Supplemental Information for

Regional Drivers of Stream Chemical Behavior: Leveraging Lithology, Land Use, and Climate Gradients across the Colorado River, Texas USA

G. M. Goldrich-Middaugh[1]

Ma, L.[2]

Engle, M.A.[2]

Soto-Montero, P.[2]

Ricketts, J.W.[2]

Sullivan, P. L.[1]

[1]Oregon State University, College of Earth, Ocean, and Atmospheric sciences.

[2]University of Texas at El Paso, Department of Earth, Environmental, and Resource Sciences.

Contents of this file

Text S1

Site selection and Data

Data for all WQP sites within the Colorado River basin (defined as HUC codes "120800", "120901", "120902", "120903", "120904") was downloaded and included 490 sites and 27,7433 observations. Charge balance was calculated and only sites with a charge balance error of less than 10% were retained. This included 155 sites, and 8,066 observations with a mean absolute charge balance error of 1.32%. These observations were used for concentration-discharge analysis as measurements were assumed to be relatively representative and it was not necessary to include observations of every solute at each site.

Samples were further subset to include dates and locations where Ca^{2+}, Cl^{-}, K^{+}, Na^{+}, HCO_3^{-}, SO_4^{2-}, Si, and Mg^{2+} were all measured (termed complete cases). This subset included 117 sites and 4,686 observations. These observations were used for piper diagrams, PCA, and random forest analysis because it was necessary to include observations for each variable at every site and date.

Analysis of seasonal, tributary, and temperature variations

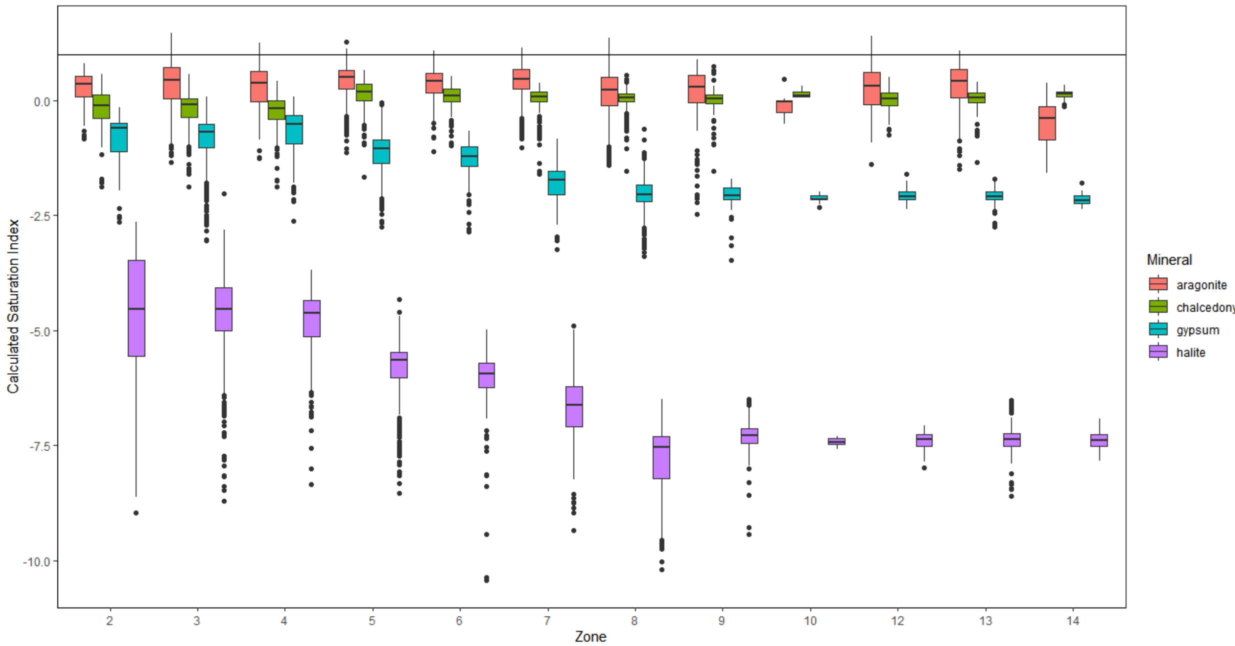

Figure S1 Box plots of the calculated saturation index across the 14 river zones for Aragonite (red), Chalcedony (green), Gypsum (blue), and Halite (purple). Box plots illustrating the median and inter quartile range as well as outlying points for calculated SI in each zone for observations with temperature measurements. A saturation index >0 indicates the potential for mineral precipitation, while a value <0 indicates the potential for mineral dissolution.

