*, 2005]: here we use meq/L. Once a sequential binary partition is described the *i*th ilr balance is computed as

$$\mathrm{zi} = \sqrt{\frac{r_i s_i}{r_i + s_i}} \ln \frac{\left(\prod_{+} x_j\right)^{\frac{1}{r_i}}}{\left(\prod_{-} x_j\right)^{\frac{1}{s_i}}}, \tag{4}$$

where r_i and s_i are the number of parts coded in the sequential binary partition as $+1$ and -1, respectively. The products subscripted $+$ and $-$ are extended to the parts x_j respectively marked with $+1$ and -1 in the sequential binary partition.

Isometric log ratios were calculated using *R*, and corresponding plots were produced using the "R" [*R development Core Team 2003*, 2015] package "ggplot2" [*Wickham*, 2009].

3.2.3. Data Centering for Use in Ternary Diagrams

The assessment of minor ion components can be challenging because of their relatively smaller concentration. One way of addressing this problem, when using ternary diagrams, is to center data to a common element (typically the inverse of the sample center) to better visualize data structure [*Pawlowsky-Glahn and Buccianti*, 2002; *von Eynatten et al*., 2002].

This data centering is performed using a perturbation operation (the symbol for which is $\oplus$); it is an operation that replaces addition and which acts as a scaling operator. If *x* and *y* are compositions, their perturbation is

$$\mathrm{x} \oplus \mathrm{y} = \mathrm{C}\,[x_1 y_1, x_2 y_2, \ \ldots., x_D y_D], \tag{5}$$

where C is the closure operation. The sample center (or compositional mean) of x is a composition computed as

$$\mathrm{cen(x)} = C\,[g(x_1), g(x_2), \ldots\ldots g(x_D)], \tag{6}$$

where $g(x_j)$ denotes the geometric mean of the values of the part x_j across the available sample. In order to center x, the composition y in equation (6) is identified with the opposite for the center then yielding the centered sample

$$\mathrm{C}\left[{}^{x_1}/_{g(x_1)}, {}^{x_2}/_{g(x_2)} \ldots\ldots {}^{x_D}/_{g(x_D)}\right]. \tag{7}$$

In this study, ternary diagrams representing subcompositional data are centered as in equation (8) and were computed using CodaPack 2.10 [*Comas-Cufí and Thió-Henestrosa*, 2011].

4. Results and Discussion

4.1. Characterizing Hydrochemistry of Coal Seam Gas Groundwater Using a Traditional Approach

Traditional Piper plots for data from gas-bearing aquifers and surrounding aquifers from the Powder River Basin and the Surat and Clarence-Moreton basins do not permit for hydrochemical discrepancies within and between these CSG groundwaters to be investigated, other than to allow visual interpretations (Figures 2a–2c). The sheer volume of data in both data sets creates a messy arrangement of overlapping data, making meaningful interpretations about hydrochemical evolution difficult.

The addition (or amalgamation) of Cl and SO_4 in Piper plots is a typical example of both the potentially spurious effects caused by these types of plots, as well as their inappropriateness for investigating CSG groundwater types. This amalgamation is counterintuitive for CSG groundwater because Cl and SO_4 ions behave very differently as CSG water types evolve. Similarly, Ca and Mg concentrations, relative to other ions, are difficult to interpret because all groundwaters are evolving to be dominated by Na cations. Since we already know that the dominant ions in CSG groundwater are Na, HCO_3, and Cl, we must ask how these plots improve our understanding of the hydrochemistry of CSG groundwater because these three ion components will always dominate the result/plot. There is potentially more to be learned about CSG groundwater evolution by understanding subtle changes in other ions, such as Ca, Mg, K, SO_4, and F.

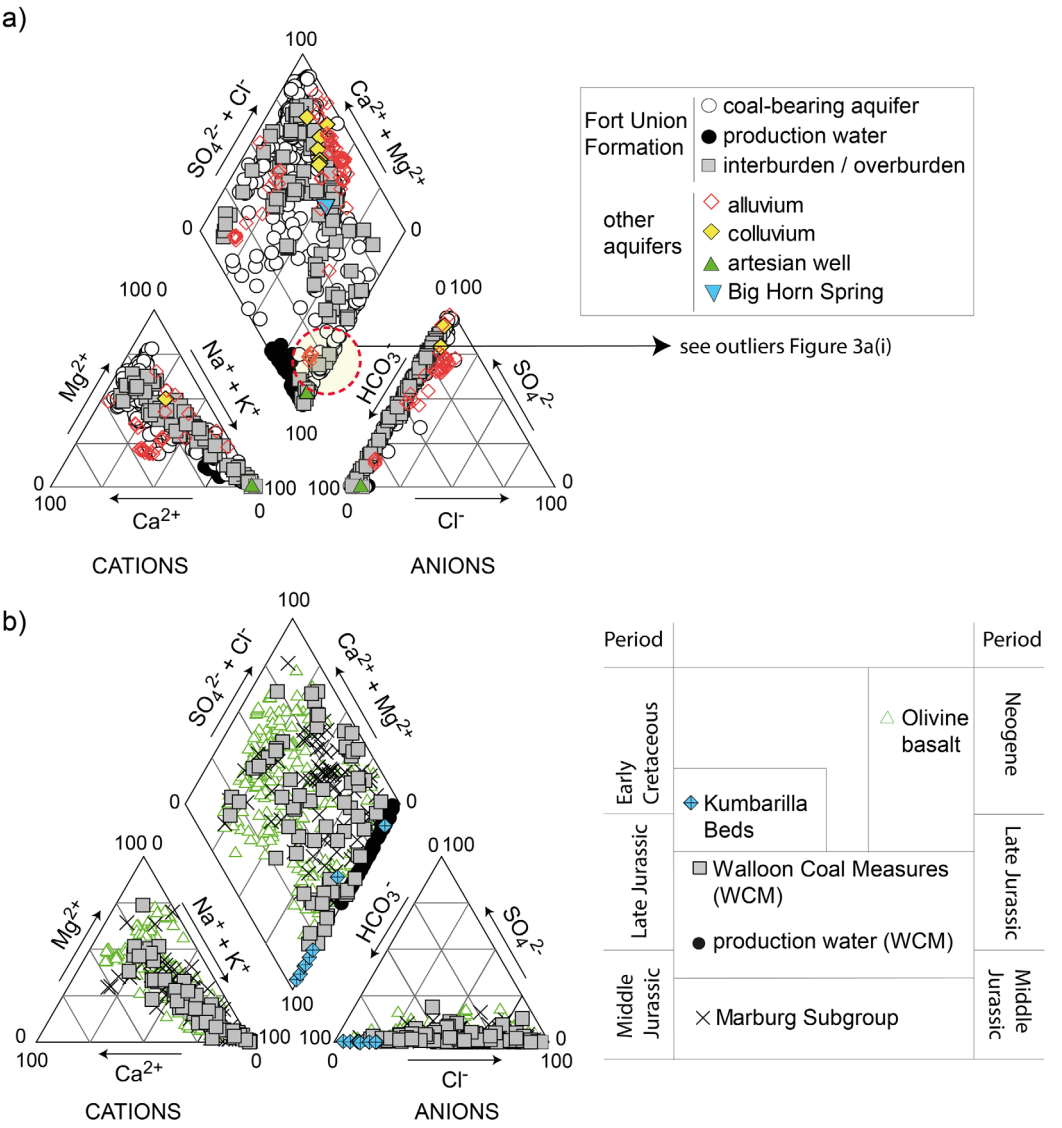

Figure 2. Traditional Piper plots for hydrochemical data for CSG production water and surrounding aquifers in (a) The Powder River Basin and (b) Surat and Clarence-Morton basins (Condamine catchment). Conceptualization of stratigraphic columns for the Condamine catchment modified from *Baublys et al*. [2015].

4.2. A Sulfate Reduction Pathway in the Powder River Basin Described by an Intuitive Set of Isometric Log Ratios

A simple way to test the influence of bacterial reduction of SO_4 on the evolution of CSG water types (Figure 1a) is to build a sequential binary partition and derive appropriate ilr-coordinates. The sequential binary partition for this exercise partitions Ca, Mg, SO_4 (the ions that are depleting), from Na, K, HCO_3 (the ions that are enriching) (Table 2). The use of Cl in the partition then allows the depletion of Ca, Mg, and SO_4, and subsequent enrichment of Na, K, and HCO_3 to be related to the relative enrichment of Cl down gradient. The first partition (b1) expresses the enrichment of Na and HCO_3 that is associated with SO_4 reduction and Ca and Mg depletion. The second partition (b2) captures the behavior of the third dominant ion component of CSG groundwater, Cl, with respect to the depleting ions. Essentially, this approach allows two ilr-coordinates to be derived; i.e., without the partition of Cl in b1 and b2, we could only derive a single ilr that partitions the enriching ions (Na and HCO_3) from the

Table 2. Sequential Binary Partition for ilr-Coordinates b1 and b2

Balance	Na	K	HCO3	Cl	Ca	Mg	SO4
b1	1	1	1	−1	−1	−1	−1
b2				1	−1	−1	−1

depleting ions (Ca, Mg, and SO_4) without any comprehension of how these changes pertain to more conservative ions, such as Cl.

If we apply these two partitions, b1 and b2, to equation (5), we derive two ilr-coordinates for each sample that can be plotted (Figure 3a). This ilr plot shows that most data in the Powder River Basin data set can be explained by this Ca and Mg depletion and SO_4 reduction (b1 and b2): the shape of this hydrochemical trend suggests a highly constrained system with two conspicuous end-members.

Using these ilr-coordinates (b1 and b2), a number of outliers occurs (Figure 3a (i)): these samples are already very low in SO_4, Ca, and Mg, and as a result the relationship between the Na, HCO_3 and Cl ions dominate compositional behavior. This relationship can be easily described using a different sequential binary partition that considers these three dominant ions only (Table 3).

These two processes can now be modeled using the appropriate ilr-coordinates; Figures 3b and 3d, respectively. However, note that because this sequential binary partition examines a different relationship with Cl, the sequential binary partition needs to be rebuilt, and these two models, therefore, are related by rotation, i.e., one can be obtained from the other through a rotation in the space of coordinates or, equivalently, in the simplex.

Overall, b1 and b2 can explain most of the hydrochemical variability in the Powder River data set: in this basin, *Van Voast*'s [2003] theory that SO_4 reduction drives the depletion of Ca and Mg and enrichment of HCO_3 as CSG water types evolve holds true (Figure 3a). This hydrochemical pathway reflects the dissolution of minerals such as gypsum and pyrite in shallow recharge areas, particularly in clinker deposits where groundwater is dominated by Ca and SO_4, but deeper groundwaters downdip, particularly in coal bed aquifers, are depleted in these ions and are dominated by Na and HCO_3 [*Bartos and Ogle*, 2002, and references therein]. This depletion of Ca and SO_4 and the evolution of Na-HCO_3 water types away from the basin margins are consistent with an increase in microbial methanogenic activity [*Bartos and Ogle*, 2002; *Flores et al*., 2008; *Quillinan and Frost*, 2014].

We propose that this quadratic model between ilr-coordinates b1 and b2 replaces the description of the CSG water type, i.e., a Na-HCO_3 water type, with a mathematical description that is both real and represents orthogonal coordinates in the simplex. Recall the variability of data in the Powder River data set as shown in the Piper plot in Figure 2a: there is no way of understanding this variability, let alone making any sense of it with respect to the evolution of CSG water types using this plot. By developing an appropriate sequential binary partition and associated ilr-coordinates, we have characterized a distinct hydrochemical pathway that explicitly describes the evolution of a particular hydrochemical end-member, despite a high degree of hydrochemical variability in the data set.

