Carnegie Mellon University

Carnegie Institute of Technology

THESIS

Submitted in partial fulfillment of the requirements

for the degree of **Doctor of Philosophy**

Measurement and Recovery of Rare Earth Elements from Hypersaline Fluids

Clinton W. Noack

July 10, 2015

This page intentionally left blank.

'

Measurement and Recovery of Rare Earth Elements from Hypersaline Fluids

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Civil & Environmental Engineering

Clinton W. Noack

B.S., Environmental Systems Engineering, Pennsylvania State University

M.S., Civil & Environmental Engineering, Carnegie Mellon University

Carnegie Mellon University

Pittsburgh, PA

December, 2015

Contents

List of Tables

List of Figures

Introduction, problem identification, and research goals

1.1   Introduction

1.2   Problem identification

1.3   Research goals

Chapter 2

Rare earth element distributions and trends in natural waters with a focus on groundwater

This chapter is adapted from a publication by the same name, co-authored by David A. Dzombak and Athanasios K. Karamalidis. This paper is citable as:

My contributions to this work were the collection, analysis, and visualization of the data; development of `R` scripts; interpretation of results; and drafting of the manuscript.

Abstract

Systematically varying properties and reactivities have led to focused research of the environmental forensics capabilities of the rare earth elements (REE). Increasing anthropogenic inputs to natural systems may permanently alter the natural signatures of REEs, motivating characterization of natural REE variability. We compiled and analyzed reported dissolved REE concentration data over a wide range of natural water types (groundwater, ocean-, river-, and lake water) and groundwater chemistries (e.g. fresh, brine, and acidic) with the goal of quantifying the extent of natural REE variability, especially for groundwater systems. Quantitative challenges presented by censored data were addressed with non-parametric distributions and regressions. Reported measurements of rare earth elements in natural waters range over nearly ten orders of magnitude, though the majority of measurements are within two to four orders of magnitude, and are highly correlated with one another. Few global correlations exist among dissolved abundance and bulk solution properties in groundwaters indicating the complex nature of source-sink terms and the need for care when comparing results between studies. This collection, homogenization, and analysis of a disparate literature facilitates inter-study comparison and provides insight into the wide range of variables that influence REE geochemistry.

2.1   Introduction

In the natural sciences, predictable thermodynamic differences between the rare earth elements (REE) allow for interpretation of natural geologic and chemical processes [1, 2]. Rare earth lithogeochemistries have long been used to infer depositional environments of geologic strata [1, 3, 4]. Similarly, REE serve as benign analogs to the transuranic actinides for nuclear waste disposal studies [5, 6] and for studying mixing and metal cycling in the oceans [7, 8]. These characteristics make REE attractive tools for environmental forensic applications such as pollutant source identification and apportionment [9, 10].

Based on atomic number, the REE are segregated into light and heavy REE (LREE and HREE, respectively) with the division occurring between Eu and Gd [11]; some studies also distinguish middle REE (MREE), though the specific elements are inconsistently defined between authors [12–15]. These

"weight" distinctions allow for simplified description and quantification of the inter-element relationships, typically ratios of normalized concentrations, which are exploited in REE analysis. Similarly, anomalies of certain REE – due to redox lability for Ce and Eu [15] and large anthropogenic emissions for Gd [16] – are used to interpret geochemical processes. Y and Sc exhibit similar properties to the lanthanides and are thus included in the suite of REE with Y being most similar to HREE and Sc being most similar to LREE in solution [15].

Aquatic geochemists apply the same principles used by geologists, the inter-element ratios and anomalies described previously, to infer water-rock interactions, hydrologic connectivity between geologic units, and groundwater mixing members [17–20]. Interactions with different mineral phases have been shown to alter REE patterns predictably. For example, an MREE enrichment is observed for fresh waters in contact with phosphate-rich minerals [12] while HREE enrichment is found in carbonate-rich waters [21]. Tester et al. [22] showed that groundwater end-members, especially at shallow depths, could be established by interpretation of REE patterns. Similarly, REE concentrations have been used to calculate end-member contributions to groundwater [17, 19].

The capabilities of REE to serve as tracers of groundwater migration and mixing have potential applications to the study of hydrology and geochemistry of shale gas development or the capture and sequestration of carbon dioxide (CO_2) gas in geologic formations. For example, characteristic REE profiles could possibly enable detection of brines displaced from shale or CO_2 storage zones and into overlying groundwater aquifers [23–26]. Capable tools for contaminant detection, as well as contaminant source apportionment, are critical to the long-term environmental feasibility of these technologies. Increasing anthropogenic inputs of REE to the environment threaten to obfuscate natural signals, which would complicate these applications [27, 28].

A primary challenge in analyzing REE data, like other geochemical compositional data [29], is the prevalence of censored observations, defined as values below the analytical method detection limit (MDL) or a value excluded due to excessive analytical interferences. Appropriate statistical methodologies, largely borrowed from survival analysis (or reliability analysis), allow for analysis of these data without relying on substitution or interpolation [30]. Utilizing these techniques removes the bias of left-censored data and of uneven sample sizes from summary statistics and statistical inference

[30, 31].

Significant contributions to the understanding of REE geochemistry in natural waters have been made through critical review of thermodynamics [15, 32], catchment-scale case studies [33–36], and groundwater flow system studies [17, 37–40] among others. Lacking thus far in the REE literature is the aggregation and analysis of the numerous and disparate studies of the REE and the origin of their concentrations. Such a compilation would enhance inter-study comparison through internal consistency, allow for investigation of broader trends, and enhance statistical analysis by increasing sample sizes.

In this work, a compilation and analysis of data from independent studies of REE in natural waters was performed, focusing on trends in groundwaters. The study focused on shallow groundwaters well as some deeper groundwaters capable of impacting surface water or shallow groundwater. The compiled data were used to develop a consistent database of REE concentrations and their associated major solute chemistry and to explore interelement relationships, examine trends in REE abundance, and test hypotheses related to REE abundance as functions of major solution chemistry parameters. The objectives of this work were: (1) to ascertain an expected range of dissolved REE concentrations in waters of variable chemistries, deriving unbiased estimates of REE distributions and (2) to investigate trends in REE abundance in groundwater in relation to other available chemical parameters (e.g. pH, ionic strength, and major solution species).

2.2   Methods

2.2.1   Data assimilation and criteria

The low-temperature, aqueous REE systematics were investigated by extracting available REE concentration data in reported studies of ground-, sea-, river-, and lake waters. Data were gathered from these studies for dissolved REE concentrations (defined here as constituents passing through a 0.45 µm filter). Many studies filter at finer pore sizes in order to study colloidal association [33, 41, 42], however 0.45 µm was chosen because it is a common size for groundwater analysis and maximized the number of samples available. While REE are commonly defined to include Y and Sc, very few studies measured Y and almost none examined Sc. Therefore, our

analyses of REE focused on the 14 naturally abundant lanthanides (excluding the short-lived, radioactive Pm). In total, 31 studies with 619 samples were collected for groundwater,14, 19, 20, 22, 33-35, 37-40, 43-63 16 studies with 178 samples for seawater,7, 45, 64-77 16 studies with 259 samples for river water,12, 19, 48, 67, 71, 78-89 and 8 studies with 74 samples for lake water.19, 56, 57, 82, 85, 90-92

A significant literature exists for other water types, for example mine drainage [43–49] and hydrothermal waters/hot springs [50–55]. However, these water types have been excluded from this study in order to focus on waters minimally impacted by anthropogenic activity (in the case of mine drainage) and that occur broadly with chemistry dominated by near-surface pressure, temperature, and mineralogical conditions (in the case of hydrothermal waters/hot springs).

2.2.2 Analysis of censored data

In the accumulated dataset, 18% of all REE measurements were either below detection or missing. For individual elements in different waters, up to 52% of measurements were censored (Tm in groundwater). The censoring rates of each element, in each water type, are given in Figure 2.1. Given this high frequency, it was necessary to employ modified statistical methods to handle this type of information31 and avoid creating a false data signal either by substitution (i.e., one-half MDL) or interpolation.

An additional complexity of this data set was the uneven distribution of data points between data sets (Figure 2.2). For example, some data sets included more than seventy samples, while other contained fewer than ten. This uneven distribution can bias calculated statistics towards studies with a higher number of samples, giving larger studies more weight than smaller studies [56].

Censored data were stored at the study-specific MDL with a separate binary variable indicating that the data were censored. Missing data – that is where REE were either not measured, not reported, or where an MDL was not specified – were excluded from calculations because nothing was known about the data point in relation to the rest of the data.

Summary statistics were calculated using a weighted Kaplan-Meier (KM) estimator of the survival function accounting for both left-censored and source-size bias. A routine based on the methodology of Singh, et al. [56] was writ-
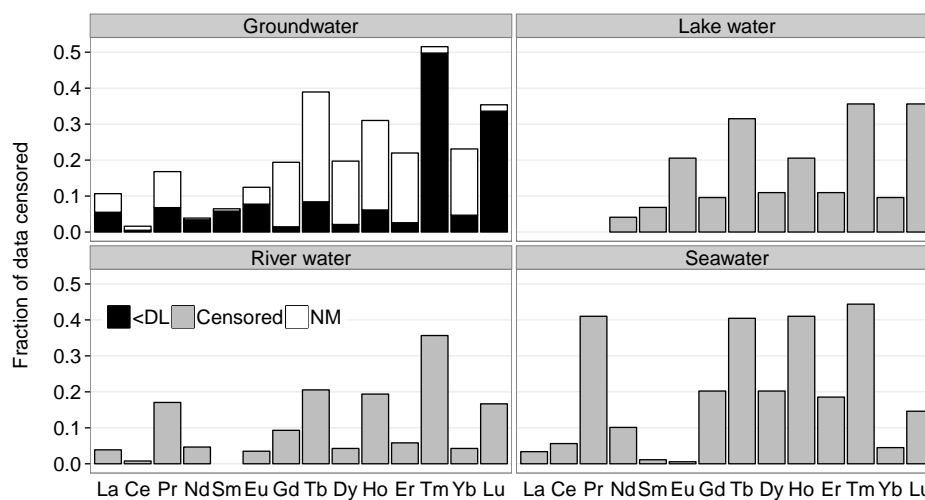
Figure 2.1: Fraction of below detection limit (<DL) or missing data (NM) for 14 REE in data sets for water types studied. For water types other than groundwater, censored data were not differentiated as below detection or missing, and are lumped together as "censored".
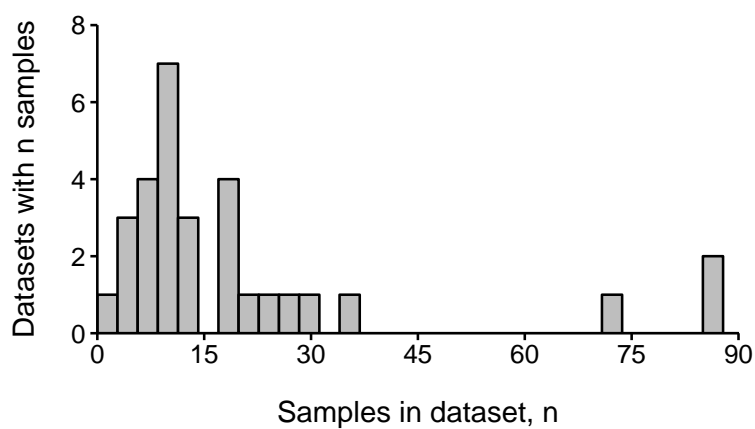
Figure 2.2: Histogram of sample distribution in groundwater dataset (compiled from 31 articles).

ten in R [57] to perform the necessary calculations. Distribution percentiles were estimated as the first value with a calculated percentile less than the percentile of interest; stated differently, if the calculated survival quantiles, $S(x)$, for adjacent observations were $S(x_1) = 0.94$ and $S(x_2) = 0.96$, then x_1 would be noted as the 95th percentile. For data sets with a limited number of samples (e.g. brines) where the sample size or censoring frequency limited the number of calculable percentiles, a log-normal distribution was fit to the KM survival curve by regression on order statistics (ROS) and was used to estimate those percentiles. Mean values, μ_{KM}, were calculated from the area under the survival curve and standard deviations, σ_{KM}, from the variance of the mean [58].

The survival curve was approximated using a weighted KM estimator, which considers non-detect data and provides robustness to the uneven distribution of samples from the reviewed literature. Equation 2.1 represents the weighted KM estimator, where $\hat{S}(x_i)$ is the probability any observation from the dataset will be *less than* the measured concentration, x_i. In this equation d_i^w represents the weighted count of uncensored observations at concentration x_i and Y_i^w represents the weighted count of all (censored and uncensored) observed concentrations less than x_i. The weight, w_i, of each observation is the inverse of the number of samples from the data source (n_i), or: $w_i = n_i^{-1}$. In the absence of weighting or censoring, this formula reduces to the empirical cumulative distribution function. These calculations, and the relevant R code, are discussed algorithmically in Appendix **A**.

$$\hat{S}(x_i) = \prod_{x \geq x_i} \left(1 - \frac{d_i^w}{Y_i^w} \right) \tag{2.1}$$

Uncertainty in this estimate $(\sigma_{\hat{S}}(x_i))$, expressed using Greenwood's formula [58], is defined by Equation 2.2. As with Equation 2.1, d_i^w represents the weighted count of uncensored observations at concentration x_i and Y_i^w represents the weighted count of all observed concentrations less than x_i.

$$\sigma_{\hat{S}}(x_i) = \hat{S}(x_i) \sqrt{\sum_{j=1}^{k} \frac{d_j^w}{Y_j^w (Y_j^w - d_j^w)}} \tag{2.2}$$

Total dissolved REE concentrations were calculated with consideration of censored data. Rather than calculating the KM survival curve and μ_{KM} for an individual element (where each sample is an observation) the survival curve was calculated for individual samples (where each REE is an observation). The sum was determined as follows [58]:

$$\sum[REE] = 14 \times \mu_{KM} \tag{2.3}$$

In addition to univariate summary statistics, inter-element correlations are significantly influenced by the high degree of censoring within the REE data. Using Pearson's r (linear, least-squares correlation) as an estimate of the correlation would require substitution for censored or missing values [30], while also making assumptions about the relationship between the variables. Instead, the Kendall's τ statistic enabled incorporation of censored data, diminished the influence of outliers (by being robust to log-transformation), and assessed monotonic relationships beyond simple linear relationships. Thus Kendall's τ was used for assessing relationships between any variables with multiple censoring levels (detection limits). The Kendall's τ method will typically produce lower numeric values (often 0.2 units lower for the same strength of correlation) than other measures, which may make interpretation difficult [58]; however, Kendall's τ is the only technique that allows for inclusion of multiple detection limits non-parametrically without substitution or imputation. Correlations between variables without censoring were evaluated non-parametrically with Spearman's ρ.

2.2.3   Reference-normalized element anomalies and ratios

Several reduced dimension variables (RDVs) are calculated from reference normalized data including ratios of the LREE to MREE and HREE, the Ce anomaly, and the Eu anomaly. RDVs provide a simplified, univariate presentation of multivariate REE data, quantitatively capturing key features of the REE profile, the plot of reference-normalized concentrations versus atomic number. The Post Archaean Average Shale (PAAS) of Nance and Taylor [3] was used for normalization.

The Ce anomaly (Ce_N^*), the ratio of the normalized Ce concentration to an expected normalized value from interpolation of normalized La and Pr

concentrations (calculated in this work with Equation 2.4), is commonly discussed in the REE literature.

$$Ce_N^* = \frac{2 \cdot [Ce]_N}{[La]_N + [Pr]_N} \tag{2.4}$$

The Eu anomaly (Eu_N^*) is calculated similarly using measured Eu concentrations and the concentrations of directly neighboring REE (Sm and Gd). Both anomalies are commonly log-transformed, a convention that was applied throughout this work.

The ratios of reference normalized REE concentrations were here taken to be an average over multiple elements as opposed to single weight group representatives [42]. The LREE included in ratio analyses were La, Pr, Nd, and Sm; the MREE were Gd, Tb, and Dy; and the HREE were Ho, Er, Tm, Yb, and Lu. Both Ce and Eu were excluded from calculation of ratios.

These anomalies and interelement ratios cannot be reliably calculated for missing or censored data, thus our analyses of RDVs pertain only to the portion of the dataset where all necessary REE have been reported. This excludes many datasets utilizing isotope dilution mass spectrometry, which cannot measure mono-isotopic elements, and many studies with higher detection limits or poor analytical precision.

2.2.4   Geochemical modelling

A primary goal was to examine the relationship of REE abundance with various bulk solution chemistries in groundwater. The REE abundance and concentration data from the literature varied by their sample chemical characterization, water sample origin, sampling location, as well as the nomenclature used to classify geologic environments. To overcome this heterogeneity we employed a quantitative criterion for the classification of data before any statistical analysis was performed. For groundwater samples, the total dataset ($N = 619$ samples) was screened under the criterion that the samples were fully characterized for major ions (Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, and HCO_3^-/CO_3^{2-}) and pH. These major solution chemistry data were used to calculate ionic strength and activity of the carbonate ion (CO_3^{2-}) using the geochemical equilibrium model PHREEQC with the Pitzer database

(pitzer.dat) [59, 60]. Considering the wide range of unique systems represented by this dataset we have not pursued modeling of REE solubility limiting reactions. While such analyses are useful in well-defined systems, and REE speciation is commonly used to investigate mechanisms of REE abundance [40, 61, 62], significantly more information than was reported for most studies would be needed to conduct the analyses and moreover this type of modeling was deemed beyond the scope of the study.

2.3 Results and discussion

2.3.1 Occurrence of REE in aqueous media

Measured concentrations of REE in natural waters are summarized in Figure 2.3. Reported concentrations of REE in surface- and groundwater spanned ten orders of magnitude, while individual elements fluctuate with atomic number (i.e., the observed "zig-zag" pattern). This fluctuation of abundance is attributable to the Oddo-Harkins rule, which stipulates that, other than hydrogen, elements with even atomic numbers are more abundant than adjacent elements having odd atomic numbers [63]; this fluctuation gives rise to the common practice of normalizing REE data to a reference standard, for example an average shale or chondritic meteorite, which also exhibits the Oddo-Harkins effect [64]. Individual REE are distributed log normally (Figure REF) and are highly correlated with one-another (Figure 2.4).