The outliers are also Na-HCO_3/Na-HCO_3-Cl water types that are associated with a low-Cl water source (Figure 3a (i)). These outliers may be related to areas of the coal-bearing aquifer where reduced conditions have already evolved or there is a limited SO_4 source, where there is vertical migration from overlying sandstone units, or where flow direction has changed due to anisotropy and cleat orientation in the coal seams [*Bartos and Ogle*, 2002]. We do not propose to resolve these hydrochemical phenomena here, but the application of compositional data analysis techniques demonstrates a robust method of understanding hydrochemical variability in large data sets with respect to the evolution of two very similar water types. These Na-HCO_3 water types can be compared to other similar water types using the modeled ilr-coordinates (Figures 3b–3d). Referring back to the Piper plot (Figure 2a), the outliers cannot be observed or differentiated from the CSG water types: in fact, there is no way of knowing that this other Na-HCO_3 water type exists.

A number of alluvial samples as well as two samples from artesian wells plot near production water samples (see (ii) in Figure 3a). These samples are compositionally distinct from other alluvial water types and their similarity to the coal-bearing aquifer suggests that these wells *may* have some relationship with the zones of the Fort Union Formation that are coal-bearing and contain gas. From the Piper plot in Figure 3a we can deduce that these samples have some hydrochemical similarity to the CSG groundwater types, but, without describing the SO_4 reduction pathway using ilr-coordinates in an appropriate linear model, there is no way of knowing if they are compositionally similar to those along the SO_4 reduction pathway, or another pathway, e.g., the outliers in (i). The well ID numbers of the outliers identified in Figure 3a (i) and (ii) are listed in Supporting Information.

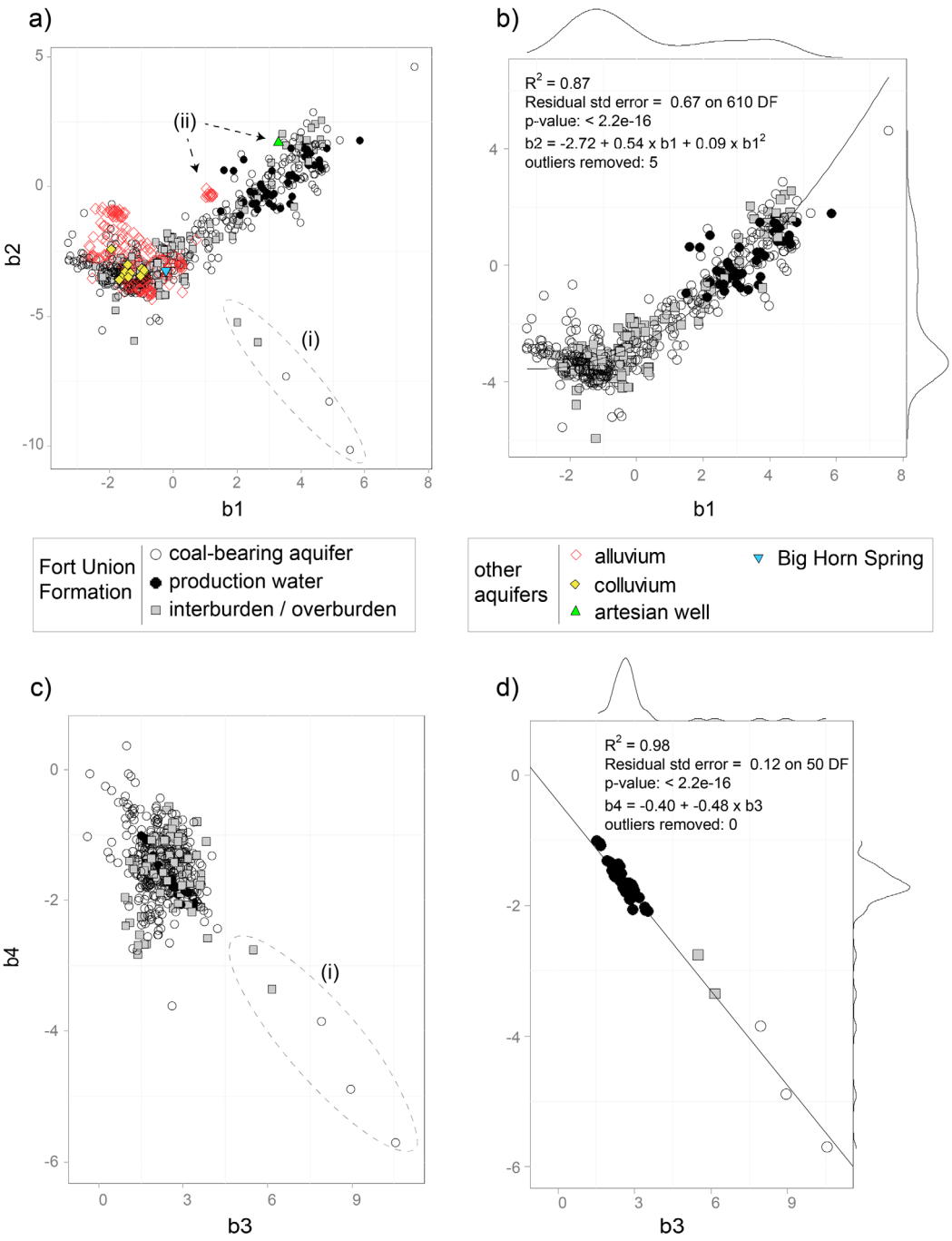

Figure 3. Isometric log ratios for Powder River Basin showing: (a) b1 versus b2 (SBP based on Table 2), with outliers in the Fort Union Formation circled (i) (see Supporting Information for well IDs), and alluvial/artesian wells water samples with compositional hydrochemistry similar to CSG groundwater marked as (ii) (see Supporting Information for well IDs); (b) quadratic model of b1 versus b2 for Fort Union Formation water (after removing (i) and alluvium, colluvium, artesian, and spring wells), with Kernel density estimations shown for *x* and *y* axis; (c) b3 versus b4 (SBP based on Table 3); and (d) linear model of b3 versus b4 for production water and Fort Union Formation water identified as outliers in Figure 3a (i), with Kernel density estimations shown for *x* and *y* axis.

Table 3. Sequential Binary Partition for ilr-Coordinates b3 and b4

Balance	Na	HCO3	Cl
b3	1		−1
b4	1	−1	1

4.3. Evidence of More Complex Pathways of CSG Groundwater Evolution in Other Basins

Using the b1 and b2 ilr-coordinates for the Surat and Clarence-Moreton basins, a similar trend toward CSG water types is evident, but with some data scatter, including some variability of production

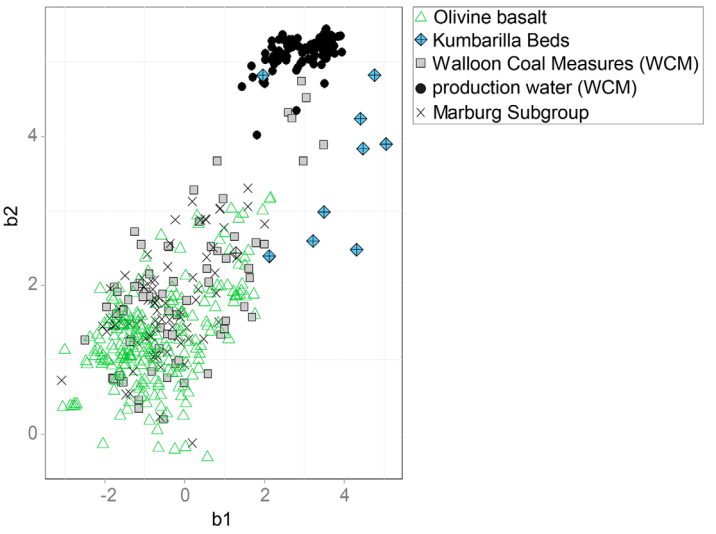

Figure 4. Isometric log ratios for b1 versus b2 (sequential binary partition as per Table 2) for the Surat and Clarence-Moreton basins (QLD, Australia).

(CSG) water (Figure 4). This scatter is caused by a range of different processes that influence the relative abundance of ions.

In this basin, the lithology of the sedimentary aquifers is similar, and carbonate dissolution and plagioclase weathering, in combination with cation exchange, are also likely to be contributing to the evolution of Na-HCO_3 water types. Unlike the Powder River Basin, there is not a significant source of SO_4 in the Surat and Clarence-Moreton basins. SO_4 reduction is occurring but it is not a major control on hydrochemistry. By better describing the compositional characteristics of the CSG end-member in the Surat and Clarence-Moreton basins, we can improve the ability to characterise groundwater within the coal measures and within other aquifers.

4.4. Delineating CSG Water Types in Large, Highly Variable Data Sets: A Case Study Example From the Surat and Clarence-Moreton Basins

4.4.1. Surat and Clarence-Moreton Basins (Condamine River and Teviot Brook Catchments): A Brief Overview

The Surat and Clarence-Moreton basins are two large, adjoining sedimentary basins in eastern Australia. For this study, we use two data sets: one from the Condamine River catchment (eastern Surat basin and north-western Clarence-Moreton Basin) and another from the Teviot Brook catchment (north-western Clarence-Moreton basin). The majority of the sedimentary formations in the upper Surat and Clarence-Moreton basins were deposited between the Late Triassic to Late Jurassic in a nonmarine, fluvial, and lacustrine environment, with some deposition extending into the Early Cretaceous [*Exon*, 1976; *Cook and Draper*, 2013; *Jell et al*., 2013]. The most important sedimentary units considered in this study are (listed as youngest to oldest): the Kumbarilla Beds (consisting of the Springbok Sandstone, Mooga Sandstone, Orallo Formation, Gubberamunda Sandstone, and Bungil Formation, herein undifferentiated), the Walloon Coal Measures (WCM), and the Marburg Subgroup (consisting of the Gatton Sandstone and the Koukandowie Formation, herein undifferentiated). These bedrock stratigraphic units in the study area have a variable composition, containing sandstones, siltstones, mudstones, and coal, with the most abundant coal seams found in the WCM [*Exon*, 1976; *Cadman et al*., 1998]. While some marine transgressions are inferred in the central-western parts of the Surat Basin during the early Cretaceous [*Cook and Draper*, 2013], these have not extended to the Condamine catchment or Teviot Brook catchment.

The Condamine River and Teviot Brook catchments share a boundary in the headwaters, where Neogene olivine basalt extrusions are dominant. Some of the sedimentary successions contained within the Surat and Clarence-Moreton basins are continuous across both these sedimentary basins, including the WCM and the Marburg Subgroup, with the WCM being a dominant shallow sedimentary formation in the headwaters of both catchments [*Cook and Draper*, 2013; *Duvert et al*., 2015]. Commercial gas production from reserves in the WCM is concentrated in the Condamine River catchment, although coal seam gas exploration has also previously occurred in the Teviot Brook catchment (production is not planned for the Teviot Brook catchment).

4.4.2. Understanding Variability Within and Between CSG Water Types

A quick way to explore the compositional variability in large data sets is to perform a centered-log ratio (clr) covariance biplot. The clr-biplot is a projection of the compositional space, with the links and rays providing

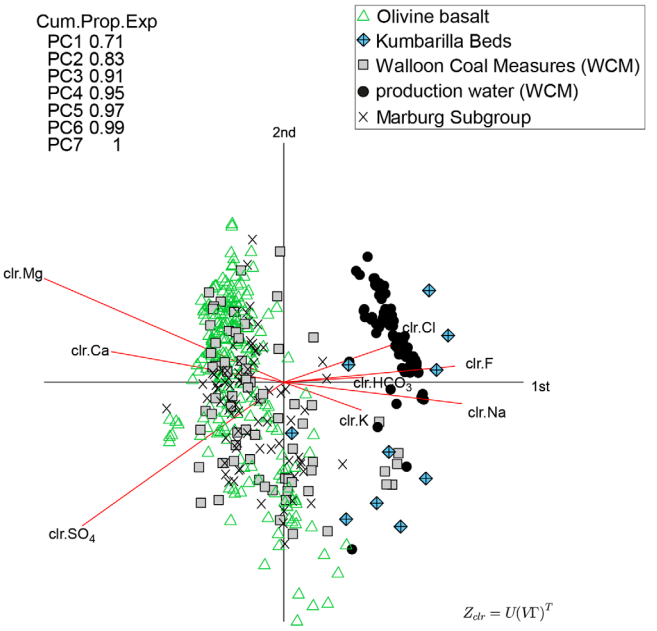

Figure 5. A clr-biplot of the Surat and Clarence-Moreton (Condamine River catchment) data set. Cum.Prop.Exp = the cumulative percentage of explained variance.

information about the variability of log ratios [*Aitchison and Greenacre*, 2002; *Pawlowsky-Glahn et al.*, 2015] (see section 3.2.1 for more detail).