Groundwater

In groundwater, the KM-calculated interquartile ranges (KM-IQR, first and third quartile) of overall REE concentrations vary between 5.7 and 410 pmol/kg solution (pmol/kg; Figure 2.3A). The mean concentration of the groundwater dataset was 3.5×10^4 pmol/kg and the median was 53 pmol/kg. As shown in the disparity between the mean and median, REE concentrations in groundwater were heavily right-tailed with many "outlier" values and a maximum of 9.7×10^6 pmol/kg [65]. The REE profile of a groundwater is thought to arise either from water-rock interactions [40, 66] or by changes in the aqueous chemistry [33], e.g., in relation to the residence time within the aquifer [18, 19].
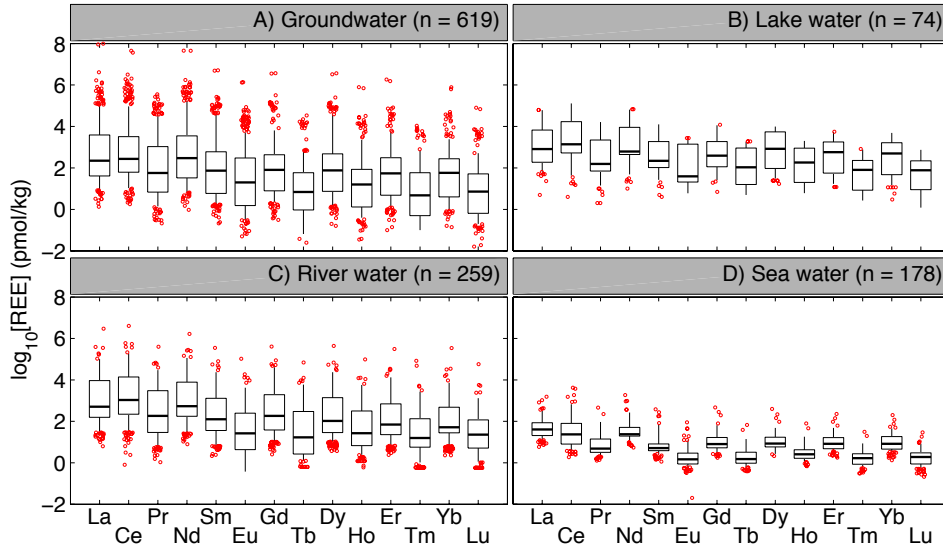
Figure 2.3: REE concentration distributions in waters of different types. Box-plots are based on source-size weighted Kaplan-Meier estimates of REE concentration percentiles and represent the median (thick, black line), inter-quartile range (IQR; first to third quartile; boxed range), 5th to 95th percentile range (whiskers), and remaining outliers (red circles). In each subfigure heading, "n" in the parenthesis denotes the total number of measurements in each water type The number of detected measurements varies between elements. These data are inclusive, representing unique water sources, spatial and temporal variations within an individual water, and any duplicate analyses.
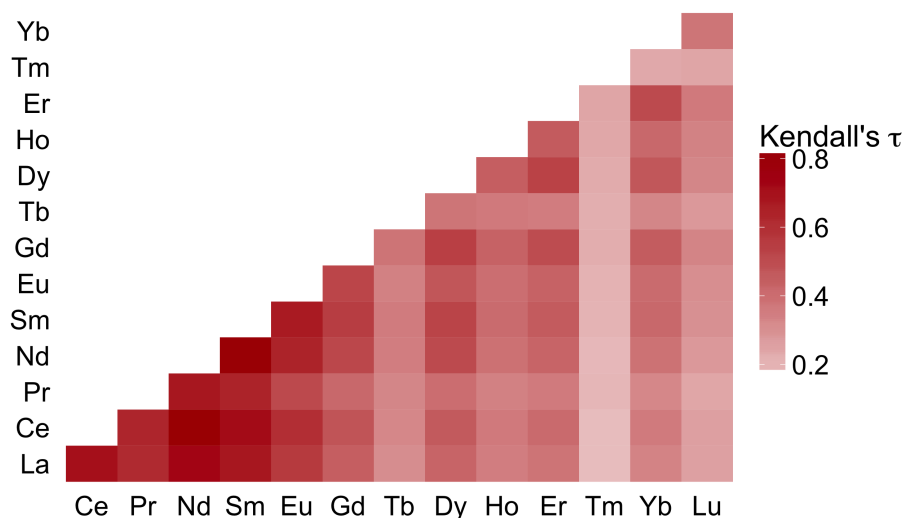
Figure 2.4: Heatmap visualization of Kendall's τ correlation coefficients among the REE in the groundwater dataset. All correlations are statistically significantly positive.

By comparison, substitution of censored data between 0 and 100% of the detection limit values (without source weighting) results in medians ranging from 4.4 to 13 pmol/kg and mean values ranging from 4.7×10^4 to 4.9×10^4 pmol/kg. Alternatively, deletion of below detection data roughly doubles all KM summary statistic estimates: 120 and 5.3×10^4 pmol/kg for median and mean respectively. Linear interpolation to fill in censored data also leads to significant error (Figure 2.5). This highlights the criticality of applying appropriate tools in calculating relevant summary statistics, particularly for quantitative analysis.

Terrestrial surface water

In lake water, the KM-IQR of overall REE concentrations varied between 30 and 1.4×10^3 pmol/kg (Figure 2.3B). The mean concentration of the lake water dataset was 3.8×10^3 pmol/kg and the median was 170 pmol/kg.

In river water, the KM-IQRs of overall REE concentrations varied between 15 and 270 pmol/kg (Figure 2.3C). The mean concentration of the river water dataset was 3.3×10^3 pmol/kg and the median was 71 pmol/kg.
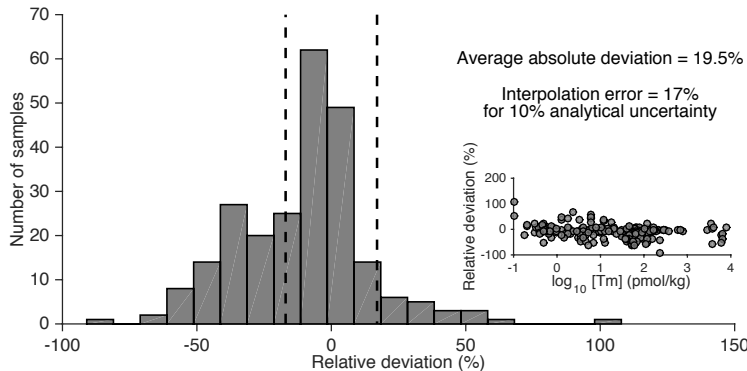
Figure 2.5: Histogram of relative Post Archaean Averaged Shale normalized interpolation error of Tm in groundwater dataset and (inset) dependence of error on measured Tm values. PAAS normalized Er and Yb were used to "guess" the Tm concentration and then compared to the actual measured Tm concentrations. Dashed lines mark the bounds of 10% analytical uncertainty propagated through the interpolation. In this analysis 42% of 241 samples exhibit interpolation error outside the bounds of analytical uncertainty.

Seawater

In seawater, the KM-IQR of reported REE concentrations varied between 1.6 and 13 pmol/kg (Figure 2.3D). The mean concentration of the seawater dataset was 19 pmol/kg and the median was 5 pmol/kg. The disparity between the dissolved REE in seawater and its influent waters is likely attributable to particulate scavenging [67, 68] at a rate that exceeds mass fluxes from surface- and groundwater sources. In groundwater by comparison, continuous weathering of the aquifer rock counteracts sorption and co-precipitation loss mechanisms. Dubinin [69] notes that dissolved fractions of REE increase with distance off-shore as terrigenous material settles, which removes REE mass faster than it can be replenished by other sources such as aeolian deposition or hydrothermal venting

Shale-normalized anomalies and ratios

The groundwaters in this dataset primarily exhibited mild negative and positive shale-normalized anomalies for Ce and Eu respectively, though, as with abundance, these RDVs varied widely (Figure 2.6). Both terrestrial surface waters reported here exhibit similar Ce* and Eu* (Figure 2.6A and 2.6B),

which are mildly negative and nearly zero respectively. The Ce* of seawater is starkly different that any of the three water types, developing consistent, negative anomalies with instances of large positive anomalies.

Nearly all waters in this study are LREE depleted (Figure 2.6C and 2.6D). As with element anomalies, the mean and IQR of interelement ratios among groundwaters and terrestrial surface waters are similar while seawater exhibits distinctly lower ratios. Both anomalies in groundwater exhibit large outlier values, similar to REE abundance, which underscores the extreme variability in groundwater systems.
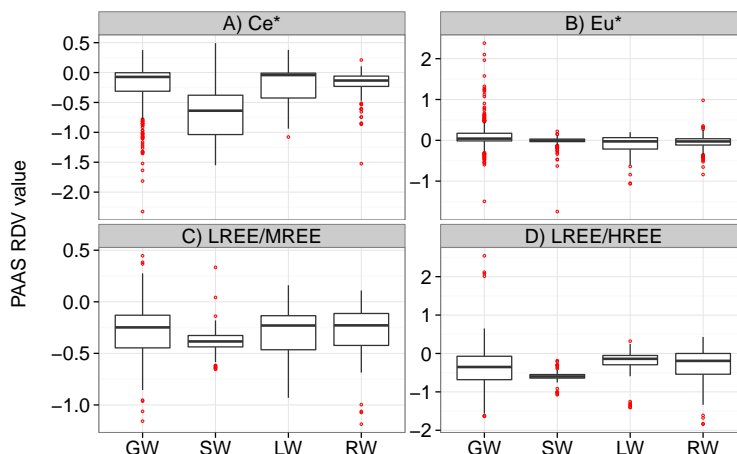
Figure 2.6: Log-transformed, Post-Archaean Average Shale (PAAS)- normalized, reduced dimension variables (RDVs) calculated for groundwater (GW), seawater (SW), lake water (LW), and river water (RW) data. RDVs are the cerium anomaly (Ce*), europium anomaly (Eu*), LREE / MREE ratio, and LREE / HREE ratio. Boxplots represent the median (thick, black line), interquartile range (IQR; first to third quartile; boxed range), ± 1.5 times the IQR (whiskers), and remaining outliers (red circles).

The anomalies of Ce and Eu are of interest because they are the two REE with redox lability within the Eh-pH regime of water stability [70]. Upon oxidation – Ce(III) to Ce(IV) – cerium readily precipitates as $CeO_2(s)$; alternatively, europium becomes less soluble when reduced from Eu(III) to Eu(II) based on preferential incorporation into other minerals due to the lower valency and higher ionic radius of the reduced state (similar to Sr) [15]. Where anomalies in rock samples have been used to infer paleo-redox conditions, the anomalies in aqueous samples are often used to analyze water-rock in-

teractions [40, 66].

Divalent europium is energetically favorable under extremely reducing conditions or at high temperatures. Reduced Eu(II) exhibits strong partitioning to various minerals during crystallization from magmatic fluids. Thus many source rocks are either highly enriched (e.g. plagioclase [71]) or highly depleted (e.g. olivine [72] and pyroxene [4]) in Eu, enhancing or limiting availability during weathering [73]. In hot, reduced, alkaline waters fracture filling calcite commonly contains notable Eu enrichments indicating that the source fluids developed negative Eu* [74, 75]. However, such environments are representative of hydrothermal waters [73], which have specifically excluded from this analysis. Thus the Eu* in this dataset are thought to result from mineral weathering [73].

While aquatic redox couples are commonly in disequilibrium for measured redox potential (Eh) [76], Figure 2.7A illustrates a negative correlation ($\rho = -0.33$; $P < 1 \times 10^{-6}$) between Ce_N^* and Eh in groundwater, as expected for oxidative scavenging. However, the low strength of the correlation reflects the mechanism being a function of both source composition and aqueous chemistry. Europium anomalies do not exhibit the expected positive trend with redox potential in groundwater, though a weak negative correlation is observed ($\rho = -0.18$, $P = 0.006$; Figure 2.7B). Stability diagrams with measure Ce* and Eu* are included in Figure 2.8 and show the lack of agreement between measured and thermodynamically predicted anomalies.

2.3.2   REE abundance trends in groundwater

Of all the bulk solution chemistry parameters for groundwater, pH exerted the greatest control as an independent variable over dissolved REE abundance (Figure 2.9). Generally, more acidic waters contained the most REE, either via acidification-enhanced weathering or from an enrichment of REE in the acid source [33, 65, 66, 77–80]. REE are effectively scavenged from more neutral and basic waters via sorption onto oxides and clays or coprecipitation with carbonates and phosphates by replacing calcium, which has a comparable atomic radius [81–86]. Over the pH range 4-8, dissolved REE abundance follows an approximate -1 slope with increasing pH (95% confidence interval of -1.02 ± 0.1), indicative of a single-proton (i.e. monodentate) surface complexation reaction or calcite precipitation when carbonate far exceeds calcium [87]. Geochemical modeling showed that nearly all
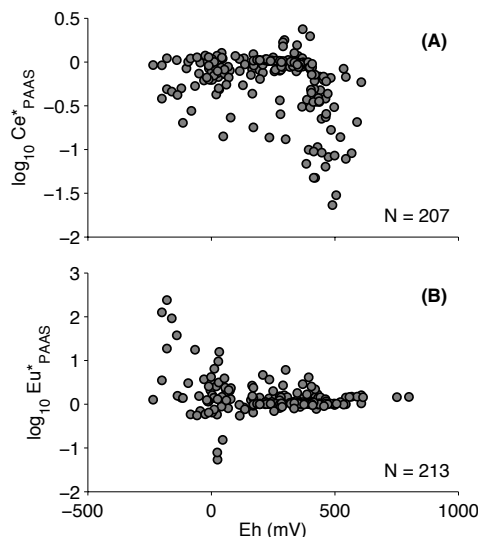
Figure 2.7: Groundwater Post-Archaean Average Shale (PAAS) Ce anomaly (Ce*, A, Eq. 2.4) and Eu anomaly (Eu*, B) as a function of reported redox potential (Eh). Sample size denoted by "N".

samples were significantly undersaturated with respect to either pure-phase REE hydroxides or carbonates (Figure 2.10).

Strong complexation with carbonate, particularly as a dicarbonato complex $(\mathrm{REE(CO_3)_2^-})$ is thought to account for the apparent amphoteric behavior as pH becomes more [6, 32, 62]. The REE form progressively stronger carbonate and dicarbonato complexes with increasing atomic number [6, 32], which would suggest an increase of MREE or HREE relative to LREE as a function of pH or carbonate ion activity. However, such a trend is not observed within the current dataset (Figure 2.11) though the breadth of the current dataset may simply obscure such trends seen in individual studies [e.g. Ref. 14].

Kendall's τ correlation analysis of REE with bulk solutes shows inconsistent trends, however most of the meaningful relationships are between the solutes and the LREE and are weakly negative (Figure 2.12). This behavior is contrary to many other metals that correlate positively with dissolved solids. This likely highlights the heterogeneity of the dataset and the importance of surface reactions in controlling dissolved REE abundance.
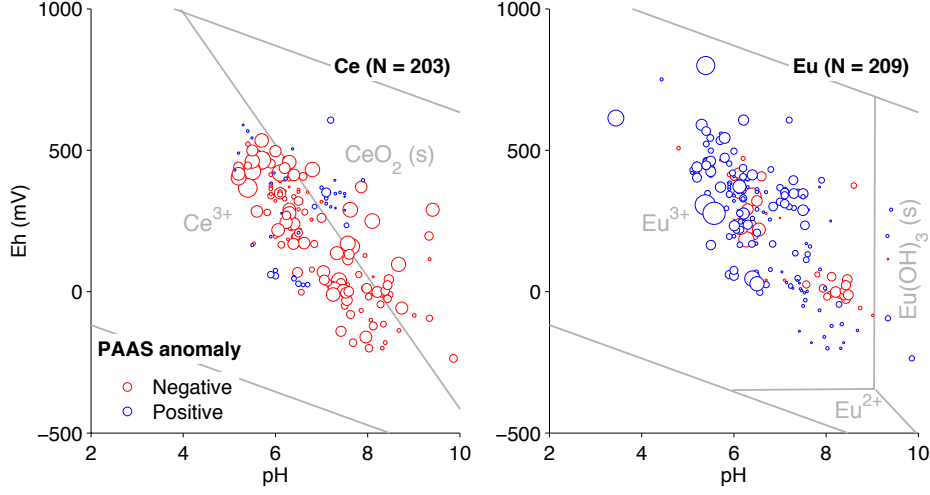
Figure 2.8: Post-Archaean Average Shale (PAAS) Cerium (Ce; left) and Europium (Eu; right) anomalies as functions of redox potential (Eh) and pH. Marker color corresponds to the sign of the anomaly and marker size corresponds to the magnitude of the anomaly. Stability fields, reproduced from Brookins [15], for relevant species in the REE-H_2O system underlie the data for $TOTCe=TOTEu = 10^{-11}$ M at 298 K
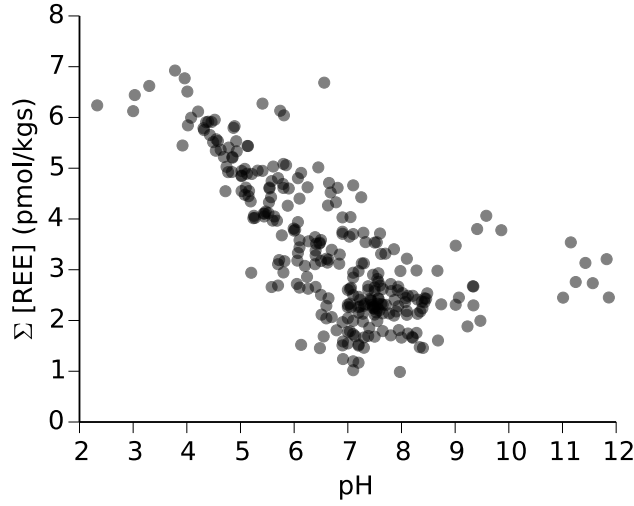
Figure 2.9: Sum of dissolved REE vs pH for groundwater data from Kaplan-Meier estimates (Eq 2.3). Markers are semitransparent to illustrate data density. Data presented for 281 samples.
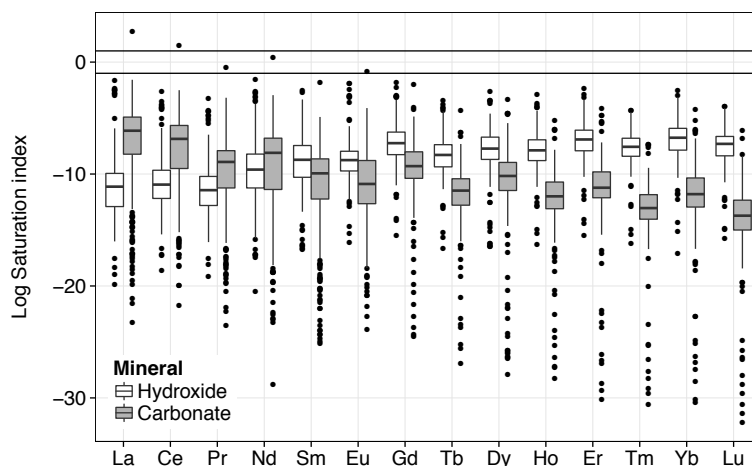
Figure 2.10: Modeled, log-transformed saturation indices of pure phase REE hydroxides and carbonates for groundwater data. Saturation indices calculated using PHREEQC and the LLNL database. Solid lines at -1 and 1 indicate range of potentially solubility controlling phases [88].

Major element chemistry of the groundwater data

For the groundwater data, the screening process reduced the total number of samples from 619 to 328. The representative chemistries of the screened dataset samples varied widely, with several orders of magnitude range in the major ion concentrations (Figure 2.13). Plotting the major ion concentrations using a Piper diagram [89] showed that the major anions were predominantly Cl/SO_4, while the predominant cations varied between monovalent and divalent (Figure 2.14).

Based on the available data three qualitative, rule-based water types were classified: fresh water ($N = 122$; $6.5 \leq pH \leq 8.5$; $I \leq 0.1$ mol/kg), brines ($N = 20$; $5.3 \leq pH \leq 8.0$; $I \geq 1.0$ mol/kg), and acidic waters ($N = 63$; pH < 5.3); ionic strength and pH values were chosen to maximize data points within the various water classes while maintaining a reasonable bound on parameters. Further subdivision of the freshwater data was possible (e.g. based on relative cation/anion composition) but was not pursued in favor of retaining a larger sample size. The medians of the REE distributions of the classified groundwaters are shown in Figure 2.15.

Utility of REE as tracers is reliant on identifying distinctive trends in different water types. Of note in Figure 2.15 is the markedly higher concentrations
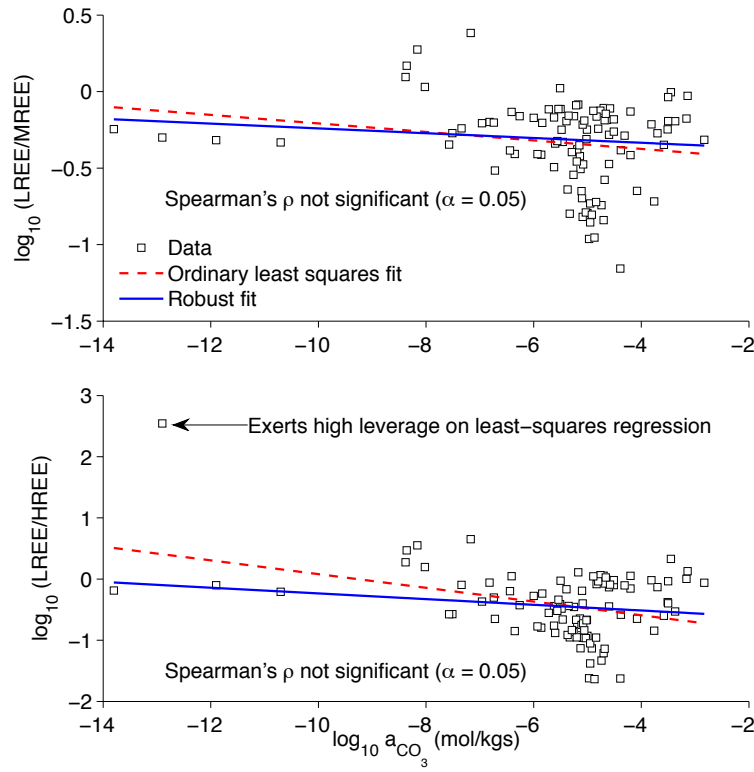
Figure 2.11: Light-middle REE fractionation (top) and light-heavy REE fractionation (bottom) as a function of modeled carbonate ion activity (a_{CO_3}). For L/M fractionation neither regression is significant, while L/H ordinary least squares regression is significant due to a high outlier value. Each plot contains 99 data points.
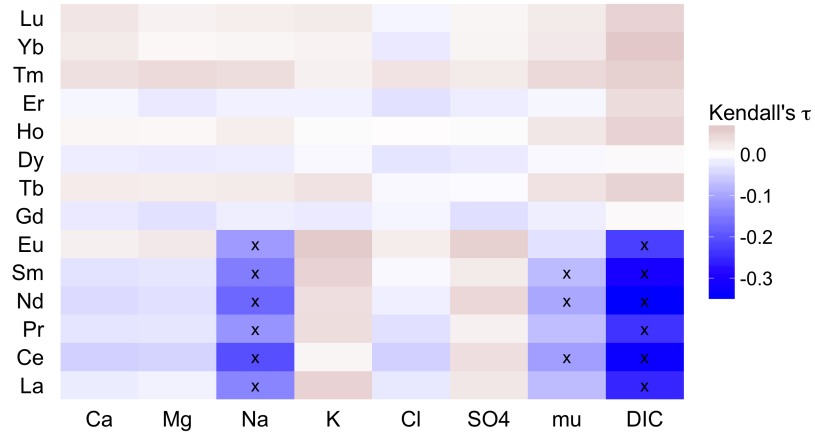
Figure 2.12: Heatmap visualization of Kendall's τ coefficients between REE abundance and total bulk solutes and calculated ionic strength ("mu") in groundwater. Statistically significant correlations ($\alpha = 0.05$) are noted with an "x".
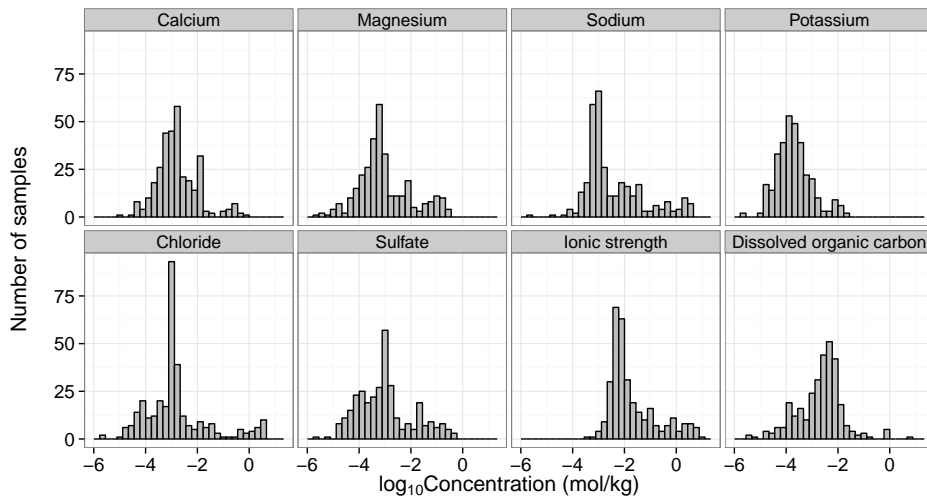
Figure 2.13: Major solution chemistry histograms for groundwater dataset. Total dissolved inorganic carbon was either given or estimated from alkalinity data using PHREEQC. Ionic strength was calculated using PHREEQC.
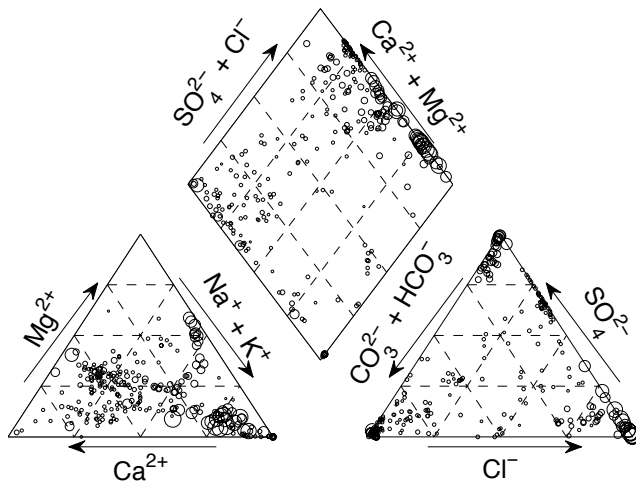
Figure 2.14: Piper diagram of ground water chemistry for circumneutral (5.5 < pH < 8.5) groundwater. Points scaled based on total dissolved solids. All plots range from 0 – 100%.

of REE in acidic waters than either fresh water or brines, which supports the observations from Figure 2.9, but also a deviation from the standard Oddo-Harkins effect pattern where Er and Yb are both less concentrated than their odd-numbered neighbors. This effect was attributed to high degrees of censored or missing data for these elements (87% for both Er and Yb in the acidic dataset) which leads to low-biased percentile estimates (as with Er and Yb) and large uncertainty (seen in all the HREE). Despite showing little or no correlation with ionic strength (Figure 2.12) the REE, in particular the LREE and MREE, were more concentrated in brines than freshwater.

Understanding of the aqueous systematics of the REE is dominated by studies of fresh groundwaters, surface waters, and seawater. Only 14% of the groundwater data gathered had a calculated ionic strength of 1.0 mol/kg or greater (Figure 2.16). While thermodynamic models have been developed to predict REE aqueous speciation in brines [6], the energetics of precipitation-dissolution and sorption-desorption in the REE system are not well defined for high ionic strength, chemically complex brines [85]. Moreover, increased utilization of groundwater [90] in response to population growth, economic development, and climate change and the potential to exploit saline aquifers [91] underscores the need to better understand the trace metal chemistry of
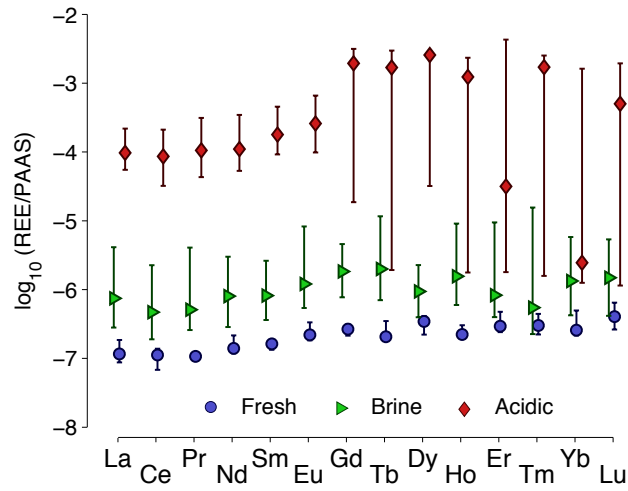
Figure 2.15: Source-size weighted Kaplan-Meier estimates of median dissolved REE concentrations, normalized to Post-Archaean Average Shale (PAAS) values, in three characteristic groundwater types. Error bars represent the 95% confidence interval of the median. Definitions of water classes are found in the text. Estimates were made for individual elements, thus results do not represent a "typical" water sample across the REE suite.

brines.
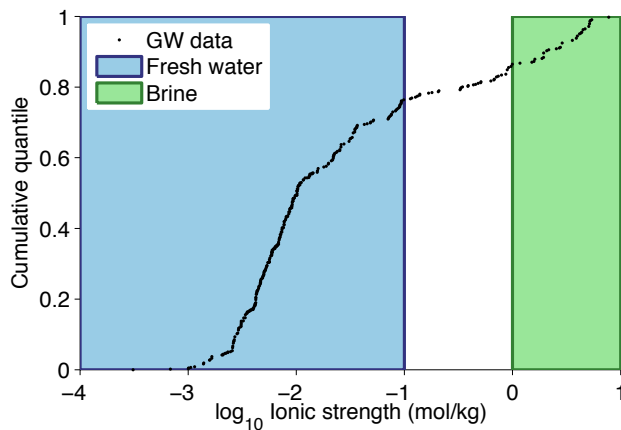
Figure 2.16: Source-size weighted, empirical cumulative distribution of calculated ionic strength in groundwater data set. Ionic strength calculated using PHREEQC [59].

Comparison of the HREE/MREE ratio to the MREE/LREE ratio succinctly captures the general shape of the REE profile of a sample, greatly simplifying visual comparison of samples. Figure 2.17 plots this scatter for classified groundwater data where all 14 REE were detected (39% of the dataset). For freshwater, more than 90% of the samples exhibited MREE/LREE ratios greater than 1, with a median ratio of 1.74 and an IQR from 1.36 to 2.64. HREE-MREE fractionation was more balanced with 53% of samples having HREE/MREE ratios greater than 1, with a median ratio of 1.25 and an IQR from 0.64 to 1.6. These data closely resemble the trends observed in the unclassified data set (not pictured), likely because freshwaters account for a large portion of the total data. Fewer brine and acidic samples are available to plot, however brines appear to cluster with HREE enrichment, with all samples having MREE/LREE ratios greater than 1.0 and only 1 sample having HREE/MREE ratio less than 1.0, while acid samples exhibit convex-down profiles (MREE/LREE > 1 and HREE/MREE < 1). However, it is difficult to ensure these are representative patterns given the small sample size for these water classes.
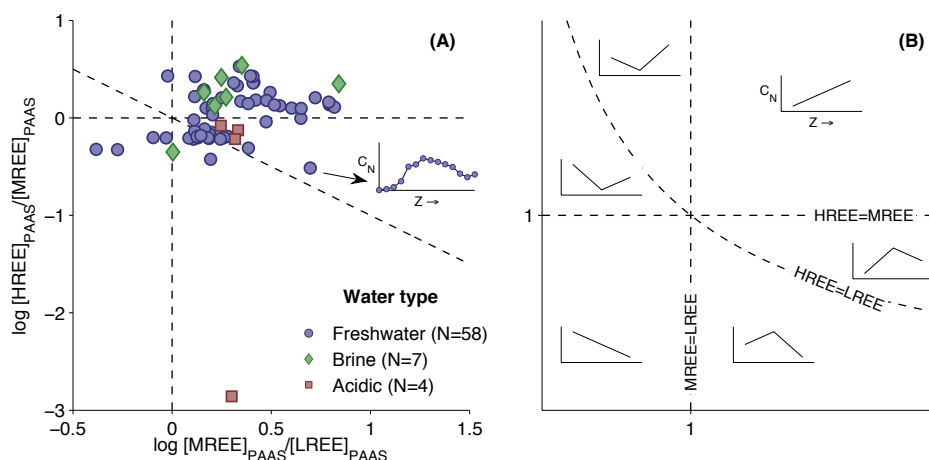
Figure 2.17: (A) Averaged Post-Archaean Average Shale (PAAS) element ratio biplot for chemistry-classified groundwater data with no censoring, and (B) a schematic intended for interpretation, illustrating the general shape of the REE profile for a given sample. Because each marker in (A) is a summary of the REE profile for an individual sample, an inset has been included to show the raw data (PAAS-normalized concentration, C_N, vs. elements in order of increasing atomic number, Z) for one sample. The schematic in (B) is modified from Stolpe et al. [42] with permission. Copyright 2013 Elsevier. Note that the values in (A) have been log-transformed while the axes of the schematic are linear.

2.3.3 Limitations of assembled dataset

The primary goal of this analysis was to understand large-scale REE variability and trends. However, researchers have understood that, because of complex geologies and watershed characteristics, there is significant merit in studying local, small-scale REE systematics. This leads to specialized studies, for example, on the importance of colloids,[33, 35, 42, 84, 87] phosphate complexation,[12, 21, 91] or organic compounds.[41, 85, 132-134] While this approach yields crucial understanding of these processes, it also leads to a paucity of consistent data for inter-study comparison. Moreover, it could motivate spurious extrapolation of thermodynamic or case study data to unique systems.

References

[1] Richard W. Murray, Marilyn R. Buchholtz ten Brink, David L. Jones, David C. Gerlach, and G.Price Russ. Rare earth elements as indicators of different marine depositional environments in chert and shale. *Geology*, 18(3):268–271, 1990.

[2] C Laveuf and S Cornu. A review on the potentiality of rare earth elements to trace pedogenetic processes. *Geoderma*, 154(1):1–12, 2009.

[3] W. B. Nance and S. R. Taylor. Rare earth element patterns and crustal evolution—i. australian post-archean sedimentary rocks. *Geochimica et Cosmochimica Acta*, 40(12):1539–1551, 1976.

[4] G. N. Hanson. Rare-earth elements in petrogenetic studies of igneous systems. *Annual Review of Earth and Planetary Sciences*, 8:371–406, 1980.

[5] K. B. Krauskopf. Thorium and rare-earth metals as analogs for actinide elements. *Chemical Geology*, 55(3–4):323–335, 1986.

[6] Frank J. Millero. Stability constants for the formation of rare earth-inorganic complexes as a function of ionic strength. *Geochimica et Cosmochimica Acta*, 56(8):3123–3132, 1992.

[7] H. J. W. De Baar, M. P. Bacon, and P. G. Brewer. Rare-earth distributions with a positive Ce anomaly in the western North-Atlantic ocean. *Nature*, 301(5898):324–327, 1983.

[8] H. Elderfield, M. Whitfield, J. D. Burton, M. P. Bacon, and P. S. Liss. The oceanic chemistry of the rare-earth elements [and discussion].

Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 325(1583):105–126, 1988.

[9] Pranav Kulkarni, Shankararaman Chellam, and Matthew P Fraser. Lanthanum and lanthanides in atmospheric fine particles and their apportionment to refinery and petrochemical operations in Houston, TX. *Atmospheric Environment*, 40(3):508–520, 2006.

[10] Pranav Kulkarni, Shankararaman Chellam, and Matthew P Fraser. Tracking petroleum refinery emission events using lanthanum and lanthanides as elemental markers for $pm_{2.5}$. *Environmental Science & Technology*, 41(19):6748–6754, 2007.

[11] Stephen B. Castor and James B. Hedrick. *Rare Earth Elements*, pages 769–792. SME, 7 edition, 2006.

[12] Robyn E. Hannigan and Edward R. Sholkovitz. The development of middle rare earth element enrichments in freshwaters: weathering of phosphate minerals. *Chemical Geology*, 175(3–4):495–508, 2001.

[13] Jianwu Tang and Karen H. Johannesson. Rare earth elements adsorption onto carrizo sand: Influence of strong solution complexation. *Chemical Geology*, 279(3–4):120–133, 2010.

[14] Hyeon-Su Choi, Seong-Taek Yun, Yong-Kwon Koh, Bernhard Mayer, Seong-Sook Park, and Ian Hutcheon. Geochemical behavior of rare earth elements during the evolution of co2-rich groundwater: A study from the kangwon district, south korea. *Chemical Geology*, 262(3–4): 318–327, 2009.

[15] D. G. Brookins. Aqueous geochemistry of rare earth elements. *Reviews in Mineralogy and Geochemistry*, 21(1):201–225, 1989.

[16] Michael Bau and Peter Dulski. Anthropogenic origin of positive gadolinium anomalies in river waters. *Earth and Planetary Science Letters*, 143 (1):245–255, 1996.

[17] Kevin H. Johannesson, Klaus J. Stetzenbach, and Vernon F. Hodge. Rare earth elements as geochemical tracers of regional groundwater mixing. *Geochimica et Cosmochimica Acta*, 61(17):3605–3618, 1997.

[18] Kevin H. Johannesson, Klaus J. Stetzenbach, Vernon F. Hodge, David K. Kreamer, and Xiaoping Zhou. Delineation of ground-water flow systems in the southern great basin using aqueous rare earth element distributions. *Ground Water*, 35(5):807–819, 1997.

[19] S. Bwire Ojiambo, W. Berry Lyons, Kathy A. Welch, Robert J. Poreda, and Karen H. Johannesson. Strontium isotopes and rare earth elements as tracers of groundwater–lake water interactions, lake naivasha, kenya. *Applied Geochemistry*, 18(11):1789–1805, 2003.

[20] Christian Siebert, Eliahu Rosenthal, Peter Möller, Tino Rödiger, and Miki Meiler. The hydrochemical identification of groundwater flowing to the bet she'an-harod multiaquifer system (lower jordan valley) by rare earth elements, yttrium, stable isotopes (h, o) and tritium. *Applied Geochemistry*, 27(3):703–714, 2012.

[21] Kevin H. Johannesson, Klaus J. Stetzenbach, Vernon F. Hodge, and W. Berry Lyons. Rare earth element complexation behavior in circum-neutral ph groundwaters: Assessing the role of carbonate and phosphate ions. *Earth and Planetary Science Letters*, 139(1–2):305–319, 1996.