Using a clr-biplot, it is evident that the majority of the compositional variability for the Condamine River catchment data set (Surat and Clarence-Moreton basins) can be explained by one component (PC1) (Figure 5). In combination with a second component (PC2), the CSG groundwaters represent a distinct end-member. Samples with positive PC1 and negative PC2 components generally have higher TDS, while fresher samples are those with negative PC1 values.

4.4.3. Categorizing Variability Using Hierarchical Cluster Analysis of ilr-Coordinates

A hierarchical cluster analysis (HCA) was applied to the data represented in ilr-coordinates derived from an intuitive sequential binary partition that partitions ions relative to the clr biplot (Table 4). Euclidean distance was used as the distance metric. This robust method of performing a HCA means the clustered groups represent compositional similarities in the simplex, rather than simply similarities in individual ion concentrations. We use it as a tool to describe different groups of variable compositional structure that can then be used to explore different hypotheses using intuitive and appropriate ilr-coordinates.

4.4.4. Major Ion Controls on CSG Groundwater (Surat and Clarence-Moreton Basins)

A clr-variation array confirms the major controls on the CSG groundwater from the Surat and Clarence Moreton basins are reflected by variability of Ca, Mg, F, Cl, and HCO_3 (Table 5). In contrast, variability in the Powder River CSG groundwater is relatively low (Table 5).

In the Surat and Clarence-Moreton basin the relationships between these ions can be easily modeled in ternary diagrams (Figures 6a and 6c). Substituting Mg for Ca, similar relationships were also observed. The modeled variability in the data set coincides with a shift between two subclusters identified through the ilr HCA (CSG clusters A1 and A2).

The contrasts between the Surat/Clarence-Moreton basin and Powder River data shown here are peculiar, especially considering coal type is similar and the dominant methanogenic pathway is the same. In the Powder River Basin, the relatively higher Ca and SO_4, derived from shallower sources, may create a more pronounced control on hydrochemistry, with significant SO_4 reduction having a pronounced effect on HCO_3. In the Surat and Clarence-Moreton basins, relationships between Cl and other ions may reflect more complex influences on hydrochemistry (see section 4.5.3 for more discussion).

Table 4. Sequential Binary Partition Used to Describe ilr-Coordinates for the Hierarchical Cluster Analysis

Na	Ca	Mg	K	HCO3	Cl	F	SO4
1	−1	−1	1	1	−1	1	−1
1			1	−1		−1	
1			−1				
				1		−1	
	1	1			−1		−1
	1	−1					
					1		−1

Using a traditional approach, both of these CSG end-members in the Powder River and Surat/Clarence-Moreton basins would have the same classification: a Na-HCO_3 or Na-HCO_3-Cl water type, yet important between-basin distinctions cannot be described. For example, concentrations of Ca and Mg ions in the Surat and Clarence-Moreton CSG are too low to be observed on the Piper plot, yet these ions have a major control on compositional variability in this basin. The models in Figures 6a and 6b condense the compositional variability into a descriptive tool for identifying CSG

Table 5. The clr-Variation Array for CSG Groundwater From the Surat and Clarence-Moreton Basins (a) and the Powder River Basin (b); These Arrays Represent the Log-Transformed Ratios, With the Sample Mean $\ln(x_i/x_j)$ Shown Below the Diagonal, and Estimates of var[-$\ln(x_i/x_j)$] Shown Above the Diagonal, and the clr-Variances Representing the Sum of Log Ratio Variances for Each Part [*Thió-Henestrosa and Comas*, 2011]*

a) CSG Groundwater: Surat and Clarence-Moreton Basins

		Variance $\ln(x_i\backslash x_j)$								
	$x_i\backslash x_j$	Na	Ca	Mg	K	HCO_3	Cl	F	SO_4	clr Variances
Mean $\ln(x_i/x_j)$	Na		0.29	0.27	0.03	0.34	0.05	0.22	0.29	0.01
	Ca	−5.19		0.06	0.26	1.17	0.14	0.95	0.74	0.28
	Mg	−5.47	−0.27		0.25	1.13	0.12	0.93	0.69	0.25
	K	−6.02	−0.82	−0.55		0.39	0.07	0.27	0.31	0.02
	HCO_3	−1.37	3.82	4.10	4.65		0.62	0.05	0.42	0.34
	Cl	−0.32	4.87	5.14	5.69	1.05		0.45	0.40	0.05
	F	−6.62	−1.43	−1.15	−0.60	−5.25	−6.29		0.34	0.23
	SO_4	−8.09	−2.90	−2.63	−2.07	−6.72	−7.77	−1.47		0.22
						Total variance				1.40

b) Powder River Basin

		Variance $\ln(x_i\backslash x_j)$								
	$x_i\backslash x_j$	Na	K	Ca	Mg	HCO_3	Cl	F	SO_4	clr Variances
Mean $\ln(x_i/x_j)$	Na		0.26	0.38	0.34	0.45	0.36	0.60	4.61	0.19
	K	−4.03		0.16	0.13	0.69	0.39	0.25	5.25	0.21
	Ca	−2.05	1.99		0.07	0.77	0.38	0.49	4.55	0.17
	Mg	−2.30	1.74	−0.25		0.73	0.47	0.56	4.79	0.20
	HCO_3	0.10	4.14	2.15	2.40		0.89	1.11	4.39	0.44
	Cl	−3.60	0.43	−1.55	−1.30	−3.70		0.49	4.87	0.30
	F	−5.52	−1.48	−3.47	−3.22	−5.62	−1.92		5.40	0.43
	SO_4	−7.10	−3.07	−5.05	−4.81	−7.21	−3.50	−1.59		3.55
						Total variance				5.48

*Caution needs to be applied to results for SO_4 because CSG groundwater is typically highly reduced and values below detection limit represent imputed values.

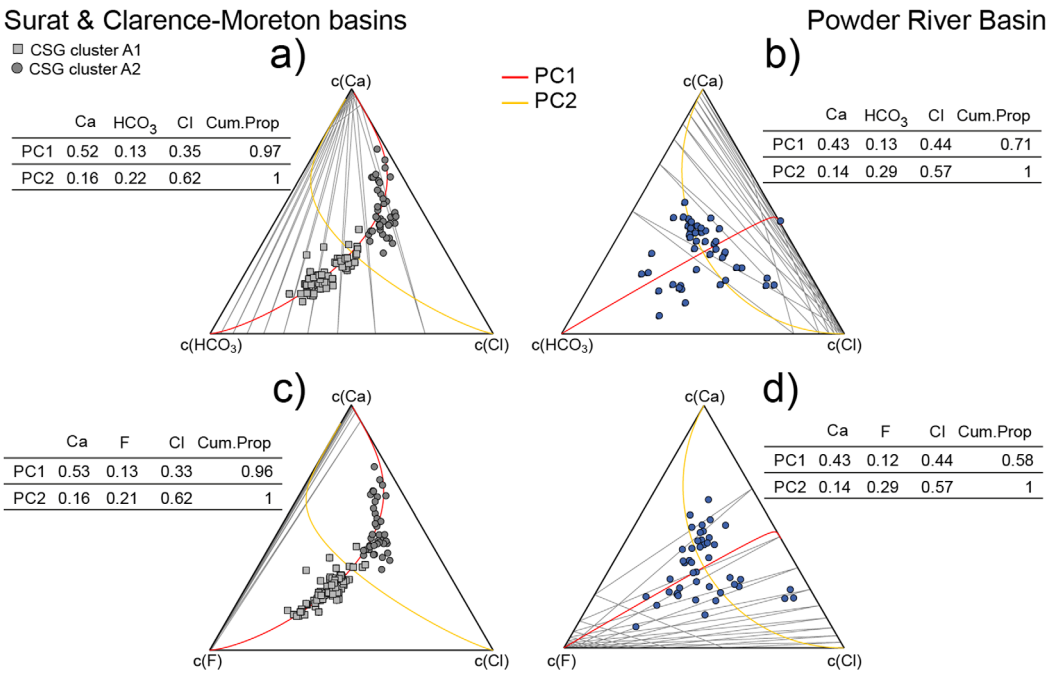

Figure 6. Ternary models for CSG groundwater (a, c) from the Surat and Clarence-Moreton basins (Condamine River catchment) and (b, d) from the Powder River basin, showing the principal components for each model. Samples in the ternary diagram represent compositions that are perturbed by the closed inverse of the geometric mean of each sample. Cum.Prop = cumulative proportion of explained variance.

groundwater types that is unique to this basin. This new approach to describing hydrochemistry of CSG end-members opens up the door to a different philosophy for characterizing these water types based on the relationships between the relative proportion of certain ions.

4.4.5. Compositional Distinctions Between CSG Groundwater and Similar Na-HCO_3 Water Types

The HCA results showed that a number of groundwater samples from the overlying Kumbarilla Beds and the shallower parts of the coal measures are also dominated by Na and HCO_3 ions and exhibit some compositional similarities to CSG groundwater. We can investigate these similarities further using ilr-coordinates that describe relationships between Ca, Mg, F, and HCO_3, and Ca, Mg, F, HCO_3, and Cl, respectively (Table 6), which were modeled in the ternary diagrams (Figures 6a and 6c). The positive linear relationship between b5 and b6 values shows that, for these samples, increases in the Ca and Mg components with respect to F and HCO_3 have a major control on compositional variability. The result shows that these Na-HCO_3 water types may be evolving rapidly in shallower zones, without a precursor water type with higher Ca and Mg components.

A key difference between the samples from the Kumbarilla Beds, coal measures and CSG aquifers in Figure 7a is the lower F concentrations in the CSG groundwater, causing a shift toward higher b6 values. Similar lithology between the Kumbarilla Beds and the coal measures may allow similar water types to evolve independently, with varying concentrations of F depending on the source rock.

If we rebuild the sequential binary partition and include Cl (Table 6 and Figure 7b), we can see that the relationship observed in Figure 7a is maintained for CSG groundwater but this is not the case for other coal measure samples or those from the Kumbarilla Beds. This distinction shows that the controls on Cl in the CSG groundwater are different to surrounding aquifers and up-gradient in the coal measures, even where very similar hydrochemical compositions occur. This is an important distinction to make because the concentrations of Ca, Mg, and F are only minor components of these water types, and the compositional differences shown in Figure 7b cannot be observed using conventional techniques (Figure 2b). The relationships between Cl in the CSG groundwater and its distinction from the controls on Cl in the overlying Kumbarilla Beds suggest that there is a source of Cl in the coal measures that is associated with high gas concentrations which may not be related to overlying aquifers.

4.4.6. Developing a New ilr Descriptor for Na-HCO_3 Water Types in the Kumbarilla Beds

Here we use ilr dendrograms [*Pawlowsky-Glahn and Egozcue*, 2011] to identify subtle compositional traits that are unique to the Na-HCO_3 water type in the Kumbarilla Beds.

This process calculates the ilr-coordinates for each aquifer based on a common sequential binary partition. The development of this SBP requires some trial and error, but suitable sequential binary partitions can be developed intuitively by referring back to the clr-biplot (Figure 5) and using information from other plots, such as the ternary plots and ilr plots in Figures 6 and 7. From the clr-biplot it is evident that Na-HCO_3 groundwater in the Kumbarilla Beds is explained by K and Na, and a possible inverse relationship with Cl. Table 7 describes the sequential binary partition used to construct an intuitive ilr dendrogram. This sequential binary partition also considers the potential relationships with other ions, but SO_4 is excluded: we found that the inclusion of SO_4 masked relationships between other ions in these water types because SO_4 concentrations are very low, but with high variance.