[22] M. Tesmer, P. Möller, S. Wieland, C. Jahnke, H. Voigt, and A. Pekdeger. Deep reaching fluid flow in the North East German Basin: origin and processes of groundwater salinisation. *Hydrogeology Journal*, 15(7):1291–1306, 2007.

[23] Sally M Benson and David R Cole. Co_2 sequestration in deep sedimentary formations. *Elements*, 4(5):325–331, 2008.

[24] Athanasios K Karamalidis, Sharon G Torres, J Alexandra Hakala, Hongbo Shao, Kirk J Cantrell, and Susan Carroll. Trace metal source terms in carbon sequestration environments. *Environmental Science & Technology*, 47(1):322–329, 2012.

[25] S. Chaudhuri, M. Totten, N. Clauer, J. Miesse, G. Riepl, S. Massie, and K. Semhi. A study yielding the first demonstration that rare-earth elements could be a useful geochemical tracer in formation hydraulic fracturing schemes for enhanced gas and oil production. *The Journal of the Oklahoma City Geological Society*, pages 214–223, 2011.

[26] Katrina Cheung, Hamed Sanei, Patrick Klassen, Bernhard Mayer, and Fariborz Goodarzi. Produced fluids and shallow groundwater in coalbed methane (cbm) producing regions of alberta, canada: Trace element and rare earth element geochemistry. *International Journal of Coal Geology*, 77(3–4):338–349, 2009.

[27] Serkan Kulaksız and Michael Bau. Rare earth elements in the rhine river, germany: First case of anthropogenic lanthanum as a dissolved

microcontaminant in the hydrosphere. *Environment International*, 37 (5):973–979, 2011.

[28] Serkan Kulaksiz and Michael Bau. Anthropogenic dissolved and colloid/nanoparticle-bound samarium, lanthanum and gadolinium in the rhine river and the impending destruction of the natural rare earth element distribution in rivers. *Earth and Planetary Science Letters*, 362:43–50, 2013.

[29] J Palarea-Albaladejo, JA Martín-Fernández, and RA Olea. Non-detect bootstrap method for estimating distributional parameters of compositional samples revisited: a multivariate approach. In *Proceedings of the 4th international workshop on compositional data analysis, San Feliu de Guixols, Spain*, pages 1–9, 2011.

[30] Dennis R Helsel. Less than obvious-statistical treatment of data below the detection limit. *Environmental Science & Technology*, 24(12):1766–1774, 1990.

[31] Dennis R Helsel. More than obvious: better methods for interpreting nondetect data. *Environmental Science & Technology*, 39(20):419–423, 2005.

[32] Scott A. Wood. The aqueous geochemistry of the rare-earth elements and yttrium 1. Review of available low-temperature data for inorganic complexes and the inorganic REE speciation of natural waters. *Chemical Geology*, 82:159–186, 1990.

[33] Aline Dia, Gérard Gruau, Gwenaelle Olivé-Lauquet, Christine Riou, Jérome Molénat, and Pierre Curmi. The distribution of rare earth elements in groundwaters: assessing the role of source-rock composition, redox changes and colloidal particles. *Geochimica et Cosmochimica Acta*, 64(24):4131–4151, 2000.

[34] Gérard Gruau, Aline Dia, Gwenaelle Olivié-Lauquet, Mélanie Davranche, and Gilles Pinay. Controls on the distribution of rare earth elements in shallow groundwaters. *Water Research*, 38(16):3576–3586, 2004.

[35] Olivier Pourret, Gérard Gruau, Aline Dia, Mélanie Davranche, and Jérôme Molénat. Colloidal control on the distribution of rare earth elements in shallow groundwaters. *Aquatic Geochemistry*, 16(1):31–59, 2010.

[36] Lin Ma, Lixin Jin, and Susan L Brantley. How mineralogy and slope aspect affect REE release and fractionation during shale weathering in the Susquehanna/Shale Hills Critical Zone Observatory. *Chemical Geology*, 290(1):31–49, 2011.

[37] Kevin H. Johannesson, Irene M. Farnham, Caixia Guo, and Klaus J. Stetzenbach. Rare earth element fractionation and concentration variations along a groundwater flow path within a shallow, basin-fill aquifer, southern nevada, usa. *Geochimica et Cosmochimica Acta*, 63(18):2697–2708, 1999.

[38] Kevin H. Johannesson, Xiaoping Zhou, Caixia Guo, Klaus J. Stetzenbach, and Vernon F. Hodge. Origin of rare earth element signatures in groundwaters of circumneutral ph from southern Nevada and eastern California, USA. *Chemical Geology*, 164(3–4):239–257, 2000.

[39] Jianwu Tang and Karen H. Johannesson. Controls on the geochemistry of rare earth elements along a groundwater flow path in the carrizo sand aquifer, texas, usa. *Chemical Geology*, 225(1–2):156–171, 2006.

[40] Stephanie S. Willis and Karen H. Johannesson. Controls on the geochemistry of rare earth elements in sediments and groundwaters of the Aquia aquifer, Maryland, USA. *Chemical Geology*, 285(1–4):32–49, 2011.

[41] Olivier Pourret, Mélanie Davranche, Gérard Gruau, and Aline Dia. Competition between humic acid and carbonates for rare earth elements complexation. *Journal of Colloid and Interface Science*, 305(1):25–31, 2007.

[42] Björn Stolpe, Laodong Guo, and Alan M. Shiller. Binding and transport of rare earth elements by organic and iron-rich nanocolloids in Alaskan rivers, as revealed by field-flow fractionation and ICP-MS. *Geochimica et Cosmochimica Acta*, 106:446–462, 2013.

[43] J Borrego, B Carro, N Lopez-Gonzalez, J de la Rosa, JA Grande, T Gomez, and ML de la Torre. Effect of acid mine drainage on dissolved rare earth elements geochemistry along a fluvial-estuarine system: the tinto-odiel estuary (sw spain). *Hydrology Research*, 43(3):262–274, 2012.

[44] Elke Bozau, Marc Leblanc, Jean Luc Seidel, and Hans-Joachim Stärk. Light rare earth elements enrichment in an acidic mine lake (lusatia, germany). *Applied Geochemistry*, 19(3):261–271, 2004.

[45] F. Doulati Ardejanii, R. Rooki, B. Jodieri Shokri, T. Eslam Kish, A. Aryafar, and P. Tourani. Prediction of rare earth elements in neutral alkaline mine drainage from razi coal mine, golestan province, northeast iran, using general regression neural network. *Journal of Environmental Engineering*, 139(6):896–907, 2013.

[46] Francisco Martín Romero, Rosa María Prol-Ledesma, Carles Canet, Laura Núñez Alvares, and Ramón Pérez-Vázquez. Acid drainage at the inactive santa lucia mine, western cuba: Natural attenuation of arsenic, barium and lead, and geochemical behavior of rare earth elements. *Applied Geochemistry*, 25(5):716–727, 2010.

[47] Philip L. Verplanck, D. Kirk Nordstrom, Howard E. Taylor, and Briant A. Kimball. Rare earth element partitioning between hydrous ferric oxides and acid mine water during iron oxidation. *Applied Geochemistry*, 19(8):1339–1354, 2004.

[48] F. Worrall and D. G. Pearson. Water-rock interaction in an acidic mine discharge as indicated by rare earth element patterns. *Geochimica et Cosmochimica Acta*, 65(18):3027–3040, 2001.

[49] Fenghua Zhao, Zhiyuan Cong, Hongfu Sun, and Deyi Ren. The geochemistry of rare earth elements (REE) in acid mine drainage from the sitai coal mine, shanxi province, north china. *International Journal of Coal Geology*, 70(1–3):184–192, 2007.

[50] Johnson R. Haas, Everett L. Shock, and David C. Sassani. Rare earth elements in hydrothermal systems: Estimates of standard partial molal thermodynamic properties of aqueous complexes of the rare earth elements at high pressures and temperatures. *Geochimica et Cosmochimica Acta*, 59(21):4329–4350, 1995.

[51] E.A. Lepel, J. C. Laul, and M.R. Smith. Rare-earth elements in hot brines (165 to 190°c) from the salton sea geothermal field, 1988.

[52] Annie Michard. Rare earth element systematics in hydrothermal fluids. *Geochimica et Cosmochimica Acta*, 53(3):745–750, 1989.

[53] Anita J Lewis, Martin R Palmer, Neil C Sturchio, and Anthony J Kemp. The rare earth element geochemistry of acid-sulphate and acid-sulphate-chloride geothermal systems from Yellowstone National Park, Wyoming, USA. *Geochimica et Cosmochimica Acta*, 61(4):695–706, 1997.

[54] A.J. Lewis, A. Komninou, B.W.D. Yardley, and M.R. Palmer. Rare earth element speciation in geothermal fluids from yellowstone national park, wyoming, usa. *Geochimica et Cosmochimica Acta*, 62(4):657–663, 1998.

[55] Tetsuya Sanada, Nobuki Takamatsu, and Yuzo Yoshiike. Geochemical interpretation of long-term variations in rare earth element concentrations in acidic hot spring waters from the Tamagawa geothermal area, Japan. *Geothermics*, 35(2):141–155, 2006.

[56] Anshuman Singh, Edward F. Neuhauser, Nicholas A. Azzolina, Mark Distler, Krista M. Anders, Melissa A. Doroski, and Alan J. Rabideau. Statistical techniques for analyzing of soil vapor intrusion data: A case study of manufactured gas plant sites. *J. Air Waste Manage. Assoc.*, 63(2), 2013.

[57] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.R-project.org/.

[58] D.R. Helsel. *Statistics for Censored Environmental Data Using Minitab and R*. Wiley, 2 edition, 2012.

[59] David L. Parkhurst and C.A.J. Appelo. PHREEQC, 1999.

[60] Kenneth S Pitzer. Thermodynamics of electrolytes. i. theoretical basis and general equations. *The Journal of Physical Chemistry*, 77(2):268–277, 1973.

[61] Kevin H. Johannesson, W. Berry Lyons, Mary A. Yelken, Henri E. Gaudette, and Klaus J. Stetzenbach. Geochemistry of the rare-earth elements in hypersaline and dilute acidic natural terrestrial waters: Complexation behavior and middle rare-earth element enrichments. *Chemical Geology*, 133(1–4):125–144, 1996.

[62] Kevin H. Johannesson, W. Berry Lyons, Klaus J. Stetzenbach, and Robert H. Byrne. The solubility control of rare earth elements in natural terrestrial waters and the significance of po43- and co32- in limiting dissolved rare earth concentrations: A review of recent information. *Aquatic Geochemistry*, 1(2):157–173, 1995.

[63] William D. Harkins. The evolution of the elements and the stability of complex atoms. I. A new periodic system which shows a relation between the abundance of the elements and the structure of the nuclei

of atoms. *Journal of the American Chemical Society*, 39(5):856–879, 1917.

[64] Scott M. McLennan. Rare earth element geochemistry and the "tetrad" effect. *Geochimica et Cosmochimica Acta*, 58(9):2025–2033, 1994.

[65] N. Miekeley, H. Coutinho de Jesus, C. L. Porto da Silveira, P. Linsalata, and R. Morse. Rare-earth elements in groundwaters from the osamu utsumi mine and morro do ferro analogue study sites, poços de caldas, brazil. *Journal of Geochemical Exploration*, 45(1–3):365–387, 1992.

[66] Pauline L. Smedley. The geochemistry of rare earth elements in groundwater from the Carnmenellis area, southwest England. *Geochimica et Cosmochimica Acta*, 55(10):2767–2779, 1991.

[67] Edward R. Sholkovitz, William M. Landing, and Brent L. Lewis. Ocean particle chemistry: The fractionation of rare earth elements between suspended particles and seawater. *Geochimica et Cosmochimica Acta*, 58(6):1567–1579, 1994.

[68] Yigal Erel and Edward M. Stolper. Modeling of rare-earth element partitioning between particles and solution in aquatic environments. *Geochimica et Cosmochimica Acta*, 57(3):513–518, 1993.

[69] A. V. Dubinin. Geochemistry of rare earth elements in the ocean. *Lithology and Mineral Resources*, 39(4):289–307, 2004.

[70] Douglas G Brookins. Eh-pH diagrams for the rare earth elements at 25 °c and one bar pressure. *Geochem. J*, 17(5):223–229, 1983.

[71] WP Nash and HR Crecraft. Partition coefficients for trace elements in silicic magmas. *Geochimica et Cosmochimica Acta*, 49(11):2309–2322, 1985.

[72] Stephen F. Foley and George A. Jenner. Trace element partitioning in lamproitic magmas - the Gaussberg olivine leucitite. *Lithos*, 75(1–2): 19–38, 2004.

[73] Dimitri A. Sverjensky. Europium redox equilibria in aqueous solution. *Earth and Planetary Science Letters*, 67(1):70–78, 1984.

[74] Seung-Gu Lee, Dae-Ha Lee, Yongje Kim, Byong-Gon Chae, Won-Young Kim, and Nam-Chil Woo. Rare earth elements as indicators of groundwater environment changes in a fractured rock system: evidence from fracture-filling calcite. *Applied Geochemistry*, 18(1):135–143, 2003.

[75] Chunfang Cai, Kaikai Li, Hongtao Li, and Baoshou Zhang. Evidence for cross formational hot brine flow from integrated ^{87}sr / ^{8}6sr, REE and fluid inclusions of the Ordovician veins in Central Tarim, China. *Applied Geochemistry*, 23(8):2226–2235, 2008.

[76] RD Linberg and DD Runnels. Ground water redox reactions: an analysis of equilibrium state applied to Eh measurements and geochemical modeling. *Science*, 225:925–927, 1984.

[77] David C. Gosselin, Monty R. Smith, Elwood A. Lepel, and J. C. Laul. Rare earth elements in chloride-rich groundwater, palo duro basin, texas, usa. *Geochimica et Cosmochimica Acta*, 56(4):1495–1505, 1992.

[78] René P. T. Janssen and Wilko Verweij. Geochemistry of some rare earth elements in groundwater, vierlingsbeek, the netherlands. *Water Research*, 37(6):1320–1350, 2003.

[79] Fred Worrall and D. G. Pearson. The development of acidic groundwaters in coal-bearing strata: Part i. rare earth element fingerprinting. *Applied Geochemistry*, 16(13):1465–1480, 2001.

[80] Matthew I. Leybourne, Wayne D. Goodfellow, Dan R. Boyle, and Gwendy M. Hall. Rapid development of negative ce anomalies in surface waters and contrasting REE patterns in groundwaters associated with zn–pb massive sulphide deposits. *Applied Geochemistry*, 15(6): 695–723, 2000.

[81] K. H. Johannesson and W. B. Lyons. The rare-earth element geochemistry of Mono Lake water and the importance of carbonate complexing. *Limnology and Oceanography*, 39(5):1141–1154, 1994.

[82] David R Lide. *CRC handbook of chemistry and physics*. CRC press, 2012. ISBN 1439880492.

[83] M. H. Bradbury and B. Baeyens. Sorption of eu on na- and ca-montmorillonites: experimental investigations and modelling with cation exchange and surface complexation. *Geochimica et Cosmochimica Acta*, 66(13):2325–2334, 2002.

[84] Frédéric Coppin, Gilles Berger, Andreas Bauer, Sylvie Castet, and Michel Loubet. Sorption of lanthanides on smectite and kaolinite. *Chemical Geology*, 182(1):57–68, 2002.

[85] Kelly A. Quinn, Robert H. Byrne, and Johan Schijf. Sorption of yttrium

and rare earth elements by amorphous ferric hydroxide: Influence of ph and ionic strength. *Marine Chemistry*, 99(1–4):128–150, 2006.

[86] Yasutaka Terakado and Akimasa Masuda. The coprecipitation of rare-earth elements with calcite and aragonite. *Chemical Geology*, 69(1–2): 103–110, 1988.

[87] François M.M. Morel and Janet G. Hering. *Principles and Applications of Aquatic Chemistry*. John Wiley & Sons, Inc., 1993.

[88] Jeannet A Meima and Rob NJ Comans. Geochemical modeling of weathering reactions in municipal solid waste incinerator bottom ash. *Environmental science & technology*, 31(5):1269–1276, 1997.

[89] Arthur M Piper. A graphic procedure in the geochemical interpretation of water-analyses. *Transactions, American Geophysical Union*, 25:914–928, 1944.

[90] Yu Wen Chen, Liang Cheng Chang, Chun Wei Huang, and Hone Jay Chu. Applying genetic algorithm and neural network to the conjunctive use of surface and subsurface water. *Water resources management*, 27 (14):4731–4757, 2013.

[91] Katie L Benko and Jörg E Drewes. Produced water in the western united states: geographical distribution, occurrence, and composition. *Environmental Engineering Science*, 25(2):239–246, 2008.

Chapter 3

Determination of Rare Earth Elements in Hypersaline Solutions Using Low-Volume, Liquid-Liquid Extraction

This chapter is adapted from a publication by the same name, co-authored by David A. Dzombak and Athanasios K. Karamalidis. This paper is citable as:

My contributions to this work were the experimental design; execution of experiments including ICP-MS analysis; data analysis and visualization; interpretation of results; and drafting of the manuscript.

Abstract

Complex, hypersaline brines — including those co-produced with oil and gas, rejected from desalination technologies, or used as working fluids for geothermal electricity generation — could contain critical materials such as the rare earth elements (REE) in valuable concentrations. Accurate quantitation of these analytes in complex, aqueous matrices is necessary for evaluation and implementation of systems aimed at recovering those critical materials. However, most analytical methods for measuring trace metals have not been validated for highly saline and/or chemically complex brines. Here we modified and optimized previously published liquid-liquid extraction (LLE) techniques [Shabani et al., *Anal. Chem.* 1990, 62 (24) 2709-2714; Lawrence and Kamber, *Geostand. Geoanal. Res.* 2007, 31 (2) 95-103], using bis(2-ethylhexyl) phosphate as the extractant in a heptane diluent, and studied its efficacy for REE recovery as a function of three primary variables: background salinity (as NaCl), concentration of a competing species (here Fe), and concentration of dissolved organic carbon (DOC). Results showed that the modified LLE was robust to a range of salinity, Fe, and DOC concentrations studied as well as constant, elevated Ba concentrations. With proper characterization of the natural samples of interest, this method could be deployed for accurate analysis of REE in small volumes of hyper-saline and chemically complex brines.

3.1 Introduction

The rare earth elements (REE) are among the most frequently cited critical materials for clean energy and high-tech manufacturing [1, 2]. The unique and varied properties of REE have led to their application in more consumer products than nearly any other element group [3]. REE are mostly obtained from mining and processing of REE-enriched ores [2]. While economically preferred, mining is laborious with a significant environmental burden, and inexpensive alternative sources of critical materials are sought after resources.

Aqueous byproduct or waste streams, both natural and industrial, are potential sources of the REE and other critical materials. With increasing global interest in geothermal energy [4], development of unconventional oil and gas resources (e.g. hydraulic fracturing of organic rich shales) [5], and

desalination technologies [6, 7], large volumes of waste brines are being managed and processed at great expense. Development of technologies for recovery of valuable byproducts, such as the REE, from these waste streams could improve the economics of these technologies while diversifying available critical material resources. Development of such technologies requires accurate determination of the source REE concentration in order to develop and implement recovery systems. However, precise quantitation of REE in complex matrices like brines is a significant challenge for conventional instrumentation such as inductively coupled plasma mass spectrometry (ICP-MS) [8].

There exists a dearth of methodologies in the analytical literature for quantitation of REE in brines by ICP-MS. Many approaches have been applied for separation and concentration of REE from aqueous media including solid-phase extraction (SPE) [9–25], co-precipitation (co-ppt) [26–28], and liquid-liquid extraction (LLE) [29–32]. However, nearly all studies in the analytical chemistry literature have focused on fresh water or seawater matrices, neglecting hypersaline waters (i.e. more concentrated than ~ 0.7 M NaCl or seawater). Despite this deficiency, approximately 14% of published measurements of REE in groundwater constitute brine samples (greater than 1 eq/kg ionic strength) [33], with these analyses utilizing methodologies not explicitly validated for extreme salinity.

Commonly applied separation techniques such as SPE and co-ppt may lack the robustness necessary to analyze REE in hypersaline brines. For example, high dissolved organic carbon may lead to fouling of column-based SPE while high dissolved metal loads may lead to saturation of the surface sites responsible for REE binding. Oliveira et al. [34] ascribed diminished Zn recovery in 166‰ salinity produced water to competitive sorption of matrix cations on their iminodiacetate resin. Similarly, excessive cations in hypersaline solutions may screen the REE from sorption sites during co-ppt, a phenomenon noted by Nelson et al. [35] for Ra determination in produced waters from the Marcellus Shale by both $BaSO_4$ and MnO_2 co-ppt. Moreover, at the elevated pH necessary for SPE and co-ppt techniques, the formation of energetically favorable, neutral- or negatively-charged aqueous complexes of the REE (with both organic and inorganic ligands) can further limit REE-particle partitioning [36].

Liquid-liquid extractions are potentially robust to all of these conditions and represent an attractive alternative for REE separation from hypersaline solutions [37]. Since LLE of REE from highly acidic solutions has been thor-

oughly studied for separation of lanthanides and actinides during nuclear fuel reprocessing [e.g. refs. 38, 39] elevated pH is not required of LLE techniques. Moreover, electrolyte theory dictates that increased sample salinity should enhance chemical partitioning (through salting out of neutral/micellar REE-organic ligand complexes from the aqueous feed to the organic solvent) and physical phase separation (by collapsing the electric double layer of the organic droplets, hastening coalescence). A primary obstacle in extraction of hydrophilic metals to a hydrophobic, organic phase is the dehydration of the metal cations in the aqueous phase. However, increasing salt concentrations decrease the effective concentration of water in the solution available for hydration of the target metal cations (REE), improving the energetics of the extraction [37].

In this work, efficiency of REE separation and concentration from synthetic brines using a LLE method for quantitative analysis was studied. The LLE method was adapted, modified, and optimized from previously published studies [29, 30]. For the LLE, a common ligand used for REE complexation and extraction, bis(2-ethylhexyl) phosphate (HDEHP), was studied in a heptane diluent. The objectives of this work were to: (1) evaluate the feasibility of REE recovery from small volumes of hypersaline solutions by LLE, (2) optimize the LLE methodology for high salinity brines, and (3) study the influence of brine composition on REE recovery.

3.2 Modified liquid-liquid extraction procedure

The modified LLE method adheres to many of the physical steps of the methods originally developed by Shabani et al. [29] and Lawrence and Kamber [30], but following optimization, differs in the operating conditions. A schematic flowsheet of the process is shown in Figure 3.1. The process involved sample preparation, followed by three cycles of extraction, whereby the REE were complexed by the HDEHP ligand and brought into the organic phase, leaving an REE-free, waste brine. Matrix elements were rinsed from the organic phase with dilute acid, and, finally, the REE were recovered by four cycles of elution with strong acid. This REE-loaded aqueous phase was then analyzed by ICP-MS. The details of the process are described in following. Details of sample preparation, instrumentation, analytical protocols, and data analysis methods are presented in Section 3.3.

Synthetic brine solutions were adjusted to pH 2.5 in 50 mL PP centrifuge
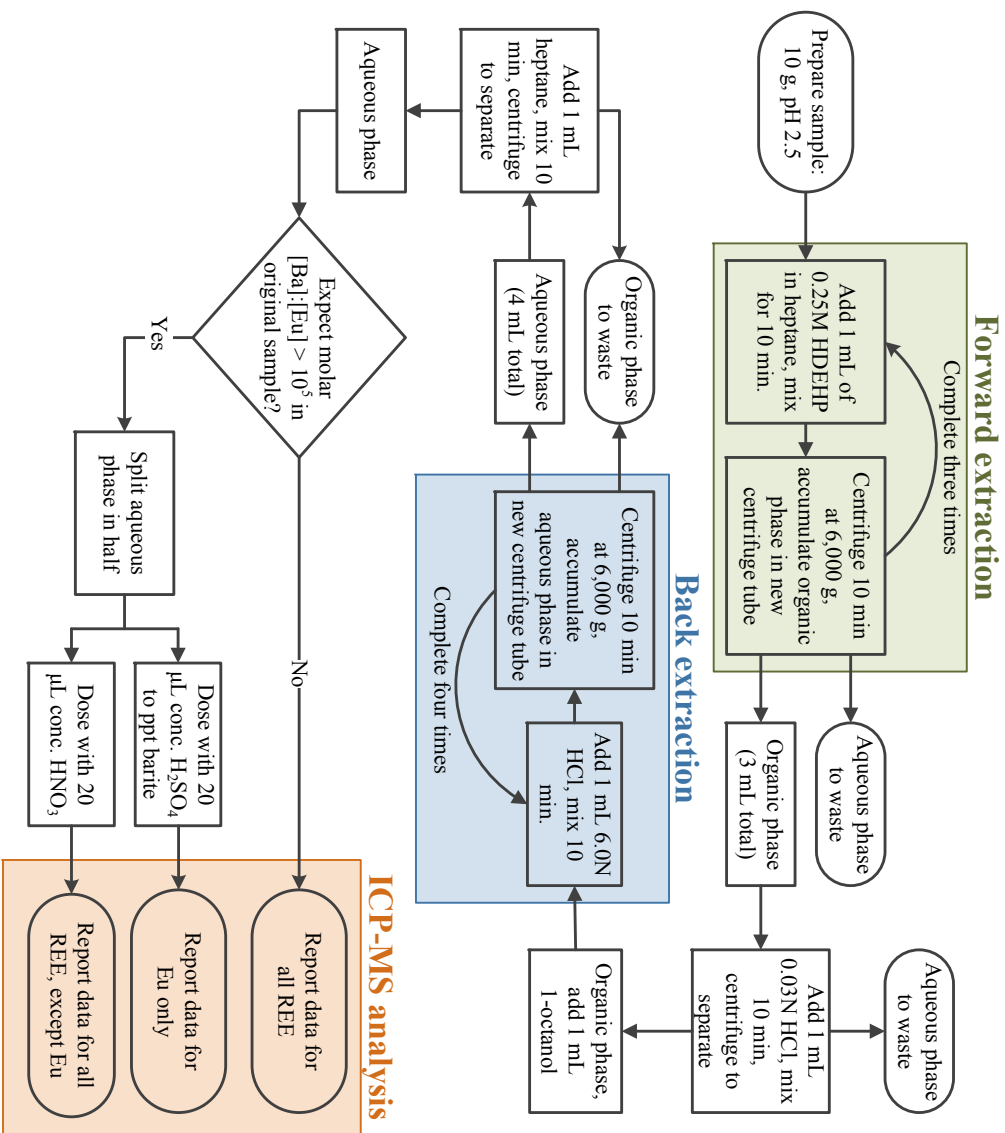
Figure 3.1: Liquid-liquid extraction method flowsheet for separation and concentration of REE from small volume, hypersaline brines. The decision node for the [Ba]:[Eu] molar ratio assumes a BaO^+ formation rate on the order of 0.1% in the ICP-MS.

tubes with HNO_3 and subsequently split into 10 g aliquots in 15 mL PP centrifuge tubes for replicate experiments. For each aliquot or sample, the process for REE extraction and recovery is as follows. One (1) mL of 0.25 M HDEHP in heptane was added to the aqueous solution. HDEHP was used as complexing agent for REE. The phases were emulsified and mixed end over end for 10 minutes. The phases were separated by centrifugation at 6,000×g for 10 minutes and the light organic phase was removed from the centrifuge tube via pipette and accumulated in a new centrifuge tube, retaining the aqueous phase in the original tube. This process, whereby the REE are complexed with HDEHP and partitioned into the heptane (termed forward extraction), was completed a total of three times. Following the third forward extraction, the aqueous phase was discarded.

To remove any matrix (Na, Fe, etc.) and interfering species (i.e. Ba) that partitioned during forward extraction, the accumulated organic phase (3 mL total) was rinsed with 1 mL of pH 1.5 HCl. This mixture was emulsified and separated by the same methods as the forward extraction. Once separated, the dense aqueous phase was removed via pipette and discarded.

A concentrated acid solution was used to dissociate the REE-HDEHP complexes and return the REE to an aqueous phase (termed back extraction). To decrease the REE-HDEHP complexation strength and encourage complete recovery [29], 1 mL of 1-octanol was added to the organic phase. Back extraction was achieved with four, sequential steps of stripping with 1 mL of 6.0 N HCl (collecting the eluted REE in a total of 4 mL acid). As with the forward extraction, the sample was emulsified and mixed end over end for 10 minutes and then separated via centrifugation at 6,000×g for 10 minutes. After centrifugation the aqueous phase was removed via pipette and accumulated in a separate centrifuge tube, retaining the organic phase in the original tube. Following the four back-extractions, the organic phase was discarded.

The collected acid volume (4 mL) was then rinsed with 1 mL of heptane to remove any dissolved organics from the aqueous phase. Phase mixing and separation were accomplished in the same manner as all other steps. Following centrifugation, the dense aqueous phase was removed and analyzed.

Preliminary experiments (Figure B.1) indicated that, while Ba was successfully removed in the course of the LLE ($> 99.9\%$ average reduction) and the HEHe-mode collision cell in the ICP-MS was successfully limiting $^{135}Ba^{16}O^+$ interferences with $^{151}Eu^+$ ($^{135}Ba^{16}O^+$:$^{135}Ba^+ \sim 0.2\%$ on average), initial Ba

concentrations were so high that ppb level, background Eu concentrations were observed (Figure B.2). Thus, in order to determine Eu accurately in these synthetic brines an additional step was tested, where an aliquot of the final, collected acid volume was dosed with 20 µL concentrated sulfuric acid (H_2SO_4) to precipitate any remaining barium as barite ($BaSO_4$). Efficiency of Ba removal after H_2SO_4 addition is compared in Figure B.1B and B.1D. It should be noted that this step was unnecessary for samples without Ba.

The methodology of Jenner et al. [40] as modified by McGinnis et al. [41] was employed to correct for matrix effects, isobaric interferences, and instrument drift during ICP-MS analysis. For each sample a 2 mL aliquot was spiked with 2 mL of a mixed element standard (5% HNO_3 background) while a separate 2 mL aliquot was diluted with 2 mL of blank 5% HNO_3. These solutions were analyzed sequentially to examine sample-specific matrix effects and were followed by a 5% HNO_3 flush. At the beginning of each analysis run and after every third sample, two separate standard solutions and an analytical blank were analyzed to monitor instrument drift and isobaric, polyatomic interferences (e.g. $^{135}Ba^{16}O^+$ on $^{151}Eu^+$ and light REE-oxides on heavy REE). Eight, serially-diluted, multi-element standard solutions ranging in concentration from 50 ppt to 100 ppb were analyzed at the beginning and end of each run to confirm the linearity assumed by the internal/external calibration. Typical analytical uncertainty was between 3 and 5%. Because of high backgrounds of Gd in our laboratory and high Ba in the experiments, oxide corrections for $^{137}Ba^{16}O^+$ interference on $^{151}Eu^+$ and $^{157}Gd^{16}O^+$ on $^{173}Yb^+$ were applied as in Aries et al. [42] after ICP-MS analysis.

3.3   Materials and methods

3.3.1   Chemicals and equipment

For the LLE, n-heptane (Chromasolv®; Lot # SHBC0837V), 1-octanol (Chromasolv®; Lot # SHBC6245V), and HDEHP (99.7% purity; Lot # MKBK0176V) were acquired from Sigma Aldrich. Nitric acid (HNO_3; BDH ARISTAR® Plus, VWR; assay 69 wt.%; Lot # 1113050) was used for sample pH adjustment and as the solvent for all analyses. Hydrochloric acid (HCl; BDH ARISTAR® Plus, VWR; 35 wt.%; Lot # 4113083) was used for matrix rinsing and REE back-extraction in the LLE. Chloride salts of

Na (Sigma Aldrich; $\geq 99\%$ purity), Ba (Alfa Aesar; $\geq 99.998\%$ purity), and Fe (Sigma Aldrich; $\geq 99.9\%$ purity, trace metal basis) and valeric acid (Alfa Aesar; 99% purity) were used for preparation of synthetic brines. Single element standard solutions (1000 µg/L) of the REE and all elements necessary for internal and external standardization were obtained from Inorganic Ventures. Polypropylene (PP) centrifuge tubes were used in the LLE and glass volumetric flasks were used to prepare organic phases. Ultrapure water (ASTM Type I, 18.2 MΩ/cm) was used for sample preparation and was prepared using a Barnstead NANOpure® water purification system. An OrionTM 8165BNWP ROSSTM Sure-FlowTM pH electrode (Thermo Scientific), coupled to an accumetTM XL600 meter (Fisher Scientific), was used for pH measurements of high-total dissolved solid (TDS) solutions. The pH meter was calibrated with pH 2.0, 4.0, and 7.0 standards daily. All samples were prepared gravimetrically using an analytical balance with 0.01 mg precision (Adam Equipment).

All analyses were performed on an Agilent 7700x ICP-MS with HEHe-mode octopole reaction cell. Operating parameters were optimized daily via the auto-tune function of the Agilent MassHunter software using 1000:1 diluted Agilent tuning solutions; typical operating parameters and monitored analyte masses are given in Table 3.1.

3.3.2 Recovery analysis by surrogate recovery

In the absence of a high-salinity, certified reference material for REE, the IUPAC recommended methodology of surrogate recovery [43] was used to study the recovery of REE by this LLE method. Synthetic brine samples were either left blank or spiked with constant amounts of REE. The difference in results between these samples yielded the mass of the spiked analytes recovered by the LLE method. Additionally, indium was included in the spike solution as it is commonly employed as an internal standard for REE separation to monitor recovery [29, 44]. Data below the instrument detection limit (DL) were assigned a value of $0.5 \times$DL for computation; this occurred for roughly 41% of all the REE in the unspiked samples and 4% in the spiked samples.

Table 3.1: Typical operating conditions for ICP-MS analysis. Analysis performed on Agilent 7700x using oxygen-free argon as the carrier and dilution gas and ultra high-purity helium in the reaction cell. Conditions determined using 1000:1 diluted Agilent tuning solution. For elements where multiple mass-to-charge ratios were monitored, ^{148}Sm, ^{151}Eu, and ^{157}Gd were used in data analysis.

	Parameter	Value
Plasma		
	RF power	1600 W
	Nebulizer pump rate	0.10 rps
	Carrier argon flow rate	0.61 L/min
	Dilution argon flow rate	0.36 L/min
Lenses		
	Extract 1	0.0 V
	Extract 2	-200.0 V
	Omega bias	-110 V
	Omega lens	9.6 V
	Cell entrance	-110 V
	Cell exit	-150 V
	Deflect	-74.8 V
	Plate bias	-150 V
Octopole reaction cell		
	Octopole bias	-100.0 V
	Octopole RF	200 V
	He flow rate	10 mL/min
	Energy discrimination	7.0 V
Data aquisition		
	Replicates	5
	Integration time	0.3 s
Masses monitored		
	^{45}Sc, ^{89}Y, ^{115}In, ^{135}Ba, ^{137}Ba, ^{139}La, ^{140}Ce, ^{141}Pr, ^{145}Nd, ^{147}Sm, ^{148}Sm, ^{151}Eu, ^{153}Eu, ^{157}Gd, ^{158}Gd, ^{159}Tb, ^{163}Dy, ^{165}Ho, ^{167}Er, ^{169}Tm, ^{173}Yb, ^{175}Lu	
Oxides and doubly charged		
	^{140}Ce^{16}O$^+$/^{140}Ce $< 2.1\%$ ^{140}Ce^{2+}/^{140}Ce $< 1.6\%$	

3.3.3 Preparation of synthetic brines

In addition to optimization of operating parameters, the methodology must be validated for complex brine solutions. The complexity of the brine is a function of background salinity and interfering compounds (both inorganic and organic). This was investigated in two stages. First, simple solutions (1 m and 5 m NaCl) were used to evaluate feasibility. Second, compositional complexity was explored via a uniform shell, or Doehlert, experimental design [45], varying NaCl, Fe, and dissolved organic carbon (DOC) concentrations.

The concentrations of background salinity and interfering compounds were chosen to match reported value ranges found in studies of produced waters from unconventional gas development in the Marcellus Shale [46–50], however the range of compositions studied is similar to other deep, basinal brines [51]. Dissolved organic carbon was modeled with pentanoic (or valeric) acid, a common component of deep, saline brines [51], with representative metal-complexing functionality. Additionally, organic acids have been shown to be a significant component of DOC in produced waters from the Marcellus Shale [52]. The parameters of interest — concentrations of NaCl, Fe, and DOC — were scaled linearly. Experimental conditions for variability in salinity, Fe concentration, and DOC concentration are given in Table 3.2. For all experiments the concentration of each REE (along with indium) was set at 500 ppt (parts per trillion), a value that falls between the 45th percentile (for Tm) and the 1st percentile (for La) of natural REE distributions in groundwater [33]. Total dissolved Ba was held constant at 2,000 ppm, roughly the average concentration observed by Barbot, et al. [47] for Marcellus Shale produced waters. Results of these experiments were analyzed by multiple linear regression (MLR) to determine which parameters of the synthetic brine influenced the recovery most strongly.

3.3.4 Optimization of liquid-liquid operation parameters

Preliminary experiments utilizing LLE conditions described by Shabani et al. [29] and Lawrence and Kamber [30] were unable to achieve high or consistent recovery of the REE (Figure 3.2A,B). These experiments showed recovery $< 80\%$ for all elements in all replicates, with significant variability, as well as preferential recovery of the MREE over the L- and HREE. It should be noted that both Shabani et al. [29] and Lawrence and Kamber [30]

Table 3.2: Doehlert experimental design matrix for LLE validation in chemically complex brines. Doehlert coding for each experiment are given in parentheses next to the parameter value. All variables were varied arithmetically.