Figure 8a shows the ilr dendrogram derived from the sequential binary partition in Table 7. Clearly, the Na-HCO_3 water type in the Kumbarilla Beds can be distinguished from CSG groundwater and other similar water types in shallower areas of the coal measures by the behaviour of K, Na, and Cl, which follows a linear model described by the second and third balances in the sequential binary partition (b10 and b11) (Figure 8b).

Table 6. Sequential Binary Partition for ilr-Coordinates b5 and b6 and b7 and b8, Respectively

ilr-Coordinates b5 and b6					
Balance	HCO_3	Ca	Mg	F	
b5	−1	1	1	1	
b6		1	1	−1	
ilr-Coordinates b7 and b8					
Balance	HCO_3	Cl	Ca	Mg	F
b7	−1	−1	1	1	1
b8			1	1	−1

Given that differentiating between the Kumbarilla Beds and underlying coal measures has proved to be difficult in this basin, due to similar lithology, a subtle but descriptive method for distinguishing between similar water types between these aquifers will be useful.

In Figures 7 and 8 one sample (well ID = 22194) from the Kumbarilla Beds consistently plots near the CSG water samples, including along the ilr trend that

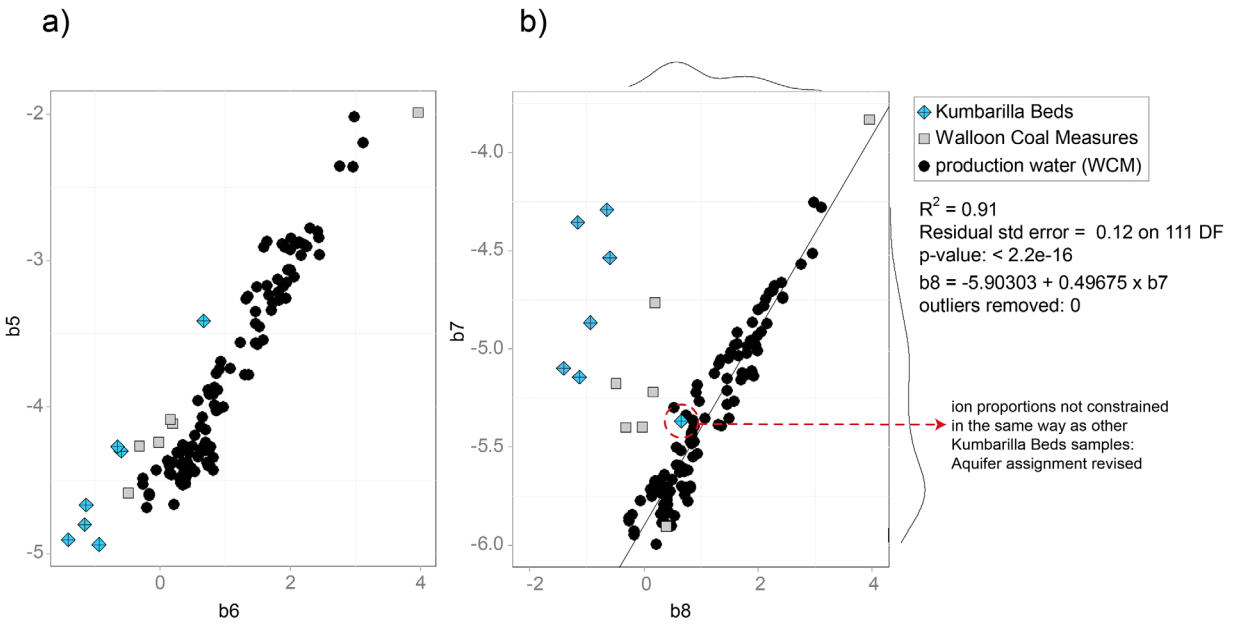

Figure 7. Isometric log ratios for similar Na-HCO_3 water types from the Surat and Clarence-Moreton basins (cluster A) showing: (a) b5 versus b6 (SBP based on Table 6), and (b) b7 versus b8 (SBP based on Table 6) showing a linear model of b7 versus b8 for production water, with Kernel density estimations shown for *x* and *y* axes.

describes the constraint of ion proportions in CSG water samples. As a result, the aquifer assignment for well 22194 of Kumbarilla Beds can be revised and reassigned to the coal measures. Review of the drill logs for these samples shows that this well occurs in the same area as the shallower well 11797 which was assigned to the coal measures. Note this exercise is not possible using conventional plots such as Piper plots.

4.4.7. A Unique, Robust Descriptor That Simplifies Complex Processes

Unique descriptions of water types using ilr-coordinates are much more effective tools for describing hydrochemical variability than descriptions of the dominant ion components because they capture a range of complex processes. Some inverse modeling (using PHREEQC version 3.1.7-9213) for the Kumbarilla Beds samples demonstrates this point. For simplicity we performed inverse geochemical modeling at three steps along this trend line (increasing b10 values) (see Figure 8b). This generally follows a process which is associated with the depletion of Cl with depth. Typical minerals used in the inverse modeling scenarios were calcite, dolomite, gypsum, fluorite, kaolinite, K-feldspar, gibbsite, albite, biotite, as well as cation exchange sites. The selection of minerals is based on the known mineralogy of the sedimentary features in the Surat Basin [*Grigorescu*, 2011]. See Supporting Information for inverse modeling results.

Inverse modeling results indicate that dissolution of albite, K-feldspar and fluorite as well as cation exchange are typical for all steps along this trend. For step 1 (highest b10 values) and step 3 (lowest b10 values), gypsum and fluorite dissolution are typical. However, for step 2, gypsum precipitation occurred and no phase mole transfers were recorded for fluorite. This shows some complexity on the relative proportions of ions for this Na-HCO_3 water type even though the dominant ion composition remains the same. This highlights the value of the ilr approach. Instead of interpreting a singular process from ion ratios, e.g., carbonate dissolution, the ilr-coordinates b10 and b11 capture real restrictions on the relative proportions of Na, K, and Cl ions that occur during a range of processes. These characteristics of the relative proportion of ions are unique to this aquifer. Given that K tends to be re-incorporated into clay minerals, and Na remains in solution [*Hem*, 1985], these characteristics must be due to cation exchange and albite weathering which promotes Na release, which, in turn, increases the proportion of Na to both K and Cl.

Table 7. Sequential Binary Partition Used to Describe Isometric Log Ratio Balances for the Hierarchical Cluster Analysis

Balance	Na	Ca	Mg	K	HCO_3	Cl	F
b9	1	−1	−1	1	−1	1	−1
b10	1			−1		1	
b11	1					−1	
b12		1	1		−1		−1
b13		1	−1				
b14					−1		1

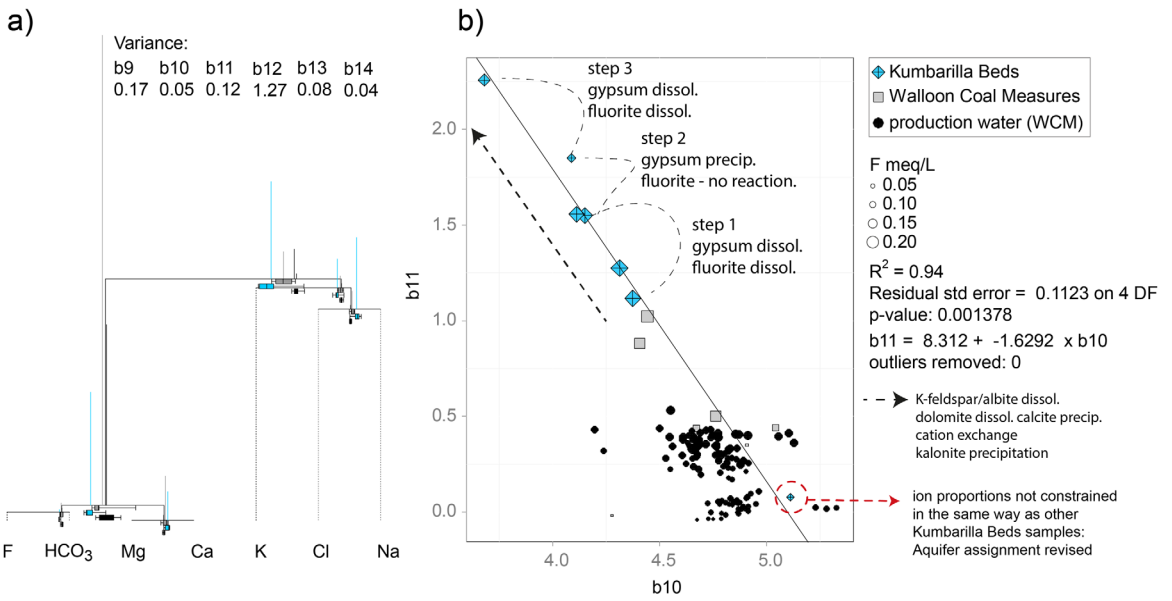

Figure 8. (a) A balance dendrogram showing distinctions in the isometric log ratios as defined in the table (b9–b14) between the Kumbarilla Beds, the coal measures where gas is not prevalent (the Walloon Coal Measures), and the coal measures where high concentrations of gas occur (production water (WCM)); and (b) comparison of isometric log ratios b11 and b10, showing a linear relationship for samples from the Kumbarilla Beds. Dashed arrow represents a general depletion of Cl with depth. Dashed lines represent a step along the linear trend used in inverse modeling scenarios. The linear model in Figure 6b could be improved by collecting more data from this aquifer.

The ilr-coordinates used in Figures 7b and 8b are useful for differentiating between CSG and other similar water types, but a small number of samples plot near CSG groundwater in these plots. It is not clear if this is the result of interactions with CSG water types, the influence of coal and/or methanogenesis, or other factors. In any case these sites are worthy of further investigation. To assist future studies, the wells where samples show the greatest similarity with CSG groundwater in Figures 7b and 8b are three wells in the coal measures with b11 values >0.25 but <0.75 (well ID: 5390, 16625, and 83519), as well as one well with b11 value <0.25, from the coal measures (well ID: 11797). Interestingly, drill logs for wells 5390, 16625, and 83519 show coal occurs at the screen depth, although spatial preference or relationship with depth were not observed. The remaining well (11797) occurs in the west of the catchment at $\sim$200 m depth.

4.5. A New Method for Describing Hydrochemical Variability Within and Between CSG Water Types Using Isometric Log Ratios

4.5.1. Defining an Appropriate Sequential Binary Partition and Orthonormal Bases

One advantage of the compositional approach is that, instead of deliberating over large data sets and attempting to explain all processes that influence hydrochemistry, we can describe suitable ilr-coordinates that target specific water types (distinct characteristics of the relative proportions of ions). A key objective is to avoid getting bogged down in differences between groups/water types that distract the investigator: here we are only interested in the evolution of CSG water types.

From earlier sections we have successfully shown how to better describe and differentiate between similar water types using ilr-coordinates. The aim of this section is to present a new plot that is capable of delineating CSG groundwater types, and of elucidating groundwaters that are compositional similar to CSG groundwater in large data sets. In doing so, the plot needs to be able to both describe compositional variability within CSG groundwater and it needs to be geochemically intuitive, in order to assist with interpretation.

To achieve this, we developed a sequential binary partition (in Table 8) that draws on the work by *Duvert et al*. [2015] where relationships between residual alkalinity and chloride were used in combination with isotopic data to assess connectivity between the coal measures and surrounding aquifers in the Teviot Brook catchment of the Clarence-Moreton Basin, QLD, Australia. Residual alkalinity describes the residual bicarbonate and carbonate ions that occur in solution in excess to any possible dissolution of carbonate precipitates, and is defined as

Table 8. Sequential Binary Partition for ilr-Coordinates b15 and b16

Balance	Ca	Mg	HCO3	Cl	SO4
b15	−1	−1	1		
b16				1	−1

$$(HCO_3 + CO_3) - (Ca + Mg), \quad (8)$$

where units are meq/L.