Exp.	[NaCl] (m)	[Fe] (ppm)	[DOC] (ppm-C)
1	2 (0)	40 (0)	200 (0)
2	3.5 (1)	40 (0)	200 (0)
3	2.75 (0.5)	74.6 (0.866)	200 (0)
4	2.75 (0.5)	51.6 (0.289)	363 (0.817)
5	0.5 (-1)	40 (0)	200 (0)
6	1.25 (-0.5)	5.4 (-0.866)	200 (0)
7	1.25 (-0.5)	28.4 (-0.289)	37 (-0.817)
8	2.75 (0.5)	5.4 (-0.866)	200 (0)
9	2.75 (0.5)	28.4 (-0.289)	37 (-0.817)
10	1.25 (-0.5)	74.6 (0.866)	200 (0)
11	2 (0)	63.1 (0.577)	37 (-0.817)
12	1.25 (-0.5)	51.6 (0.289)	363 (0.817)
13	2 (0)	16.9 (-0.577)	363 (0.817)

employed a mixture of mono- and di-ester phosphonic acids as the chelating ligand. This ligand combination was also explored, however preliminary experiments provided poor results compared to the pure diester (HDEHP) under the same conditions and was not studied further.

In order to optimize method performance, a linear model (Equation 3.1) was fit by ordinary least-squares in R [53], using the datasets of Kimura [31, 32]. The relationship between the response (organic-aqueous distribution coefficient, K_d) and each of the predictors (solution acidity $[ACY]$, and ligand concentration $[L]$) is shown to be independently log-log linear. Therefore the variables in this model correspond to log values. Data for fitting of this model were extracted from Figure 1 of Kimura [31] for K_d vs. $[ACY]$ at constant $[L]$ and from Figure 1 of Kimura [32] for K_d vs. $[L]$ at constant $[ACY]$. This estimation of parameter values reflects the dependence of $\log K_d$ on $\log[ACY]$ (β_{ACY}) and on $\log[L]$ (β_L) as well as a constant intercept (β_0).

$$\log K_d = \beta_{ACY} \cdot \log[ACY] + \beta_L \cdot \log[L] + \beta_0 \tag{3.1}$$

The extraction of REE from the aqueous to the organic phase is calculated
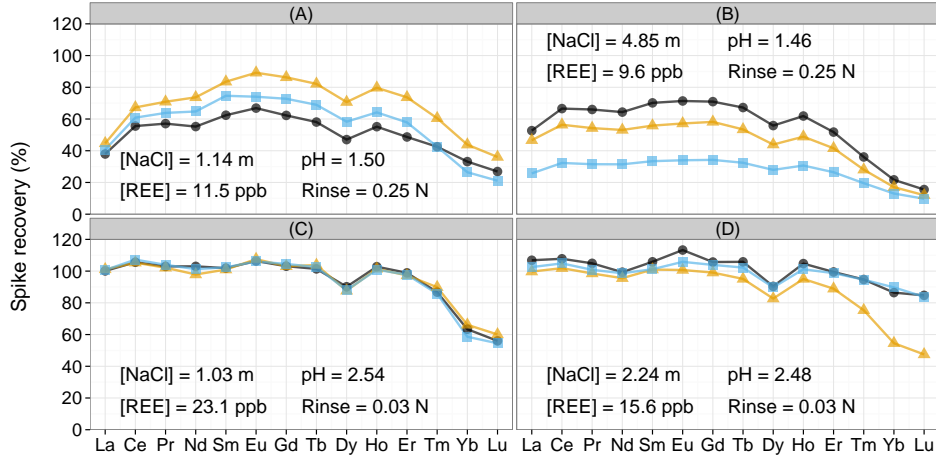
Figure 3.2: REE recovery from 10 g samples of synthetic brine solutions using LLE conditions recommended by Shabani, et al.[7] (A, B) and "optimal" conditions predicted by multiple linear regression (C, D). Each experiment was performed in triplicate (separate colors/shapes of plot) and the experimental conditions are shown in the subfigures.

using the predicted K_d values based on mass balance. The fraction of REE mass in the organic phase (R_{org}) for equilibrium between an aqueous phase (with volume V_{aq}) and an organic phase (with volume V_{org}) is calculated by Equation 3.2.

$$R_{org} = \frac{1}{1 + \frac{V_{aq}}{V_{org}} \cdot K_d^{-1}} \tag{3.2}$$

The system can be represented as independent LLE in series since the phases are separated after each extraction step. Thus, the overall partitioning of REE from the brine to the organic phase (R_{tot}) in the forward extraction can be calculated for n sequential extractions with Equation 3.3. This allows for determination of the number of extractions necessary for quantitative recovery of REE. The analysis is simply reversed to examine the elution properties of REE from the organic phase back into an aqueous phase.

$$R_{tot} = \sum_{i=1}^{n} R_{org} \left(1 - R_{org}\right)^{i-1} \tag{3.3}$$

This analysis is meant to provide a "best guess" as to the optimal method parameters without requiring additional experimentation. The inherent limitation of this approach is the uncertain extensibility of the original data to both a modified methodology (i.e. small volumes, changed organic diluent, mixed analyte solutions, low initial REE concentration) and unique matrices (i.e. acidified brines vs. HCl). Therefore post hoc analysis of preliminary experiments for parameter optimization was done qualitatively and is described in Section 3.4.1. Moreover, since K_d values were not calculated as part of this study, model validation with new experimental results was not performed.

3.4 Results and discussion

3.4.1 Multiple linear-regression of organic-aqueous distribution coefficients

Fractionation of the REE observed in preliminary experiments (Figures 3.2A and 3.2B) can be qualitatively reconciled from MLR analysis. Figure 3.3 shows that the rinse step (labeled "R" in subfigures 3.3A to 3.3D) creates strong stripping conditions ($\log K_d < -1$) for the LREE (La; Figure 3.3A) while the back extraction (labeled "B") may provide inadequate acidity to recover the HREE (Tm; Figure 3.3B) once partitioned.

Using the equilibrium mass balance model (Equations 3.2 and 3.3), the K_d required to achieve the desired partitioning was matched to the results of this regression. For each step of the LLE the desired partitioning is as follows:

1. Forward extraction: Achieve high ($> 99\%$) partitioning of the REE into the organic phase, while minimizing the partitioning of Ba and Na. This can be achieved in a series of forward extractions.

2. Matrix rinse: Minimize partitioning of Ba and Na (i.e. low K_d) while maintaining a high ($> 99\%$) partitioning of the REE.
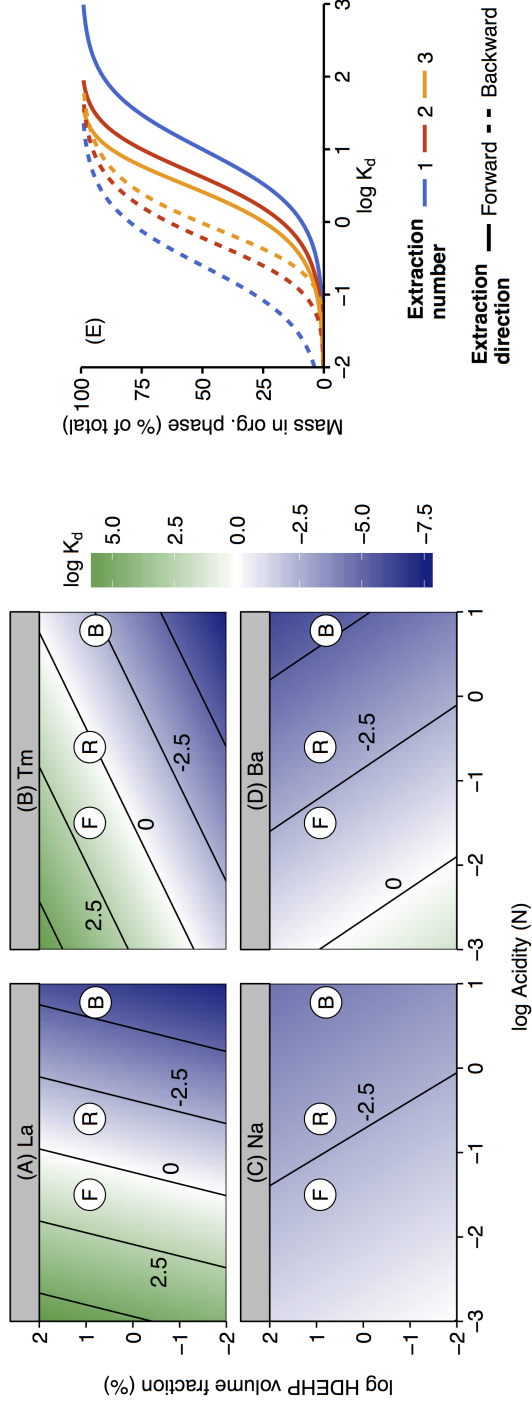
Figure 3.3: Summary of model-based optimization of LLE operating conditions. (A–D) Organic-aqueous distribution coefficient (log K_d; Equation 3.1) contours as a function of acidity and HDEHP volume fraction for La, Tm, Na, and Ba calculated with data from Kimura [31, 32]. LLE operating conditions for forward extractions (F), matrix rinse (R), and back extractions (B) suggested by Shabani et al. [29] are noted. (E) Equilibrium partitioning for triplicate forward and backward extractions as a function of organic/aqueous distribution coefficient (K_d). Partitioning calculated by Equation 3.2 with $V_{aq}/V_{org} = 10$ for forward extraction and 0.25 for backward extraction.

3. Backward extraction: Minimize REE partitioning (i.e. leave $< 1\%$ in the organic phase). This can be achieved in a series of backward extractions.

Linear regression predicts a $\log_{10} K_d$ of 1.15 for La and 1.62 for Tm when using the conditions recommended by Shabani et al. [29]. At these values, a mass balance model would predict 93% partitioning in three forward extractions for La and exactly 99% for Tm. Similarly, in the recommended matrix rinse (0.25 N HCl) the model predicts $\log_{10} K_d$ of -1.48 for La and 0.21 for Tm, which correspond to 9% retention of La in the organic phase and 83% retention of Tm. While this simplified analysis suggests that triplicate back extraction steps should achieve quantitative recovery of all REE ($\log_{10} K_d$ of -2.18 and -5.57 for La and Tm respectively), preliminary experiments (Figure 3.2C and 3.2D) showed that the HREE were incompletely recovered. From these results an additional elution step was added to the procedure.

From this analysis, it is clear that the K_d values need to increase for forward extraction and the matrix rinse in order to achieve the desired partitioning at each step. The LLE conditions, namely the solution acidity $[ACY]$, necessary to meet these goals can be found be examining the contours of Figure 3.3A,B and E . Using Figure 3.3E it can be shown that forward extractions require $\log K_d > 1.6$ to achieve $> 99\%$ partitioning to the organic phase after triplicate extractions. Further, the conditions of the rinse phase must maintain $\log K_d > 1.4$ to retain $> 99\%$ REE in the organic phase from one rinse step. Finally, to achieve $> 99\%$ recovery of REE during triplicate back extractions, conditions must create $\log K_d < -1.2$. By decreasing the initial acidity to $< 10^{-2}$ N (i.e. pH > 2) the initial extraction of REE can be enhanced without significantly increasing partitioning of Na or Ba. From this analysis, an initial sample pH of 2.5 was chosen. At a pH of 2.5 ($[ACY] = 10^{-2.5}$ N), forward extraction yields $\log K_d$ of 4.1 and 3.2 for La and Tm respectively. Dissolution of the organic phase, which has been shown to diminish partitioning, should be limited below pH 3.54 Similarly, by reducing the acidity of the rinse step to $\leq 10^{-1.5}$ N (i.e. pH ≥ 1.5) salts such as Na and Ba (Figures 3.3C and 3.3D) can be eluted while retaining the REE (Figure 3.3A and 3.3B).

3.4.2 Interferences of synthetic brine constituents

Results for REE recovery from simple solutions of 1 and 5 m NaCl, the REE are presented in Figure 3.4. In the 1 m NaCl solution, REE recovery was consistently between 90 and 110%, however, indium recovery was markedly lower: 76 and 81% in duplicate experiments. Similar results were observed in the high salinity test (5 m NaCl), except that the heaviest two lanthanides, Yb and Lu, were recovered at a much lower rate, averaging 82% for Yb and 72% for Lu. As the matrix rinse and back extraction conditions were identical between the 1 m and 5 m NaCl experiments, the diminished recovery is likely an artifact of diminished forward extraction and thus an effect of the increased salinity. Recovery of indium from the 5 m NaCl solution (mean, 76%) matched the recovery observed in the 1 m NaCl solution indicating that this diminished recovery is likely not a function of the salt concentration, and is instead endemic to indium in this extraction.
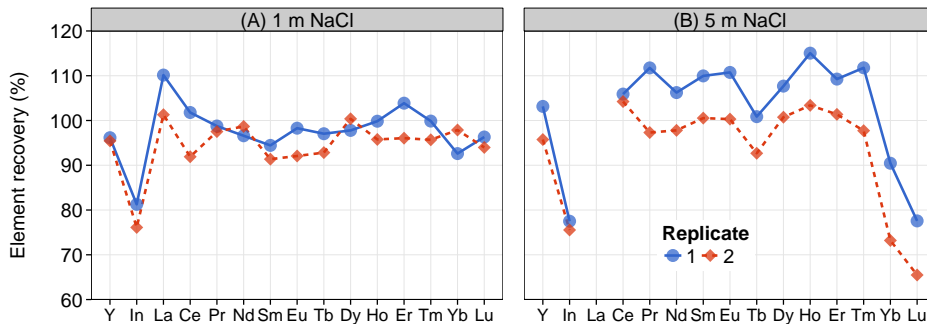
Figure 3.4: REE recovery by LLE method in simple, saline solutions. Initial REE concentrations were 500 ppt (each element) in all experiments. Duplicate experiments were conducted at each salinity and are represented by different line types, marker shapes, and colors. Data for La in the 5 m NaCl experiment and for Gd in both experiments are excluded due to excessive background concentrations ($\geq$ 250 ppt).

Figure 3.5 summarizes the Doehlert design matrix experimental results, while experiment-ordered results, with replicates, are presented in Figure B.3. Data for Sc, La, and Gd are not included because these analytes were either not recovered (Sc, which is "irreversibly bound" in the organic phase [31]) or contaminated by a high background (La and Gd; background $\geq$ 250 ppt in all experiments). Subsequent discussion excludes these elements. Across all analytes, median recovery, $\widehat{Q}(0.50) = 106\%$, was biased high (two-sided

Wilcoxon Signed Rank test; H_1: $Q(0.5) \neq 100\%$, $P < 10^{-6}$). However, these results are well within the recommended range of mean recovery for 1 ppb analytes (40 – 120% recommended) [54]. The absolute range of recoveries observed was 83-158% while the reference range for recoveries (i.e the values between which 95% of observations fell) was 92% – 122%. Experiments were generally reproducible (Figure B.3), with element-specific, replicate standard deviations ranging from 0.05% (Nd, experiment 4) to 21% (In, experiment 1). Finally, indium recovery was indistinguishable from any analyte (Wilcoxon, Rank-Sum test; $P > 0.05$), in contrast to recovery from the simple NaCl solutions (Figure 3.4). Since these data are not sufficient to determine the mechanism by which this disparity was overcome, the application of indium as a tracer of REE recovery by the LLE method in unknown samples requires further study.
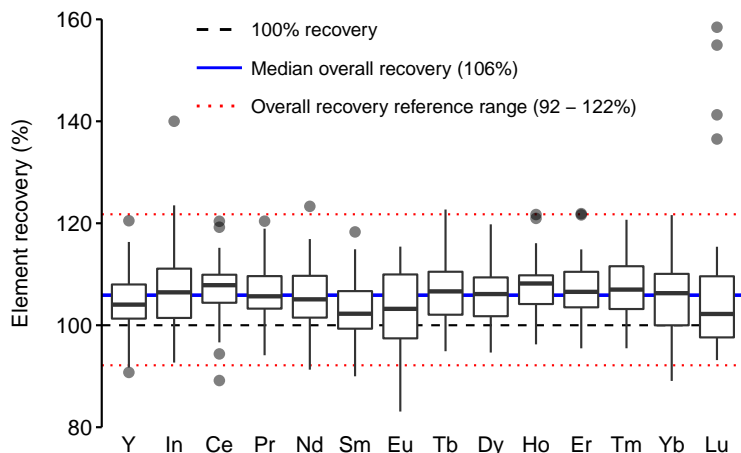
Figure 3.5: Distributions of elemental recovery in Doehlert matrix experiments by LLE methodology (see Figure B.3 for experiment-wise results). Distributions are depicted as standard box plots [55], where the thick, black line depicts the median; the boxed range represents the 25th to 75th percentile, or inter-quartile range (IQR); the thin whiskers denote all measurements within 1.5 times the IQR above or below; and the remaining observations are shown as semi-transparent grey dots. Relevant summary statistics for the overall suite of analytes are given as horizontal lines. Recovery values for Eu were determined after dosing the LLE eluent with H_2SO_4 to precipitate barite; all other recoveries were determined without barite precipitation. Data from La and Gd are not presented because the background concentration was determined to 250 ppt or greater (see Figure B.2).

The data in Figure 3.5 and Figure B.3 indicate no clear fractionation (or mass bias) of the method across the suite of REE. This result is confirmed by the Kruskal-Wallis test, which found no significant differences between any two element recoveries (H_0: No differences in element medians, $P > 0.1$). Pairwise element testing (paired sample Wilcoxon Rank-Sum test, corrected for multiple comparisons) found statistically significant differences ($P < 0.05$) between 16 element pairs. However, the differences were essentially indistinguishable from additive 3% analytical errors (Figure 3.6).
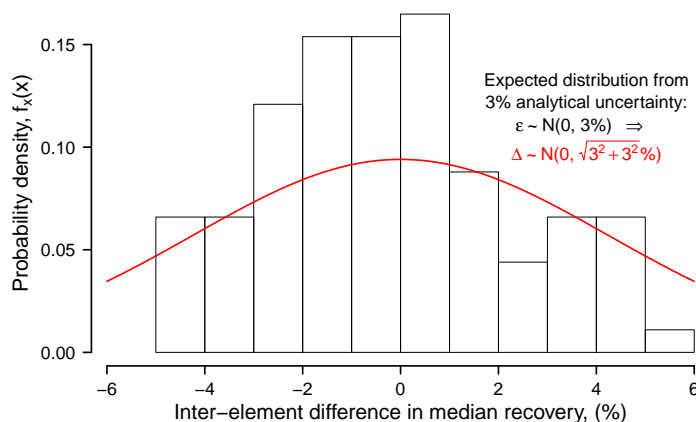
Figure 3.6: Comparison of pair-wise, inter-element differences in median recovery (Δ, estimated from the paired-sample Wilcoxon Signed Rank test) to the expected (normal) distribution based on the difference between two random variates with equal means and 3% analytical unceritnty, i.e. $\Delta = N(\mu, 3\%) - N(\mu, 3\%) = N(0, \sqrt{3^2 + 3^2}\%)$.

The lack of fractionation among REE in the more complex brines (i.e. with salinity, Fe, and DOC) differs from results obtained in experiments with simple NaCl solutions, where Yb and Lu recoveries were significantly lower at 5 m NaCl. Step-wise regression analysis of element recovery (response) against solution composition (predictors) revealed no combination of linear- or interaction-terms among the study variables that substantially influenced recovery. If the solution composition did have any impact on recovery within the range of parameters explored in the Doehlert design, it was indistinguishable from replicate variability. These results give confidence to the application of the LLE methodology for natural samples with chemical characteristics within the bounds of the variables studied here, though accurate characterization of the REE concentration of unknown samples may require multiple replicates.