The properties of residual alkalinity indicate that Na-HCO_3 water types will always have positive residual alkalinity values. However, the amalgamation and subtraction in the residual alkalinity calculation do not preserve distances within the simplex, and thus may have an unwanted, negative effect on multivariate analysis and may lead to erroneous interpretations of hydrochemical similarity [*Egozcue and Pawlowsky-Glahn*, 2005; *Engle and Rowan*, 2013]. An ilr which acts as an alternative to the conventional residual alkalinity (b15) can be developed using an appropriate sequential binary partition (Table 8). This ilr is compared with another ilr between Cl and SO_4 (b16) to describe the relationships between high HCO_3 groundwater and increases in Cl in SO_4-depleting environments that are typical of CSG water types. Figure 9 compares these two ilr-coordinates for data from the Teviot Brook (Clarence-Moreton Basin) catchment. High HCO_3 and low Ca and Mg groundwater have a b15 ≥ 1. The ilr plot is similar to that described in the residual alkalinity versus Cl plot by *Duvert et al*. [2015], but the variability of these high HCO_3 groundwaters is better elucidated when expressed as an ilr-coordinate. The relationship between high HCO_3 groundwaters (b15 >1) and increasing Cl in combination with reduction of SO_4 (b16) is orthogonal and can be modeled using an exponential equation. Increases in δ^{13}C-DIC along this modeled trend suggest an influence of methanogenesis, and samples with high b15 and b16 values showing compositional similarity to CSG groundwater from the adjoining Condamine catchment (Surat and Clarence-Moreton basins) (see shaded area in Figure 9).

Although historical isotope data are not always available for preliminary hydrochemical investigations, the use of δ^{13}C-DIC here demonstrates that the ilr-coordinates (for the modeled trend) are capturing changes in the relative proportion of ions associated with methanogenesis. In subsequent studies, the delta values of stable isotopes, which already represent a type of log ratio, can also be compared to ilr-coordinates to make further assessments. This is out of the scope of this paper, but a recent example of this application to a different hydrogeological problem is shown in *Engle et al*. [2016].

4.5.2. Applying This New Set of Isometric Log Ratios to Delineate CSG Groundwater in a Complex Hydrogeological Setting

Figure 10 compares the conventional residual alkalinity versus Cl plot (Figure 10a) with two orthonormal ilr-coordinates (Figure 10b) described in Table 8 to the Condamine River catchment (Surat and Clarence-

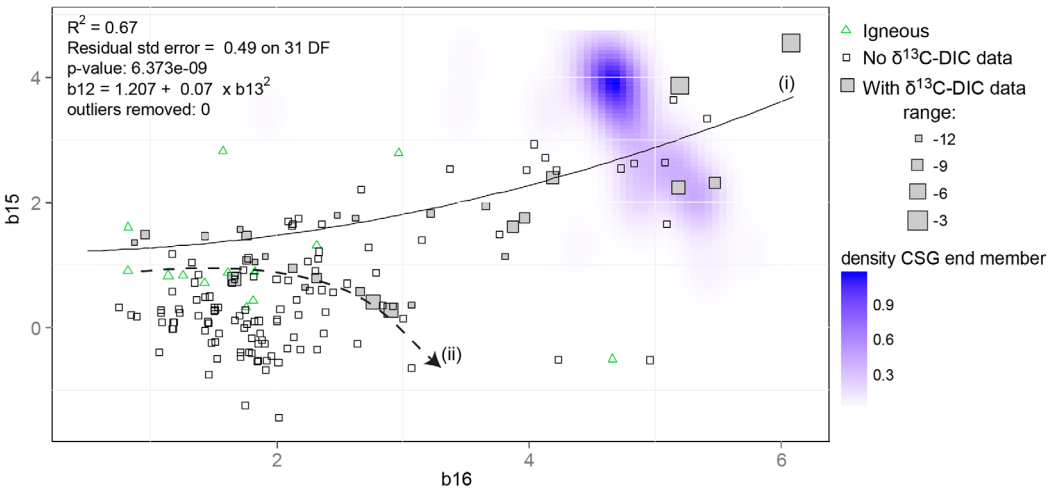

Figure 9. Isometric log ratios for groundwater samples from the Teviot Brook catchment (Clarence-Moreton Basin) based on a sequential binary partition that is intuitively derived from the residual alkalinity versus Cl plot as per *Duvert et al*. [2015, Figure 5a], and partitions HCO_3, and Ca and Mg (b15) and Cl and SO_4 (b16). Plot shows data from *Duvert et al*. [2015] (gray shaded squares) and Li (J.L. unpublished data, 2001) (unshaded squares), with: (i) a potential CSG groundwater pathway modeled using an exponential equation and (ii) a conceptual evapotranspiration pathway as proposed by *Duvert et al*. [2015]. The shaded area represents the density (abundance) of ilr values for CSG groundwater samples from the adjoining Condamine pathway River catchment (Surat and Clarence-Moreton basins). Increases in δ^{13}C-DIC along the modeled CSG pathway (i) are indicative of an influence of methanogenesis. Igneous data are from *Duvert et al*. [2015] and Li (unpublished data, 2001).

Moreton basins). Using the conventional plot (Figure 10a) the high HCO_3 of the CSG water is obvious from the high residual alkalinity values; however, in this plot the CSG end-member shows highly variable HCO_3 and Cl concentrations. In addition, many of the samples from Kumbarilla Beds appear as intermediate, between some coal measure samples and the CSG end-member, but using ilr-coordinates described in Figures 7 and 8, these water types were shown to be different.

In contrast, the ilr-coordinates compared in Figure 10b elucidate the compositional variability of the samples from this data set. A clear observation from the ilr plot in Figure 10b is that a distinct pathway in the coal measures toward CSG water types was not observed; however, two groups of coal measures samples with compositional similarities to CSG water types are obvious: (ii) shows compositional similarity with lower salinity CSG water types that are characterized by high HCO_3 and F, and low Ca and Mg; and (iii) coal measure samples with low HCO_3, relatively high Cl, and remarkably relatively higher Ca, Mg and SO_4. This variability in the CSG data shown in Figure 10b is synonymous with the variability modeled in Figures 6 and 7.

Inverse modeling results indicate that compositional changes along a hypothetical pathway from (i) to (ii), either from basalt aquifers or parts of the coal measures, can be explained by a range of common mineralogical processes, including dissolution of dolomite, fluorite, albite, and K-feldspars, precipitation of gypsum and kaolinite and some calcite, as well as cation exchange. This pathway also requires a source of salt to balance the NaCl in the phase mole transfer reaction. Similarly, the inverse modeling results for a pathway from (ii) to CSG groundwater with high HCO_3 (described as cluster A1 in Figure 6a and c) also require a

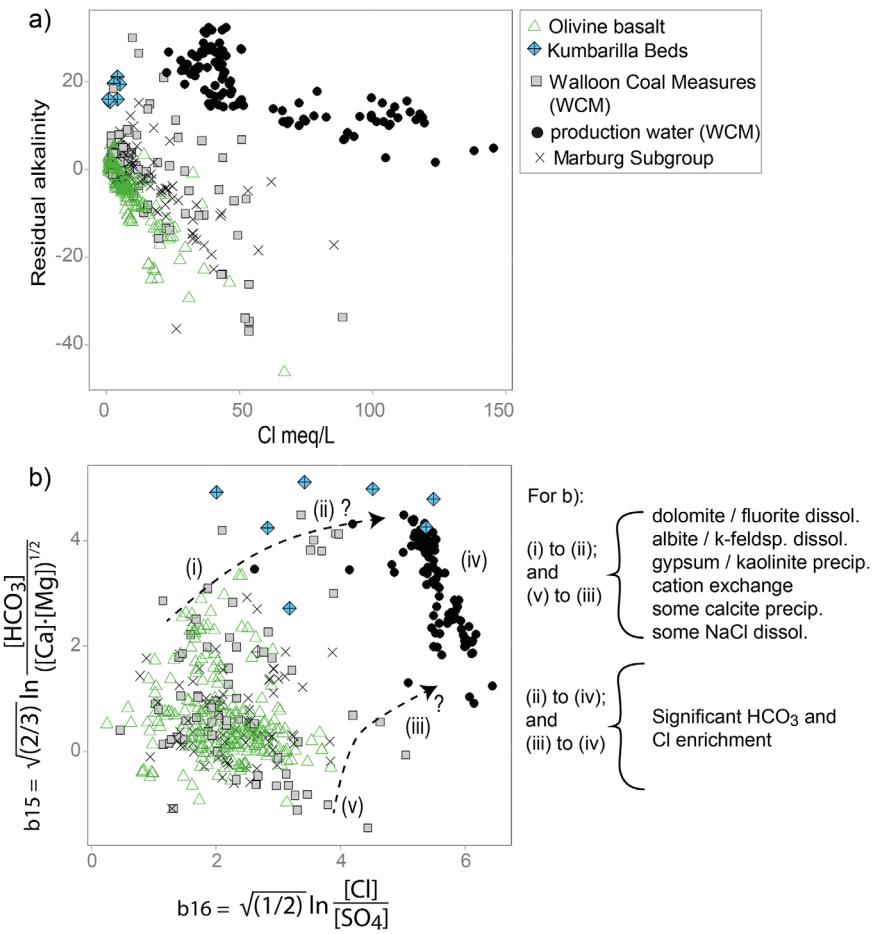

Figure 10. Data for the Surat and Clarence-Moreton basins, showing (a) residual alkalinity versus Cl, where residual alkalinity = HCO_3 − (Ca + Mg) in meq/L; and (b) isometric log ratios that partition HCO_3, and Ca and Mg (b15) and Cl and SO_4 (b16). Notation for (b): (i) = Na-HCO_3 low-salinity groundwater; (ii) = Na-HCO_3 low-salinity groundwater with compositional similarity to lower salinity CSG end-member (iv); (iii) = higher-salinity (brackish) groundwater with compositional similarity to higher-salinity CSG end-member (iv); (v) brackish groundwater low in HCO_3 with some Ca and SO_4; (iv) the CSG end-member, ranging from lower salinity (high HCO_3/b15) to higher-salinity (lower HCO_3, higher Cl/lower b15). Inverse modeling results are shown in Supporting Information.

significant addition of NaCl to balance the phase mole transfer. This shows that, even for a compositional shift from a high HCO_3 water type toward a high HCO_3 CSG end-member (i.e., CSG samples with high b15 values), there must be a significant increase in Cl. (see Supporting Information for inverse modeling results).

A subcluster of coal measure samples shows a general trend of decreasing b15 with increasing b16 values (v). The samples in (iii) are from the same subcluster, yet the b15 values are much higher: they represent a possible intermediate between groundwater that is relatively depleted in HCO_3 and higher salinity CSG water samples.

4.5.3. A Compositional Description of Groundwater Associated With Coal Seam Gas

While the compositional traits of the CSG end-member in the Condamine catchment are associated with an increase in the relative proportion of Cl, there are no brines, nor have there been any marine incursions in this part of the basin [*Cook and Draper*, 2013; *Jell et al*., 2013]. *Baublys et al*. [2015] suggested the higher Cl in some CSG groundwater may be because shallower areas of the subcrop had been influenced by recharge from more saline sources, such as from the overlying quaternary alluvium. However, a recent geochemical assessment of the alluvium does not indicate that deep alluvial groundwater is source of high Cl or Ca [*Owen and Cox*, 2015]. Similarly, the shallower zones of the coal measures do not appear as a source of high Cl either. Some results for shallower areas of the coal measures presented here also do not suggest compositional similarities with the brackish CSG end-member are necessarily related to depth. For example, one brackish coal measure sample from (iii) (Figure 10b), which is compositionally similar to the brackish CSG end-member, occurs in a shallow (∼40 m) outcrop of the coal measures in the north east of the catchment ranges, adjacent to a basalt outcrop, where recharge is expected to be relatively high. Another coal measure sample nearby, in the same outcrop and at a similar depth, was found to have b16 ∼2 and b15 ∼3: this sample is indicative of the fresher Na-HCO_3 groundwater (Figure 10b (i)). This shows large compositional variability over small distances that are not related to depth or a spatial flow path.

Baublys et al. [2015] and *Owen and Cox* [2015] also offered a number of other theories to explain the relatively higher Cl in some CSG groundwater in the Surat Basin, being the influences of variability in (a) coal permeability; (b) mixed layer clays; and (c) organic-bound Cl in the coal matrix. Spatial changes in these attributes may influence the retention of Cl in coals over small distances.

We do not propose to resolve the sources of Cl in the CSG end-member in this paper. However, what we show here is that the variability in the relative proportion of Cl is inherently related to a distinct compositional characteristic of the CSG end-member. We tested the effectiveness of the ilr-coordinates b15 and b16 (Table 8) to delineate groundwater associated with high gas concentrations via some additional sampling in both the deep CSG reservoir and the shallow coal measures in the Condamine catchment (Figure 11). The high-gas groundwater from the shallower coal measures (shaded squares) showed the

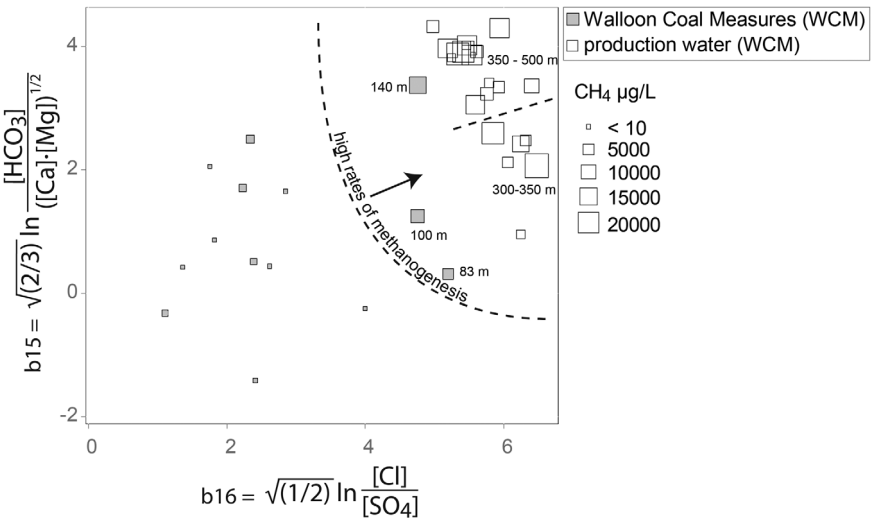

Figure 11. Additional (new) data for the Surat and Clarence-Moreton basins, showing isometric log ratios that partition HCO_3, and Ca and Mg (b15) and Cl and SO_4 (b16) for the CSG production water (WCM: 300–500 m) and shallower areas of the coal measures (WCM: <200 m). Labels in the plots represent the depth of the screens within the wells. CH_4 data (point scaling) represent dissolved gas concentrations.

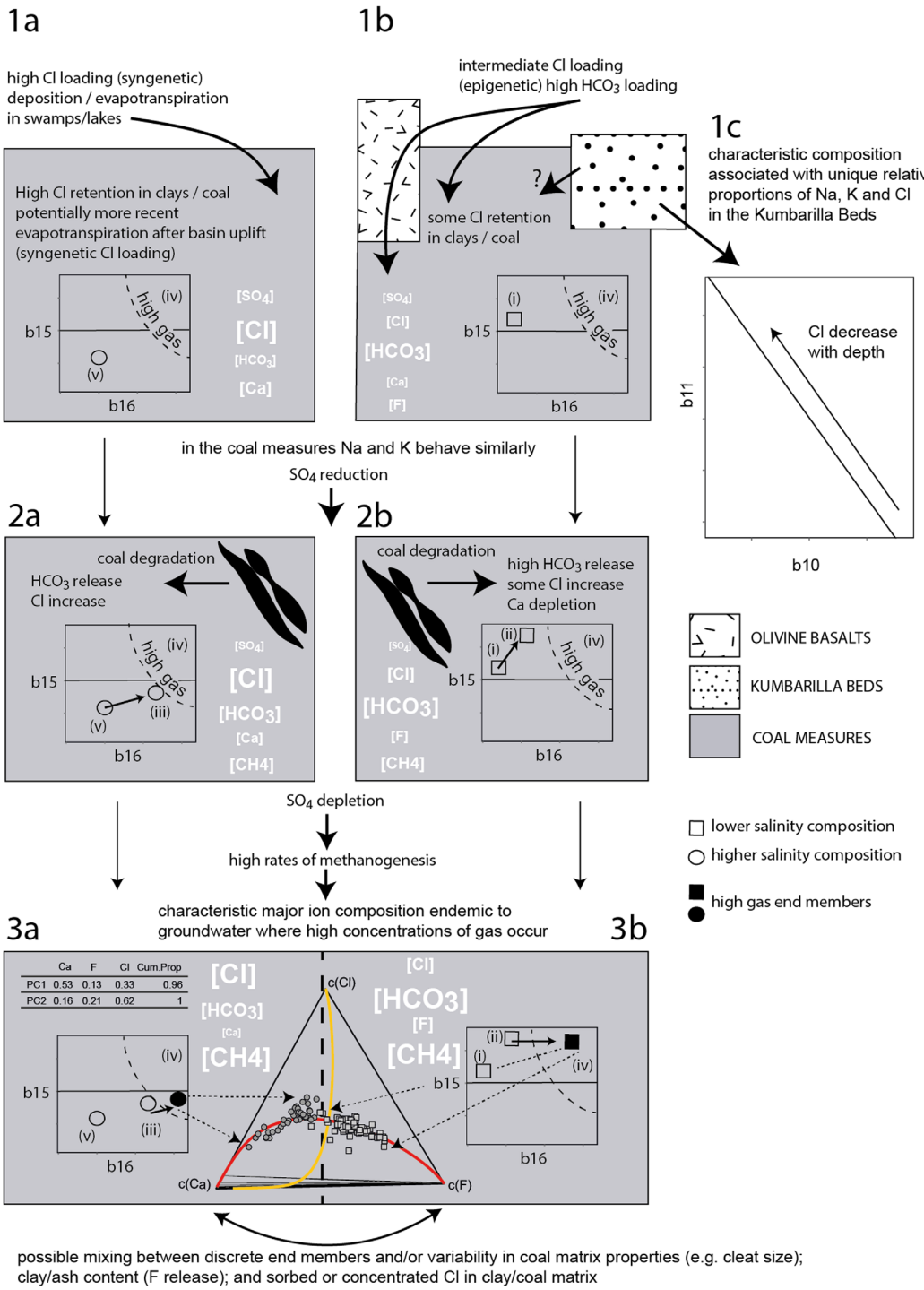

Figure 12. Conceptual process model of changes in the relative proportion of ions required for the characteristic hydrochemical composition of CSG groundwater in the Surat/Clarence-Moreton Basins to evolve. Notations (i)–(v) are as per Figure 10b. Elements in parentheses (white text) represent the relative concentration. The ilr-coordinates b15 and b16 represent the partition of HCO_3 and Ca and Mg, and Cl and SO_4, respectively (as per Table 8). The ilr-coordinates b10 and b11 represent the partition of Na, K, and Cl, and Na and Cl, respectively, as per Table 7. Possible Cl sources as per *Baublys et al*. [2015] and *Owen et al*. [2015, and references therein].

same compositional variability despite being in a much shallower zone (83–140 m). Peculiarly, the shallow coal measure sample (140 m) with the highest gas concentrations shows compositional similarities with CSG production water from depths of around 400–500 m (high b15 values). This adds some weight

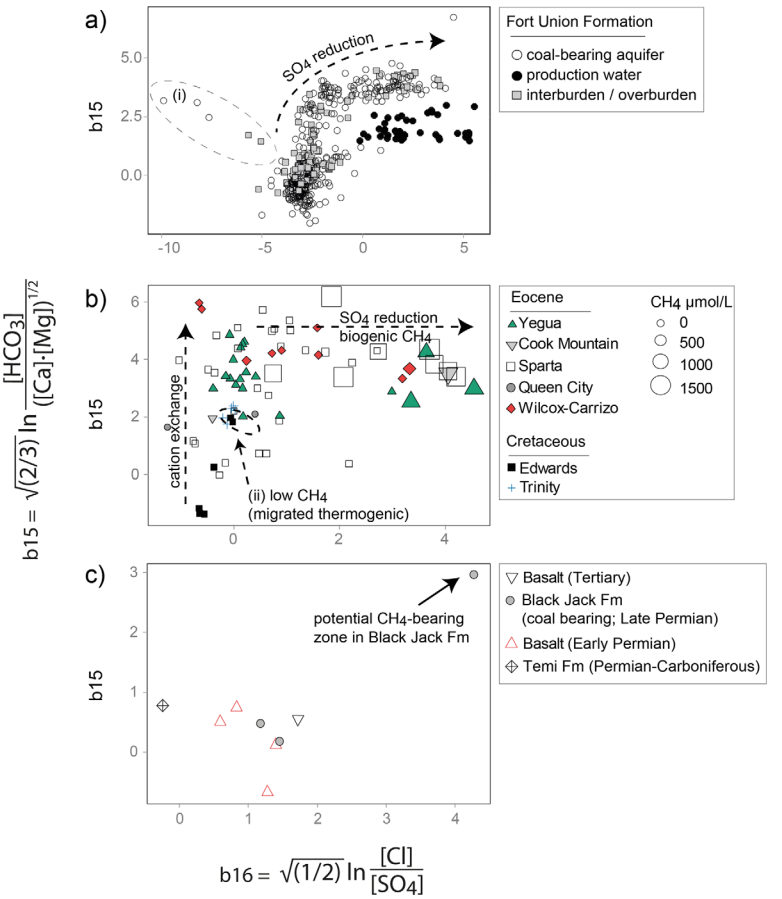

Figure 13. An isometric log ratio alternative to the RA versus Cl for: (a) Powder River Basin (USA), showing (i); the same outliers identified in Figure 3a; (b) Gulf of Mexico Basin (USA), showing high-gas (biogenic) samples (shown by large point size) and (ii) low-gas samples associated with migrated thermogenic CH_4; and (c) Gunnedah Basin (Australia).

to the theory that the compositional characteristics (the relative proportion of ions) of gas-bearing groundwater in the coal seams are related to spatial variability in the lithological/coal matrix and/or endemic methanogenesis.

4.5.4. Conceptualizing the Evolution of the Characteristic CSG End-Member (Surat and Clarence-Moreton Basins)

Taking these results and information we can build a conceptual model of the understanding of the evolution of the compositional characteristics of the CSG groundwater in the Surat/Clarence-Moreton Basins (Condamine catchment) (Figure 12). Steps 1–2 in Figure 12 conceptualize the compositional changes from a low salinity (i) or brackish (v) starting point.

At step 2 (Figure 12) increases in HCO_3 are likely to be related to the biodegradation of coal: in both scenarios (a) and (b) there must be an increase in the relative proportion of Cl also, although more so for (a). After SO_4 reduction is complete, high rates of methanogenesis proceed and the characteristic CSG composition evolves (step 3). The fact that the behavior of ions in this composition can be modeled indicates two distinct end-members (as shown in Figures 6 and 7 and the ternary model in Figure 12 3a/3b). This result suggests either mixing and/or a hydrochemical response to variable, but constrained conditions (e.g., lithology/biological processes). Given that a source of high Cl is not observed in underlying or overlying aquifers/zones, then any mixing scenario is likely to be related to in situ mixing within the same lithological matrix. Variability in the lithological matrix and methanogenic consortia, especially under closed conditions, may also, result in a hydrochemical response.