3.4.3   Advice regarding natural samples

Based on the results presented here, the modified LLE technique represents an attractive option for determination of REE in natural, hypersaline, and chemically complex brines. However, it is critical to have accurate characterization of the samples of interest, as well as the oxide formation rates for the ICP-MS. For samples with low Ba (i.e. a molar [Ba]:[Eu] $< 10^5$ in raw samples), the addition of H_2SO_4 for barite precipitation is likely unnecessary if the available analytical instrumentation can maintain BaO^+ interferences on the order of 0.1% of analyte signal. Samples with salinities and/or compositions outside of the range validated here may need to be tested with synthetic brines by the user.

While this study has explored the effects of DOC on LLE performance by way of a model compound, the character of DOC in the samples of interest should be considered more explicitly. The effects of mixed organic components were not studied here. Potent REE chelators can be found in other water sources, such as municipal wastewater discharges [56, 57], and may have uncertain effects on the efficiency of REE recovery by this and other methods.

References

[1] APS. Energy critical elements: Securing materials for emerging technologies. Report, American Physical Society, 2011.

[2] D. Bauer, D. Diamond, J. Li, D. Sandalow, P. Telleen, and B. Wanner. US Department of Energy Critical Materials Strategy. Report, US Department of Energy, 2010.

[3] Stephen B. Castor and James B. Hedrick. *Rare Earth Elements*, pages 769–792. SME, 7 edition, 2006.

[4] John W. Lund, Derek H. Freeston, and Tonya L. Boyd. Direct utilization of geothermal energy 2010 worldwide review. *Geothermics*, 40(3): 159–180, 2011.

[5] U.S. Energy Information Administration. Annual energy outlook 2014 early release overview. Report, U.S. Energy Information Administration, 2013.

[6] Mark A Shannon, Paul W Bohn, Menachem Elimelech, John G Georgiadis, Benito J Mariñas, and Anne M Mayes. Science and technology for water purification in the coming decades. *Nature*, 452(7185):301–310, 2008.

[7] Menachem Elimelech and William A. Phillip. The future of seawater desalination: Energy, technology, and the environment. *Science*, 333 (6043):712–717, 2011.

[8] Christian Agatemor and Diane Beauchemin. Matrix effects in inductively coupled plasma mass spectrometry: A review. *Analytica Chimica Acta*, 706(1):66–83, 2011.

[9] K. Benkhedda, H. G. Infante, E. Ivanova, and F. C. Adams. Determination of sub-parts-per-trillion levels of rare earth elements in natural waters by inductively coupled plasma time-of-flight mass spectrometry after flow injection on-line sorption preconcentration in a knotted reactor. *Journal of Analytical Atomic Spectrometry*, 16(9):995–1001, 2001.

[10] Qiang Fu, Limin Yang, and Qiuquan Wang. On-line preconcentration with a novel alkyl phosphinic acid extraction resin coupled with inductively coupled plasma mass spectrometry for determination of trace rare earth elements in seawater. *Talanta*, 72(4):1248–1254, 2007.

[11] Brian A. Haley and Gary P. Klinkhammer. Complete separation of rare earth elements from small volume seawater samples by automated ion chromatography: method development and application to benthic flux. *Marine Chemistry*, 82(3–4):197–220, 2003.

[12] Ludwik Halicz, Ittai Gavrieli, and Ethel Dorfman. On-line method for inductively coupled plasma mass spectrometric determination of rare earth elements in highly saline brines. *Journal of Analytical Atomic Spectrometry*, 11(9):811–814, 1996.

[13] Shizuko Hirata, Tasuku Kajiya, Masato Aihara, Kazuto Honda, and Osamu Shikino. Determination of rare earth elements in seawater by on-line column preconcentration inductively coupled plasma mass spectrometry. *Talanta*, 58(6):1185–1194, 2002.

[14] Tasuku Kajiya, Masato Aihara, and Shizuko Hirata. Determination of rare earth elements in seawater by inductively coupled plasma mass spectrometry with on-line column pre-concentration using 8-quinolinole-immobilized fluorinated metal alkoxide glass. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 59(4):543–550, 2004.

[15] Rosi Ketrin Katarina, Mitsuko Oshima, and Shoji Motomizu. High-capacity chitosan-based chelating resin for on-line collection of transition and rare-earth metals prior to inductively coupled plasma-atomic emission spectrometry measurement. *Talanta*, 79(5):1252–1259, 2009.

[16] Intae Kim, Seolwon Kim, and Guebuem Kim. Analytical artifacts associated with the chelating resin extraction of dissolved rare earth elements in natural water samples. *Aquatic Geochemistry*, 16(4):611–620, 2010.

[17] M. Kühn and M. Kriews. Improved detection of transition and rare

earth elements in marine samples with the CETAC DSX-100 preconcentration/matrix elimination system and ICP-MS. *Fresenius' Journal of Analytical Chemistry*, 367(5):440–444, 2000.

[18] Sanjukta A. Kumar, Shailaja P. Pandey, Niyoti S. Shenoy, and Sangita D. Kumar. Matrix separation and preconcentration of rare earth elements from seawater by poly hydroxamic acid cartridge followed by determination using ICP-MS. *Desalination*, 281(0):49–54, 2011.

[19] Peter Möller, Peter Dulski, and Joachim Luck. Determination of rare earth elements in seawater by inductively coupled plasma-mass spectrometry. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 47(12): 1379–1387, 1992.

[20] Klaus J. Stetzenbach, Megumi Amano, David K. Kreamer, and Vernon F. Hodge. Testing the limits of ICP-MS: Determination of trace elements in ground water at the part-per-trillion level. *Ground Water*, 32(6):976–985, 1994.

[21] O. Vicente, A. Padró, L. Martinez, R. Olsina, and E. Marchevsky. Determination of some rare earth elements in seawater by inductively coupled plasma mass spectrometry using flow injection preconcentration. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 53(9):1281–1287, 1998.

[22] Bei Wen, Xiao-quan Shan, and Shu-guang Xu. Preconcentration of ultratrace rare earth elements in seawater with 8-hydroxyquinoline immobilized polyacrylonitrile hollow fiber membrane for determination by inductively coupled plasma mass spectrometry. *Analyst*, 124(4): 621–626, 1999.

[23] S. N. Willie and R. E. Sturgeon. Determination of transition and rare earth elements in seawater by flow injection inductively coupled plasma time-of-flight mass spectrometry. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 56(9):1707–1716, 2001.

[24] Tian-hong Zhang, Xiao-quan Shan, Rui-xia Liu, Hong-xiao Tang, and Shu-zhen Zhang. Preconcentration of rare earth elements in seawater with poly(acrylaminophosphonic dithiocarbamate) chelating fiber prior to determination by inductively coupled plasma mass spectrometry. *Analytical Chemistry*, 70(18):3964–3968, 1998.

[25] Yanbei Zhu, Tomonari Umemura, Hiroki Haraguchi, Kazumi Inagaki, and Koichi Chiba. Determination of REEs in seawater by ICP-MS

after on-line preconcentration using a syringe-driven chelating column. *Talanta*, 78(3):891–895, 2009.

[26] Timothy J. Shaw, Thomas Duncan, and Bernhard Schnetger. A preconcentration/matrix reduction method for the analysis of rare earth elements in seawater and groundwaters by isotope dilution ICPMS. *Analytical Chemistry*, 75(14):3396–3403, 2003.

[27] William M. Shannon and S. A. Wood. *The Analysis of Picogram Quantities of Rare Earth Elements in Natural Waters*, page 1—37. Springer, Dordrecht, The Netherlands, 2005.

[28] Maria Raso, Paolo Censi, and Filippo Saiano. Simultaneous determinations of zirconium, hafnium, yttrium and lanthanides in seawater according to a co-precipitation technique onto iron-hydroxide. *Talanta*, 116(0), 2013.

[29] Mohammad B. Shabani, Tasuku Akagi, Hiroshi Shimizu, and Akimasa Masuda. Determination of trace lanthanides and yttrium in seawater by inductively coupled plasma mass spectrometry after preconcentration with solvent extraction and back-extraction. *Analytical Chemistry*, 62 (24):2709–2714, 1990.

[30] Michael G. Lawrence and Balz S. Kamber. Rare earth element concentrations in the natural water reference materials (NRCC) NASS-5, CASS-4 and SLEW-3. *Geostandards and Geoanalytical Research*, 31 (2):95–103, 2007.

[31] Kan Kimura. Inorganic extraction studies on the system between bis (2-ethyl hexyl)-orthophosphoric acid and hydrochloric acid (i). *Bulletin of the Chemical Society of Japan*, 33(8):1038–1046, 1960.

[32] Kan Kimura. Inorganic extraction studies on the system between bis (2-ethyl hexyl)-orthophosphoric acid and hydrochloric acid (ii). *Bull. Chem. Soc. Japan*, 34, 1961.

[33] Clinton W. Noack, David A. Dzombak, and Athanasios K. Karamalidis. Rare earth element distributions and trends in natural waters with a focus on groundwater. *Environmental Science & Technology*, 48(8): 4317–4326, 2014.

[34] Eliane Padua Oliveira, Lu Yang, Ralph E. Sturgeon, Ricardo Erthal Santelli, Marcos Almeida Bezerra, Scott N. Willie, and Ramses Capilla. Determination of trace metals in high-salinity petroleum produced for-

mation water by inductively coupled plasma mass spectrometry following on-line analyte separation/preconcentration. *Journal of Analytical Atomic Spectrometry*, 26(3):578–585, 2011.

[35] Andrew W. Nelson, Dustin May, Andrew W. Knight, Eric S. Eitrheim, Marinea Mehrhoff, Robert Shannon, Robert Litman, and Michael K. Schultz. Matrix complications in the determination of radium levels in hydraulic fracturing flowback water from Marcellus Shale. *Environmental Science & Technology Letters*, 1(3):204–208, 2014.

[36] Yigal Erel and Edward M. Stolper. Modeling of rare-earth element partitioning between particles and solution in aquatic environments. *Geochimica et Cosmochimica Acta*, 57(3):513–518, 1993.

[37] Kenneth L. Nash. A review of the basic chemistry and recent developments in trivalent f-elements separations. *Solvent Extraction and Ion Exchange*, 11(4):729–768, 1993.

[38] Boyd Weaver and FA Kappelmann. TALSPEAK: A new method of separating americium and curium from the lanthanides by extraction from an aqueous solution of an aminopolyacetic acid complex with a monoacidic organophosphate or phosphonate. Report, Oak Ridge National Lab., Tenn., 1964.

[39] Mikael Nilsson and Kenneth L Nash. Review article: A review of the development and operational characteristics of the TALSPEAK process. *Solvent Extraction and Ion Exchange*, 25(6):665–701, 2007.

[40] GA Jenner, HP Longerich, SE Jackson, and BJ Fryer. ICP-MS - a powerful tool for high-precision trace-element analysis in earth sciences: evidence from analysis of selected USGS reference samples. *Chemical Geology*, 83(1):133–148, 1990.

[41] Cathleen E. McGinnis, Jinesh C. Jain, and Clive R. Neal. Characterisation of memory effects and development of an effective wash protocol for the measurement of petrogenetically critical trace elements in geological samples by ICP-MS. *Geostandards Newsletter*, 21(2):289–305, 1997.

[42] Sébastien Aries, Michel Valladon, Mireille Polvé, and Bernard Dupré. A routine method for oxide and hydroxide interference corrections in ICP-MS chemical analysis of environmental and geological samples. *Geostandards Newsletter*, 24(1):19–31, 2000.

[43] Michael Thompson, Steven LR Ellison, Aleš Fajgelj, Paul Willetts, and Roger Wood. Harmonized guidelines for the use of recovery information in analytical measurement. *Pure and applied chemistry*, 71(2):337–348, 1999.

[44] Beata Zawisza, Katarzyna Pytlakowska, Barbara Feist, Marzena Polowniak, Andrzej Kita, and Rafal Sitko. Determination of rare earth elements by spectroscopic techniques: a review. *Journal of Analytical Atomic Spectrometry*, 26(12):2373–2390, 2011.

[45] David H. Doehlert. Uniform shell designs. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 19(3):231–239, 1970.

[46] Lara O. Haluszczak, Arthur W. Rose, and Lee R. Kump. Geochemical evaluation of flowback brine from Marcellus gas wells in Pennsylvania, USA. *Applied Geochemistry*, 28:55–61, 2013.

[47] Elise Barbot, Natasa S. Vidic, Kelvin B. Gregory, and Radisav D. Vidic. Spatial and temporal correlation of water quality parameters of produced waters from devonian-age shale following hydraulic fracturing. *Environmental Science & Technology*, 47(6):2562–2569, 2013.

[48] Kelvin B. Gregory, Radisav D. Vidic, and David A. Dzombak. Water management challenges associated with the production of shale gas by hydraulic fracturing. *Elements*, 7(3):181–186, 2011.

[49] Arvind Murali Mohan, Angela Hartsock, Kyle J. Bibby, Richard W. Hammack, Radisav D. Vidic, and Kelvin B. Gregory. Microbial community changes in hydraulic fracturing fluids and produced water from shale gas extraction. *Environmental Science & Technology*, 47(22): 13141–13150, 2013.

[50] RD Vidic, SL Brantley, JM Vandenbossche, D Yoxtheimer, and JD Abad. Impact of shale gas development on regional water quality. *Science*, 340(6134), 2013. ISSN 0036-8075.

[51] Yousif K Kharaka, Paul D Lundegard, and Thomas H Giordano. Distribution and origin of organic ligands in subsurface waters from sedimentary basins. *Rev. Econ. Geol*, 9:119–131, 2000.

[52] Robert Wolford. *Characterization of organics in the Marcellus Shale flowback and produced waters*. Thesis, The Pennsylvania State University, 2011.

[53] R Core Team. R: A language and environment for statistical computing, 2014.

[54] Isabel Taverniers, Marc De Loose, and Erik Van Bockstaele. Trends in quality in the analytical laboratory. II. Analytical method validation and quality assurance. *TrAC Trends in Analytical Chemistry*, 23(8): 535–552, 2004.

[55] Robert McGill, John W Tukey, and Wayne A Larsen. Variations of box plots. *The American Statistician*, 32(1):12–16, 1978.

[56] Michael Bau and Peter Dulski. Anthropogenic origin of positive gadolinium anomalies in river waters. *Earth and Planetary Science Letters*, 143 (1):245–255, 1996.

[57] Serkan Kulaksiz and Michael Bau. Anthropogenic dissolved and colloid/nanoparticle-bound samarium, lanthanum and gadolinium in the rhine river and the impending destruction of the natural rare earth element distribution in rivers. *Earth and Planetary Science Letters*, 362:43–50, 2013.

Appendix A

Supporting information for Chapter 2

A.1 Weighted Kaplan-Meier estimation of survival function for analyzing left-censored data

The weighted Kaplan-Meier (KM) estimator $(\widehat{S}(x_i))$ is expressed as:

$$\widehat{S}(x_i) = \prod_{x \geq x_i} \left(1 - \frac{d_i^w}{Y_i^w}\right)$$

where, for application to aqueous chemistry data, $\widehat{S}(x_i)$ is estimate of the probability of any measured concentration, x, from the population being *less* than x_i. It is calculated using the weighted count of uncensored observations at x_i (d_i) and the weighted count of all (censored or uncensored) observed concentrations less than x_i (Y_i^w). The weighting of each data point is calculated here as the inverse of the number of samples from the data source, or $w_i = n_i^{-1}$. This curve can be used to estimate the quantiles of the data distribution as well as make estimates of the mean and variance of the population. The computational methods here are adapted from Xie and Liu [1] and Singh et al. [2].

To demonstrate these calculations, random data, drawn from log-normal distributions, will be used, followed by analysis of the groundwater REE dataset. This analysis makes use of functions from the `plyr` (V 1.8.1), `dplyr` (V 0.3.0.2), and `tidyr` (V 0.1) packages which must be installed to run this analysis. Aside from the pipe operator (`%>%`, loaded via the `dplyr` namespace, but part of the `magrittr` package), functions from these namespace are denoted as `package_name::function_name`, e.g. `dplyr::mutate`.

Code, written in R, is shown with a grey-shaded background:

```
# This is an R comment
X <- 10
Y <- rnorm(1000)
```

Conversely, output from R calculations will be shown as:

```
##
## These are results
##
```

A.1.1   Generation of random data

```
## If these packages are not installed,
## un-comment and run these lines
# install.packages('plyr')
# install.packages('dplyr')
# install.packages('tidyr')
library(plyr)
library(dplyr)

# Data are drawn from three separate log-normal distributions
# Group means
means <- c(1,1.5,2.5)

# Samples in each group
N <- 100

# Generate random data
set.seed(8675309)
R <- data.frame(sapply(means, function(mu) rlnorm(N, mu, 1)))
colnames(R) <- c('Grp1','Grp2','Grp3')
```

```
head(R) # Look at first few values of random data, R

# 'mutate' adds censoring and site ID
R.g <- tidyr::gather(data = R, key = Dataset,
                     value = Concentration,
                     contains('Grp')) %>%
        # Assume that 60% of samples were analyzed by
        # method with DL = 1 ppb, 40% with DL = 10 ppb
  dplyr::mutate(DL = sample(c(1, 10),
                            size = nlevels(Dataset)*N,
                            replace = T,
                            prob = c(0.6,0.4)),
        # Flag censored values and store at detection limit
        Censored = ifelse(test = Concentration < DL,
                          yes = 1, no = 0),
        Concentration = ifelse(Censored == 1,
                               DL, Concentration),
        # Randomly assign to sites w/ non-uniform probability
        Site = sample(LETTERS[1:10],
                      nlevels(Dataset)*N,
                      replace = T,
                      prob = 1:10/sum(1:10))) %>%
  dplyr::select(Dataset, Concentration, Censored, Site)

# Look at first few rows of fictional, partially censored data.
dplyr::tbl_df(R.g)
```

```
## Source: local data frame [300 x 4]
##
##    Dataset Concentration Censored Site
## 1    Grp1         10.000        1    D
## 2    Grp1         10.000        1    J
## 3    Grp1         10.000        1    I
## 4    Grp1         20.685        0    H
## 5    Grp1          7.889        0    J
## 6    Grp1         10.000        1    I
## 7    Grp1          2.794        0    H
## 8    Grp1         10.000        1    I
## 9    Grp1          4.817        0    G
```

```
## 10    Grp1        10.000        1     F
## ..     ...          ...         ...  ...
```

A.1.2   Functions of the weighted Kaplan-Meier routine

To utilize the weighted KM, the weights of each observation must be calculated.

```
calc_weights <- function(site_vector){
  counts <- data.frame(Site = site_vector) %>%
    dplyr::group_by(Site) %>%
    dplyr::summarise(weight = 1/n())
  return(counts)
}


# Weights of sites are very similar across groups
# Note that no samples were assigned to "Site" A in Grp3,
# thus the weight is NA
plyr::ddply(R.g, .(Dataset), function(df){
    calc_weights(df$Site)
  }) %>%
  tidyr::spread(Dataset, weight)
```

```
##      Site     Grp1     Grp2     Grp3
## 1       A 0.25000 0.20000       NA
## 2       B 1.00000 0.20000 1.00000
## 3       C 0.16667 0.16667 0.20000
## 4       D 0.14286 0.16667 0.14286
## 5       E 0.25000 0.14286 0.08333
## 6       F 0.08333 0.08333 0.08333
## 7       G 0.12500 0.06667 0.12500
## 8       H 0.05000 0.10000 0.05263
## 9       I 0.06667 0.06667 0.05556
## 10      J 0.04348 0.05263 0.05556
```