Overall, these exact mechanisms and controls on compositional variability need resolving. The advantage of using the compositional data analysis techniques here is that the large data set (over 500 samples with

~180 from the coal measures) can be rapidly assessed and CSG and other groundwater that is likely to contain gas that has been generated in situ can be easily delineated. This is vital information. Previously it could not be obtained using major ions alone and it cannot be derived from conventional plots, such as Piper plots. Note that the conceptual model in Figure 12 is much more informative than assuming the evolution of the CSG end-member in these basins follows a theoretical and universal SO_4 reduction pathway (Figure 1a).

4.5.5. Delineating CSG Pathways in a Range of Other Basins

The ilr-coordinates b15 and b16 are also effective in delineating groundwater associated with high gas concentrations in other basins (Figures 13a–13c). In the Powder River Basin, increasing b15 and b16 values show the depletion of SO_4, Ca, and Mg, and enrichment of HCO_3 along a SO_4 reduction pathway (Figure 13a). The outliers shown in Figure 11a (i) are the same outliers observed in Figure 3a. In the Gulf of Mexico Basin (Texas, USA), a steep increase in b15 values is indicative of the cation exchange that occurs in the Eocene aquifer [*Grossman et al*., 1989; *Zhang et al*., 1998]. Subsequent increases in b16 values for samples show the SO_4 reduction required prior to microbial methanogenesis. Samples which contain thermogenic gas that has migrated from underlying petroleum reservoirs can be distinguished from biogenic samples (Figure 13b, (ii)) [*Grossman et al*., 1989; *Zhang et al*., 1998]. Data for the Gunnedah Basin (NSW, Australia) where a CSG end-member has not been described, show a single sample from the coal-bearing Black Jack Formation is distinct (Figure 13c) and is likely to be associated with coal and may also contain gas.

5. Conclusions

The typical Na-HCO_3 or Na-Cl-HCO_3 water type of coal seam gas (CSG) groundwater is universal and by now well-known, but while the dominant ions may be similar, the compositional variability may be different. In this study, using examples from a number of basins in the USA and Australia, we use sequential binary partitions to define orthogonal isometric log ratio (ilr-coordinates) coordinates that can describe particular relationships between ions that are unique to different water types or aquifers.

These techniques allow the structure of the compositional simplex to be comprehensively investigated with respect to both major and minor ion components, and the identification of relationships between the relative proportion of ions that cannot be observed using conventional techniques such as Piper plots, Stiff diagrams, or classic ion ratios. Relationships between isometric log ratios, expressed as linear or quadratic models, derived from appropriate sequential binary partitions are a more robust and informative description of water types than the traditional means of classifying water types by their dominant ions. Appropriate sequential binary partitions can be developed using information derived from clr-bivariate plots, covariance matrices of log-ratios and geochemical intuition.

Using these compositional data analysis techniques the evolution of CSG water types in the Powder River Basin (USA) via a SO_4 reduction pathway was modeled. In the Surat and Clarence-Moreton basins relationships between Ca, Mg, F, HCO_3, and Cl ions were found to be the major controls on CSG water types. Using appropriate ilr-coordinates, Na-HCO_3 water types in non-gas-bearing zones/aquifers could be differentiated from CSG water types by compositional relationships between F, Cl, and K. These distinctions are unlikely to have been observed using conventional techniques and offer important insights into the variability between similar Na-HCO_3water types.

Finally, this paper developed two ilr-coordinates that can be used to delineate CSG water types in large data sets. These ilr-coordinates provide mathematical descriptions of the relationships between proportions of ions that are inherently related to complex lithological and biological processes in areas where high concentrations of biogenic gas occur. Future studies in these basins that are concerned with aquifer interaction or the occurrence of gas could benefit by designing strategic sampling programs based on results of this study. These subsequent studies can apply the same techniques to larger data sets that also contain minor ion, trace element or isotope data to make further hydrochemical assessments. More broadly, the results presented here demonstrate the value of compositional data analysis techniques in describing particular hydrochemical water types, and delineating unique water types in large data sets.

Acknowledgments
Thanks are extended to entities that provided hydrochemical data for use in this study including: Ethan L. Grossman at the University of Texas, the Montana Bureau of Mines and Geology, Elizabeth Meredith and John Wheaton at the Montana Bureau of Mines and Geology, Clément Duvert from QUT, the QLD Department of Natural Resources and Mines, and Arrow Energy. Thanks are also extended to David Lovell, QUT, for his advice on data analysis, Malcolm E. Cox for general project support, and Clément Duvert from QUT for comments on the figures, discussion, and paper structure. V. Pawlowsky-Glahn and J. J. Egozcue have been supported by the Ministerio de Economía y Competitividad ("CODA-RETOS" (Ref. MTM2015-65016-C2-1-R; MINECO/FEDER, UE); and by the Generalitat de Catalunya (project AGAUR:COSDA:2009SGR424). Thanks are also extended to an Associate Editor at *Water Resources Research*, and two reviewers, Madalyn S. Blondes and an anonymous reviewer, for their constructive comments and suggestions to improve the paper. Data used in this study can be accessed via the following: Powder River basin—publicly available data from the Montana Bureau of Mines and Geology database (http://mbmggwic.mtech.edu/) and production water data from *Rice et al.* [2000]; Gulf of Mexico Basin—published data from *Grossman et al.* [1989] and *Zhang et al.* [1998] (raw data provided by Grossman E. L., Texas A&M University, USA, e-grossman@geos.tamu.edu); Surat and Clarence-Moreton basins—publicly available data from the QLD government groundwater database (data available from QLD Department of Natural Resources and Mines https://www.dnrm.qld.gov.au/our-department/natural-resources-and-mines-data. Surat and Clarence-Moreton basins: CSG data provided by Arrow Energy, St John Herbert StJohn.Herbert@arrowenergy.com.au. Surat and Clarence-Moreton basins: new data (CSG and shallow coal measures—see Supporting Information). Teviot Brook catchment: published data from *Duvert et al.* [2015] and unpublished data from J. Li (data provided by Queensland University of Technology: contact School of Environmental, Biological and Earth Sciences sef.eebsenquiry@qut.edu.au), and Gunnedah Basin—unpublished data provided by J. Bradd, University of Wollongong, Australia (see Supporting Information). This study is part funded by a scholarship provided by Arrow Energy to the Queensland University of Technology and subsequently awarded to Daniel D. R. Owen. The funding provider was not involved in the preparation of this paper, including in the analyses of data, the development of the results or discussion, or the decision to submit the paper for publication.

References

Aitchison, J. (1982), The statistical analysis of compositional data (with discussion), *J. R. Stat. Soc., Ser. B*, *44*(2), 139–177.
Aitchison, J. (1986), *The Statistical Analysis of Compositional Data, Monogr. Stat. Appl. Probab.*, Chapman and Hall, London, U. K.
Aitchison, J., and M. Greenacre (2002), Biplots of compositional data, *J. R. Stat. Soc., Ser. C*, *51*(4), 375–392.
Atkins, M. L., I. R. Santos, and D. T. Maher (2015), Groundwater methane in a potential coal seam gas extraction region, *J. Hydrol. Reg. Stud.*, *4*, 452–471, doi:10.1016/j.ejrh.2015.06.022.
Bacon-Shone, J. (2006), A short history of compositional data analysis, in *Compositional Data Analysis: Theory and Applications*, vol. 378, edited by V. Pawlowsky-Glahn and A. Buccianti, John Wiley, Chichester, U. K.
Bartos, T. T., and K. M. Ogle (2002), Water quality and environmental isotopic analyses of ground-water samples collected from the Wasatch and Fort Union formations in areas of coalbed methane development—Implications to recharge and groundwater flow, eastern Powder River Basin, Wyoming, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, 02-4045, pp. 1–89.
Baublys, K. A., S. K. Hamilton, S. D. Golding, S. Vink, and J. Esterle (2015), Microbial controls on the origin and evolution of coal seam gases and production waters of the Walloon Subgroup; Surat Basin, Australia, *Int. J. Coal Geol.*, *147–148*, 85–104, doi:10.1016/j.coal.2015.06.007.
Bicocchi, G., G. Montegrossi, G. Ruggieri, A. Buccianti, and O. Vaselli (2011), Modeling composition of Ca-Fe-Mg carbonates in a natural CO_2 reservoir, in *Proceedings of the 4th International Workshop on Compositional Data Analysis*, edited by J. J. Egozcue, R. Tolosana-Delgado, and M. I. Ortego, 166 pp., Girona, Spain.
Buccianti, A., and V. Pawlowsky-Glahn (2005), New perspectives on water chemistry and compositional data analysis, *Math. Geol.*, *37*(7), 703–727, doi:10.1007/s11004-005-7376-6.
Cadman, S. J., L. Pain, and V. Vuckovic (1998), Bowen and Surat Basins, Clarence-Moreton Basin, Sydney Basin, Gunnedah Basin and other minor onshore basins, Qld, NSW and NT, in *Australian Petroleum Accumulations Report*, 783 pp., Dept. of Primary Ind. and Energy, Bur. of Resour. Sci., Canberra, Australia.
Chae, G.-T., S.-T. Yun, K. Kim, and B. Mayer (2006), Hydrogeochemistry of sodium-bicarbonate type bedrock groundwater in the Pocheon spa area, South Korea: Water–rock interaction and hydrologic mixing, *J. Hydrol.*, *321*(1–4), 326–343, doi:10.1016/j.jhydrol.2005.08.006.
Chayes, F. (1960), On correlation between variables of constant sum, *J. Geophys. Res.*, *65*(12), 4185–4193.
Chebotarev, I. I. (1955), Metamorphism of natural waters in the crust of weathering—1, *Geochim. Cosmochim. Acta*, *8*(1–2), 22–48, doi:10.1016/0016-7037(55)90015-6.
Comas-Cufí, M., and S. Thió-Henestrosa (2011), CoDaPack 2.0: A stand-alone, multi-platform compositional software CoDaWork'11, in CoDaWork'11: 4th International Workshop on *Compositional Data*, edited by J. J. Egozcue, R. Tolosana-Delgado, and M. I. Ortego, 28 pp., Girona, Spain.
Conrad, R. (1999), Contribution of hydrogen to methane production and control of hydrogen concentrations in methanogenic soils and sediments, *FEMS Microbiol. Ecol.*, *28*(3), 193–202, doi:10.1111/j.1574-6941.1999.tb00575.x.
Conrad, R. (2005), Quantification of methanogenic pathways using stable carbon isotopic signatures: A review and a proposal, *Org. Geochem.*, *36*(5), 739–752, doi:10.1016/j.orggeochem.2004.09.006.
Cook, A. G., and J. J. Draper (Eds.) (2013), *Geology of Queensland*, Geol. Surv. of Queensland, Brisbane, Australia.
Draper, J. J., and C. J. Boreham (2006), Geological controls on exploitable coal seam gas distribution in Queensland, *APPEA J.*, *46*, 343–366.
Duvert, C., M. Raiber, D. D. R. Owen, D. I. Cendón, C. Batiot-Guilhe, and M. E. Cox (2015), Hydrochemical processes in a shallow coal seam gas aquifer and its overlying stream–alluvial system: Implications for recharge and inter-aquifer connectivity, *Appl. Geochem.*, *61*, 146–159, doi:10.1016/j.apgeochem.2015.05.021.
Egozcue, J. J., and V. Pawlowsky-Glahn (2005), Groups of parts and their balances in compositional data analysis, *Math. Geol.*, *37*(7), 795–828, doi:10.1007/s11004-005-7381-9.
Egozcue, J. J., and V. Pawlowsky-Glahn (2011), Basic concepts and procedures, in *Compositional Data Analysis: Theory and Applications*, vol. 378, edited by V. Pawlowsky-Glahn and A. Buccianti, pp. 12–27, John Wiley, Chichester, U. K.
Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal (2003), Isometric logratio transformations for compositional data analysis, *Math. Geol.*, *35*(3), 279–300, doi:10.1023/A:1023818214614.
Engle, M., and E. Rowan (2013), Interpretation of Na–Cl–Br systematics in sedimentary basin brines: comparison of concentration, element ratio, and isometric log-ratio approaches, *Math. Geosci.*, *45*(1), 87–101, doi:10.1007/s11004-012-9436-z.
Engle, M. A., F. R. Reyes, M. S. Varonka, W. H. Orem, L. Ma, A. J. Ianno, T. M. Schell, P. Xu, and K. C. Carroll (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, *Chem. Geol.*, *425*, 76–92, doi:10.1016/j.chemgeo.2016.01.025.
Exon, N. F. (1976), *The Geology of the Surat Basin: Bulletin 166*, Aust. Gov. Publ. Serv., Canberra.
Feitz, A. J., T. R. Ransley, R. Dunsmore, T. J. Kuske, J. Hodgkinson, M. Preda, R. Spulak, O. Dixon, and J. J. Draper (2014), Geoscience Australia and Geological Survey of Queensland Surat and Bowen Basins Groundwater Surveys Hydrochemistry Dataset (2009-2011), http://www.ga.gov.au/metadata-gateway/metadata/record/78549/, Commonw. of Aust. (Geosci. Aust.), Canberra.
Flores, R. M., Rice, C. A., Stricker, G. D., Warden, A. and Ellis, M. S. (2008), Methanogenic pathways of coal-bed gas in the Powder River Basin, United States: The geologic factor, *Int. J. Coal Geol. 76*(1–2), 52–75, doi:10.1016/j.coal.2008.02.005.
GeoScience Australia (2016), Groundwater Hydrochemical Characterisation of the Surat Region and Laura Basin, Queensland. [Available at http://www.ga.gov.au/about/what-we-do/projects/water/groundwater-hydrochemical-characterisation-of-the-surat-region-and-laura-basin-queensland, Canberra.]
Gluskoter, H. J., and R. R. Ruch (1971), Chlorine and sodium in Illinois coals as determined by neutron activation analyses, *Fuel*, *50*(1), 65–76, doi:10.1016/S0016-2361(71)81021-9.
Golding, S. D., C. J. Boreham, and J. S. Esterle (2013), Stable isotope geochemistry of coal bed and shale gas and related production waters: A review, *Int. J. Coal Geol.*, *120*, 24–40, doi:10.1016/j.coal.2013.09.001.
Green, M. S., K. C. Flanegan, and P. C. Gilcrease (2008), Characterization of a methanogenic consortium enriched from a coalbed methane well in the Powder River Basin, USA, *Int. J. Coal Geol.*, *76*(1–2), 34–45, doi:10.1016/j.coal.2008.05.001.
Grigorescu, M. (2011), *Mineralogy of the South-Eastern Bowen Basin and North-Eastern Surat Basin, Queensland*, Geol. Surv. of Queensland, Brisbane, Australia.
Grossman, E. L., K. B. Coffman, S. J. Fritz, and H. Wada (1989), Bacterial production of methane and its influence on ground-water chemistry in east-central Texas aquifers, *Geology*, *17*, 495–499.
Hamawand, I., T. Yusaf, and S. G. Hamawand (2013), Coal seam gas and associated water: A review paper, *Renewable Sustainable Energy Rev.*, *22*, 550–560, doi:10.1016/j.rser.2013.02.030.