Next, the function for calculating the survival estimator is built.

```
Surv_weighted <- function(censored_data){
  # Data should have the following form [R data type]:
  # Column 1 = 'Concentration' [double]
  # Column 2 = 'Censored' [bool/int]
```

```
# Column 3 = 'Site' [factor]

if(!is.factor(censored_data$Site)){
  censored_data$Site <- factor(censored_data$Site)
}

# Get weights
site_weights <- calc_weights(censored_data$Site)

# Combine weights with original data
data_mod <- dplyr::left_join(censored_data,
                             site_weights,
                             by = 'Site')

# Calculate the 'at risk' observations, or
# weighted # of observations less than each concentration
data_mod <- dplyr::arrange(data_mod,
                           # Data in descending order
                           desc(Concentration)) %>%
  dplyr::mutate(Yw = sum(weight) - cumsum(weight) + weight)

# Retain only uncensored (Censored == 0) observations
observed <- dplyr::filter(data_mod, Censored == 0) %>%
  dplyr::select(-Censored)

# All observations with identical values are counted
obs_weight_tab <- dplyr::group_by(observed, Concentration) %>%
  dplyr::summarize(dw = sum(weight))

# Combine weighted counts with rest of data and calculate S
observed <- dplyr::left_join(observed, obs_weight_tab,
                             by = "Concentration") %>%
  # "incremental survival", value inside product
  dplyr::mutate(P = 1 - dw/Yw) %>%
  # Remove duplicated observations
  dplyr::filter(!duplicated(Concentration)) %>%
  dplyr:: mutate(S = cumprod(P),
                 S = ifelse(S<0,0,S))
```

```
  return(observed)
}
```

Using the simulated data, these functions can be tested:

```
# Calculate weighted KM for each dataset/group
sim_wKM <- R.g %>% plyr::ddply(.(Dataset), function(df){
  df <- dplyr::select(df,-Dataset)
  wKM <- Surv_weighted(df)
  return(wKM)
})

# Look at results
sim_wKM %>%
  dplyr::select(Dataset, Concentration, weight, S) %>%
  dplyr::tbl_df()
```

```
## Source: local data frame [207 x 4]
##
##     Dataset Concentration  weight      S
## 1      Grp1        20.685 0.05000 0.9950
## 2      Grp1        20.505 0.16667 0.9783
## 3      Grp1        19.803 0.08333 0.9700
## 4      Grp1        19.540 0.04348 0.9657
## 5      Grp1        14.310 0.06667 0.9590
## 6      Grp1        13.423 0.06667 0.9523
## 7      Grp1        13.130 0.04348 0.9480
## 8      Grp1        11.808 0.08333 0.9396
## 9      Grp1         9.981 0.25000 0.9038
## 10     Grp1         9.503 0.06667 0.8942
## ..      ...           ...     ...    ...
```

To estimate the quantiles of a dataset with censoring, `Surv_quantile` is defined, which calls `Surv_weighted`. This function will choose the observation closest to the desired quantile, without exceeding that quantile. As in the main text, if the calculated survival quantiles, $S(x)$, for adjacent observations were $S(x_1) = 0.94$ and $S(x_2) = 0.96$, then x_1 would be noted as the 95th percentile. If the desired quantile is outside of the range of calculated (e.g. for $Q(0.05)$ when $\min \widehat{S}(x) > 0.05$), the quantile is estimated parametrically through regression on order statistics (ROS). Here the data are assumed to be adequately fit by a log-normal distribution. The parameters

of this distribution are estimated from quantile-quantile regression.

```
Surv_quantile <- function(censored_data,
                            # Default to median + IQR
                            percentiles = c(0.25,0.5,0.75),
                            sig.fig = 3){
  # Fraction of data below detection
  cenfrac <- mean(as.logical(censored_data$Censored))

  # Calculate wKM and select relevant columnes
  weighted_km <- Surv_weighted(censored_data) %>%
    dplyr::select(Concentration, S)

  # Warnings about data
  if(any(cenfrac > percentiles)){
    warning(paste('The fraction of censored data is larger ',
                  'than one or more desired percentiles.\n',
                  'This may produce unreliable estimates.',
                  sep = ''))
  }

  out_of_range <- c(F,F)

  if(any(percentiles < min(weighted_km$S))){
    warning(paste('At least one desired percentile below',
                  'minimum calculated from data,',
                  'will be estimated from ROS.'), sep = ' ')
    out_of_range[1] <- T
  }

  if(any(percentiles > max(weighted_km$S))){
    warning(paste('At least one desired percentile above',
                  'maximum calculated from data,',
                  'will be estimated from ROS.'), sep = ' ')
    out_of_range[2] <- T
  }

  if(any(out_of_range)){
    ROS_data <- dplyr::filter(weighted_km, !is.na(S)) %>%
      transmute(y = log(Concentration), x = qnorm(S))
```

```
    ROS <- with(ROS_data,
                lm(y ~ x)
                )

    # Function for returning predictions based on ROS
    pred_ROS <- function(ROS_obj, percentile){
      new_dat <- data.frame(x = qnorm(percentile))
      log_pred <- predict(ROS_obj, newdata = new_dat)
      return(exp(log_pred))
  }

  fl.h <- function(percentile, S){
    h.temp <- which.min(abs(S - percentile))
    return(h.temp)
  }

  h.low <- sapply(percentiles, function(p){
      fl.h(p, weighted_km$S)
    })
  perc_df <- data.frame(Percentile = percentiles) %>%
    dplyr::group_by(Percentile) %>%
    dplyr::mutate(h.low = fl.h(Percentile, weighted_km$S),
                  use_ROS = ifelse(
                     (Percentile < min(weighted_km$S) ||
                              Percentile > max(weighted_km$S)),
                     T,F)
                  ) %>%
    dplyr::ungroup() %>%
    dplyr::mutate(Xh = ifelse(use_ROS,
                              pred_ROS(ROS, Percentile),
                              weighted_km[h.low,
                                          'Concentration']),
                      Xh = signif(Xh, sig.fig)

  return(perc_df)
}
```

This estimation is demonstrated with the simulated data. Note that in this example the rates of censoring are high in the first two groups, which elicits

a warning from the `Surv_quantile` code. However, for clarity here, those warning messages have been suppressed.

```
plyr::ddply(R.g, .(Dataset),
      function(df){
        Surv_quantile(dplyr::select(df,-Dataset),
                      percentiles = c(0.05,0.25,0.5,0.75))
                      }) %>%
  dplyr::select(-h.low)
```

```
##     Dataset Percentile use_ROS     Xh
## 1     Grp1       0.05    TRUE  0.714
## 2     Grp1       0.25   FALSE  2.110
## 3     Grp1       0.50   FALSE  2.900
## 4     Grp1       0.75   FALSE  4.910
## 5     Grp2       0.05    TRUE  1.390
## 6     Grp2       0.25   FALSE  2.790
## 7     Grp2       0.50   FALSE  6.060
## 8     Grp2       0.75   FALSE  8.650
## 9     Grp3       0.05   FALSE  1.900
## 10    Grp3       0.25   FALSE  7.500
## 11    Grp3       0.50   FALSE 10.300
## 12    Grp3       0.75   FALSE 19.400
```

References

[1] Jun Xie and Chaofeng Liu. Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in Medicine*, 24(20):3089–3110, 2005.

[2] Anshuman Singh, Edward F Neuhauser, Nicholas A Azzolina, Mark Distler, Krista M Anders, Melissa A Doroski, and Alan J Rabideau. Statistical techniques for analyzing of soil vapor intrusion data: A case study of manufactured gas plant sites. *Journal of the Air & Waste Management Association*, 63(2):219–229, 2013.

B

Supporting information for Chapter 3

B.1 Barium removal and background REE concentrations

A primary objective in analyzing REE in natural water samples is the separation of Ba, which may lead to isobaric interferences with Eu. Figure B.1A illustrates the effective rejection of Ba by the LLE method presented here, however Figure B.1C shows that even with $\sim 0.2\%$ $^{135}Ba^{16}O^+$:$^{135}Ba^+$, the $^{151}Eu^+$ is indistinguishable from $^{135}Ba^{16}O^+$ in the vast majority of samples prior to barite precipitation. Following precipitation (Figures B.1B,D), these issues are resolved.

Other experimental work in our shared lab space involves high concentrations ($\sim$mM) of Gd. We ascribe the uniformly high Gd background in all experiments to cross contamination in this shared space. In the "Blank" experiment (i.e. pH adjusted ASTM Type I water), all analytes were below detection (IDL $\sim 5-20$ ppt for 1% false negative rate) except for Ba, La, and Gd. This indicates that the high La background could either be a result of laboratory cross-contamination (as with Gd) or an impurity in the organic phases used. The latter supposition was investigated by direct contact of the mixed organic phases used (i.e. 3 mL 0.25 M HDEHP in heptane + 1 mL octanol) with 4 mL of 6 N HCl, followed by analysis of the
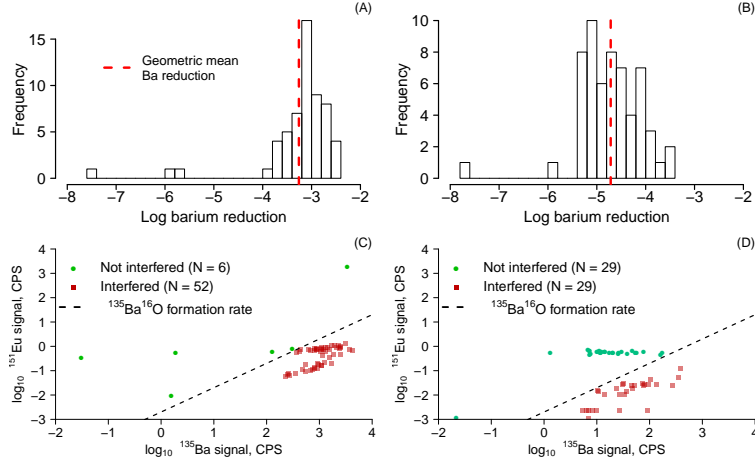
Figure B.1: Efficiency of Ba removal by LLE method (A, C) and ICP-MS octopole collision cell (B, D). Results are for samples without (A, C) and with (B, D) H_2SO_4 addition to precipitate barite. In (B) and (D), $^{135}Ba^{16}O^+$ rate is inferred from 23 replicate analyses of a 200 ppb Ba standard after blank subtraction. Interfered samples are those where accurate Eu determination could not be made due to excessive $^{135}Ba^{16}O^+$ interference. In (D) the interfered samples were all unspiked experiments.

acid phase. These results were below detection (not pictured), indicating no significant REE contamination of the organic phases. While all chemicals were purchased at high purity, we can assume that the observed background concentrations in other experiments are due to trace contamination of these reagents. Paradoxically, the level of these contaminations cannot be determined by ICP-MS without applying a separation/preconcentration technique such as the LLE method; this makes source apportionment of the observed background concentration challenging.
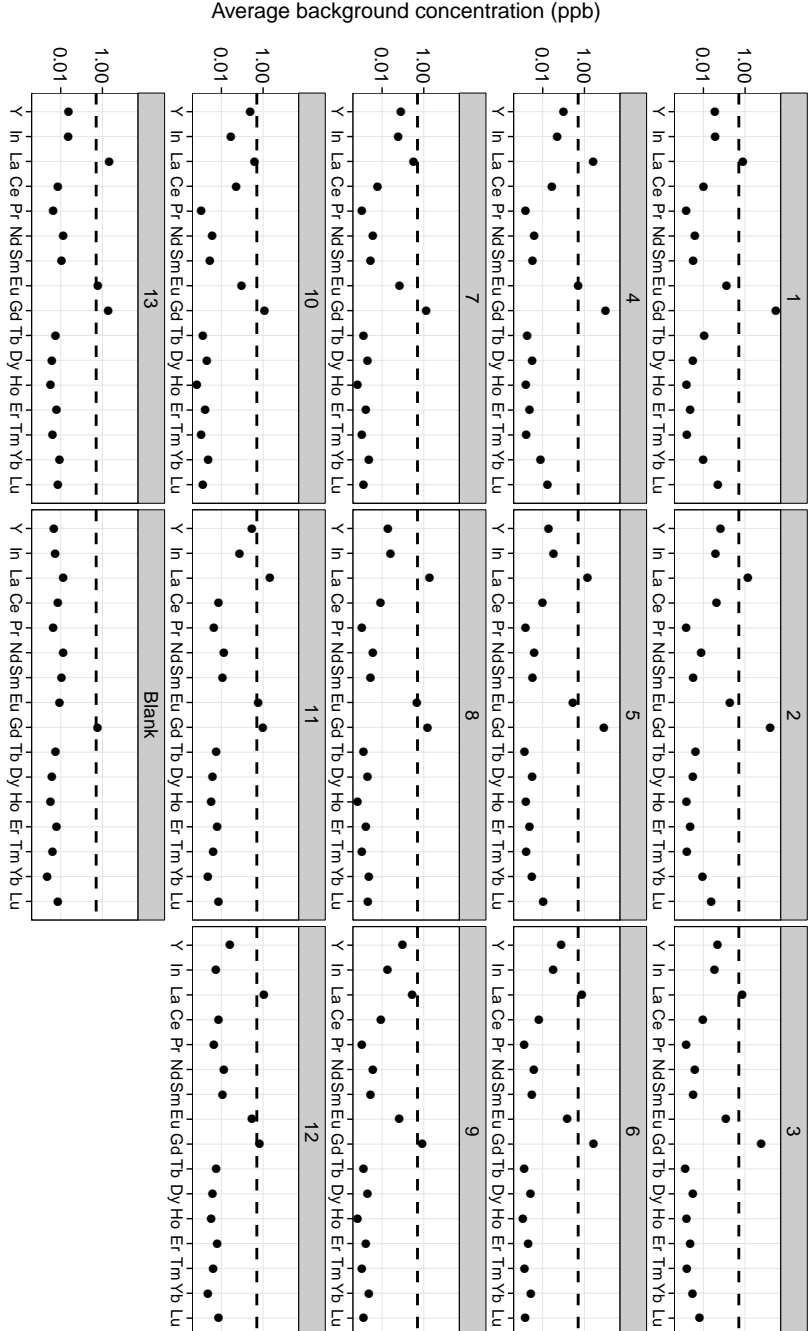
Figure B.2: Average (from $n \geq 2$ replicates, except for Blank and experiments 6, 13) background concentrations of target analytes in samples without $BaSO_4$ precipitation. Dashed line at 500 ppt indicates the input concentration for all spiked samples. Note that the y-axis is a logarithmic scale. Blank experiment represents pH adjusted ASTM Type I water.

B.2 Doehlert experimental results

These data represent the experiment and replicate ordered results of Doehlert matrix tests. Each experiment number represents a unique set of solution conditions ([NaCl], [Fe], and [DOC]); values for each of these parameters in each experiment is given in Table 3.2.
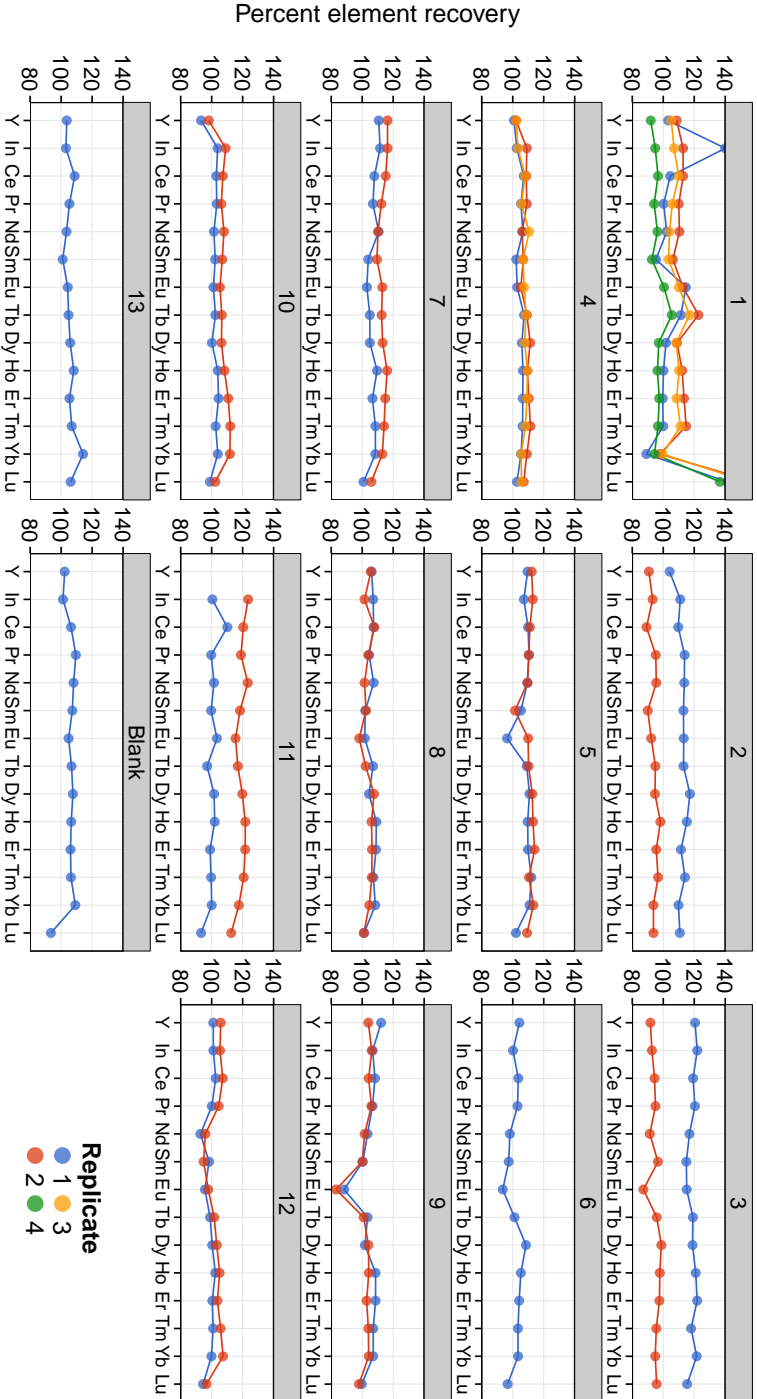
Figure B.3: Elemental recovery in Doehlert matrix samples by LLE methodology (see Table 3.2 for experimental conditions). Recovery values for Eu were determined after dosing the LLE eluent with H_2SO_4 to precipitate barite; all other recoveries were determined without barite precipitation. Elements where the background concentration was determined to 250 ppt or greater (see Figure B.2) were excluded (i.e. La and Gd in all experiments and Y in experiment 11).