Hamilton, S. K., S. D. Golding, K. A. Baublys, and J. S. Esterle (2014), Stable isotopic and molecular composition of desorbed coal seam gases from the Walloon Subgroup, eastern Surat Basin, Australia, *Int. J. Coal Geol.*, *122*, 21–36, doi:10.1016/j.coal.2013.12.003.

Hem, J. D. (1985), Study and interpretation of the chemical characteristics of natural water, *U.S. Geol. Surv. Water Supply Pap.*, 2254, 264 pp.

Hjelm, O., E. Johansson, and G. Öberg (1999), Production of organically bound halogens by the litter-degrading fungus *Lepista nuda*, *Soil Biol. Biochem.*, *31*(11), 1509–1515, doi:10.1016/S0038-0717(99)00069-3.

Huggins, F. E., and G. P. Huffman (1995), Chlorine in coal: An XAFS spectroscopic investigation, *Fuel*, *74*(4), 556–569, doi:10.1016/0016-2361(95)98359-M.

Jell, P. A., J. L. McKellar, and J. J. Draper (2013), Clarence-Moreton Basin, in *Geology of Queensland*, Geol. Surv. of Queensland, Brisbane, Australia.

Kimura, T. (1998), Relationships between inorganic elements and minerals in coals from the Ashibetsu district, Ishikari coal field, Japan, *Fuel Process. Technol.*, *56*(1–2), 1–19, doi:10.1016/S0378-3820(97)00089-1.

Kinnon, E. C. P., S. D. Golding, C. J. Boreham, K. A. Baublys, and J. S. Esterle (2010), Stable isotope and water quality analysis of coal bed methane production waters and gases from the Bowen Basin, Australia, *Int. J. Coal Geol.*, *82*(3–4), 219–231, doi:10.1016/j.coal.2009.10.014.

Lovell, D., V. Pawlowsky-Glahn, J. J. Egozcue, S. Marguerat, and J. Bähler (2015), Proportionality: A valid alternative to correlation for relative data, *PLoS Comput. Biol. 11*(3), e1004075, doi:10.1371/journal.pcbi.1004075.

Montgomery, S. L. (1999), Powder River Basin, Wyoming: An expanding coalbed methane (CBM) play, *AAPG Bull.*, *83*(8), 1207–1222.

Nisi, B., A. Buccianti, B. Raco, and R. Battaglini (2015), Analysis of complex regional databases and their support in the identification of background/baseline compositional facies in groundwater investigation: Developments and application examples, *J. Geochem. Explor.*, *164*, 3–17, doi:10.1016/j.gexplo.2015.06.019.

Öberg, G., and P. Sandén (2005), Retention of chloride in soil and cycling of organic matter-bound chlorine, *Hydrol. Processes*, *19*(11), 2123–2136, doi:10.1002/hyp.5680.

Öberg, G., M. Holm, P. Sandén, T. Svensson, and M. Parikka (2005), The role of organic-matter-bound chlorine in the chlorine cycle: A case study of the Stubbetorp Catchment, Sweden, *Biogeochemistry*, *75*(2), 241–269, doi:10.1007/s10533-004-7259-9.

Owen, D. D. R., and M. E. Cox (2015), Hydrochemical evolution within a large alluvial groundwater resource overlying a shallow coal seam gas reservoir, *Sci. Total Environ.*, *523*, 233–252, doi:10.1016/j.scitotenv.2015.03.115.

Owen, D. D. R., M. Raiber, and M. E. Cox (2015), Relationships between major ions in coal seam gas groundwaters: Examples from the Surat and Clarence-Moreton basins, *Int. J. Coal Geol.*, *137*, 77–91, doi:10.1016/j.coal.2014.11.004.

Palarea-Albaladejo, J., and J. A. Martín-Fernández (2015), zCompositions—R package for multivariate imputation of left-censored data under a compositional approach, *Chemom. Intell. Lab. Syst.*, *143*, 85–96, doi:10.1016/j.chemolab.2015.02.019.

Papendick, S. L., K. R. Downs, K. D. Vo, S. K. Hamilton, G. K. W. Dawson, S. D. Golding, and P. C. Gilcrease (2011), Biogenic methane potential for Surat Basin, Queensland coal seams, *Int. J. Coal Geol.*, *88*(2–3), 123–134, doi:10.1016/j.coal.2011.09.005.

Pawlowsky-Glahn, V., and A. Buccianti (2002), Visualization and modeling of sub-populations of compositional data: Statistical methods illustrated by means of geochemical data from fumarolic fluids, *Int. J. Earth Sci.*, *91*(2), 357–368, doi:10.1007/s005310100222.

Pawlowsky-Glahn, V., and J. J. Egozcue (2006), Compositional data and their analysis: An introduction, *Geol. Soc. Spec. Publ.*, *264*, 1–10.

Pawlowsky-Glahn, V., and J. J. Egozcue (2011), Exploring compositional data with the coda-dendrogram, *Austrian J. Stat.*, *40*(1–2), 103–113.

Pawlowsky-Glahn, V., J. J. Egozcue, and R. Tolosana-Delgado (2015), Exploratory data analysis, in *Modeling and Analysis of Compositional Data*, John Wiley, Singapore.

Pearson, K. (1897), Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs, *Proc. R. Soc. London*, *60*, 489–502.

Pratt, W. (1998), Gunnedah coalfield (North) regional geology, Map 2856484, Geol. Surv. of N. S. W., St Leonards, Australia.

Quillinan, S. A., and C. D. Frost (2014), Carbon isotope characterization of powder river basin coal bed waters: Key to minimizing unnecessary water production and implications for exploration and production of biogenic gas, *Int. J. Coal Geol.*, *126*, 106–119, doi:10.1016/j.coal.2013.10.006.

R development Core Team 2003 (2015), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna. [Available at http://www.R-project.org.]

Rice, C. A., M. S. Ellis, and J. H. Bullock Jr. (2000), Water co-produced with coalbed methane in the Powder River Basin, Wyoming: preliminary compositional data, *U.S. Geol. Surv. Open File Rep.*, 00-372, 20 pp.

Rice, C. A., R. M. Flores, G. D. Stricker, and M. S. Ellis (2008), Chemical and stable isotopic evidence for water/rock interaction and biogenic origin of coalbed methane, Fort Union Formation, Powder River Basin, Wyoming and Montana U.S.*A, Int. J. Coal Geol.*, *76*(1–2), 76–85, doi:10.1016/j.coal.2008.05.002.

Rice, D. D. (1993), Composition and origins of coalbed gas, in *Hydrocarbons From Coal, Stud. Geol.*, edited by B. E. Law and D. D. Rice, pp. 159–184, Am. Assoc. of Pet. Geol., Tulsa, Okla.

Taulis, M., and M. Milke (2007), Coal seam gas water from Maramarua, New Zealand: Characterisation and comparison to US analogues, *J. Hydrol.*, *46*(1), 1–17.

Taulis, M., and M. Milke (2013), Chemical variability of groundwater samples collected from a coal seam gas exploration well, Maramarua, New Zealand, *Water Res. 47*(3), 1021–1034, doi:10.1016/j.watres.2012.11.003.

Thió-Henestrosa, S., and M. Comas (2011), *CoDaPack v2 User's Guide, Dep. of Comput. Sci. and Appl. Math.*, Univ. of Girona, Girona, Spain.

Van Voast, W. A. (2003), Geochemical signature of formation waters associated with coalbed methane, *AAPG Bull.*, *87*(4), 667–676.

Venturelli, G., T. Boschetti, and V. Duchi (2003), Na-carbonate waters of extreme composition: Possible origin and evolution, *Geochem. J.*, *37*, 351–366.

von Eynatten, H., V. Pawlowsky-Glahn, and J. Egozcue (2002), Understanding perturbation on the simplex: A simple method to better visualize and interpret compositional data in ternary diagrams, *Math. Geol.*, *34*(3), 249–257, doi:10.1023/A:1014826205533.

Wickham, H. (2009), *ggplot2: Elegant Graphics for Data Analysis*, Springer, N. Y.

WorleyParsons (2012), Spatial analysis of coal seam gas water chemistry, in *Healthy HeadWaters Coal Seam Gas Water Feasibility Study*, Record 25, Dep. of Nat. Resour. and Mines, Brisbane, Queensland, Australia.

Yudovich, Y. E., and M. P. Ketris (2006), Chlorine in coal: A review, *Int. J. Coal Geol.*, *67*(1–2), 127–144, doi:10.1016/j.coal.2005.09.004.

Zhang, C., E. L. Grossman, and J. W. Ammerman (1998), Factors influencing methane distribution in Texas ground water, *Ground Water*, *36*(1), 58–66, doi:10.1111/j.1745-6584.1998.tb01065.x.