Temperature measurements are available for 73% of all complete, charge balanced samples. The mean measured temperature is 19.27°C. Trends in calculated SI by zone for the four minerals of interest are extremely similar when using a subset of observations with temperature measurements as compared to a constant temperature of 25°C.

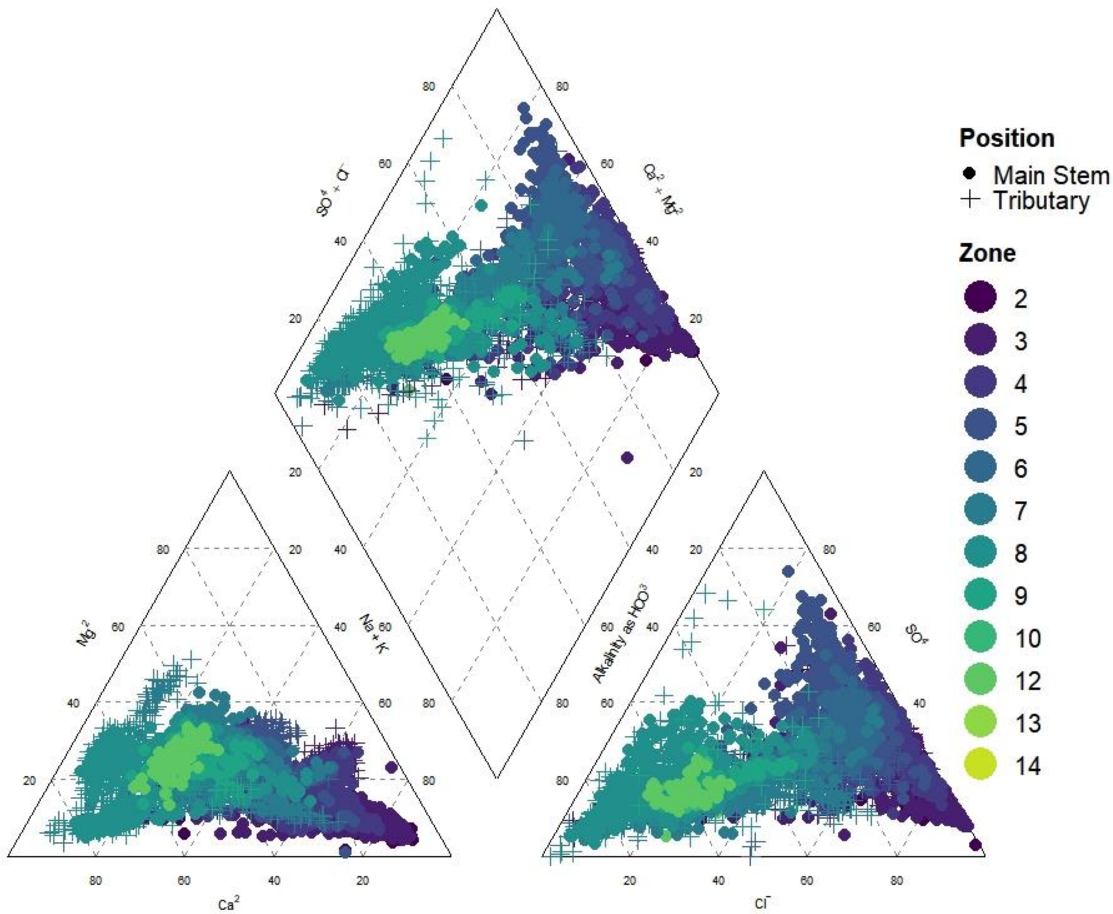

Figure S2 Piper diagram from the Colorado River showing observations based on river zone (2-14) and watershed position (tributary or main stem)

Distinguishing between tributary and non-tributary observations on the piper diagram (Figure S2) shows that overall variations are controlled by watershed position (zone) while some localized variations are influenced by tributary position. For example, samples collected on tributaries in zones 3 and 4 are somewhat higher in Mg^{2+} than samples collected from the main stem in these locations. Additionally, samples collected in tributaries within zone 8 are generally higher in Ca^{2+} and Cl^{-} than those collected from the main stem.

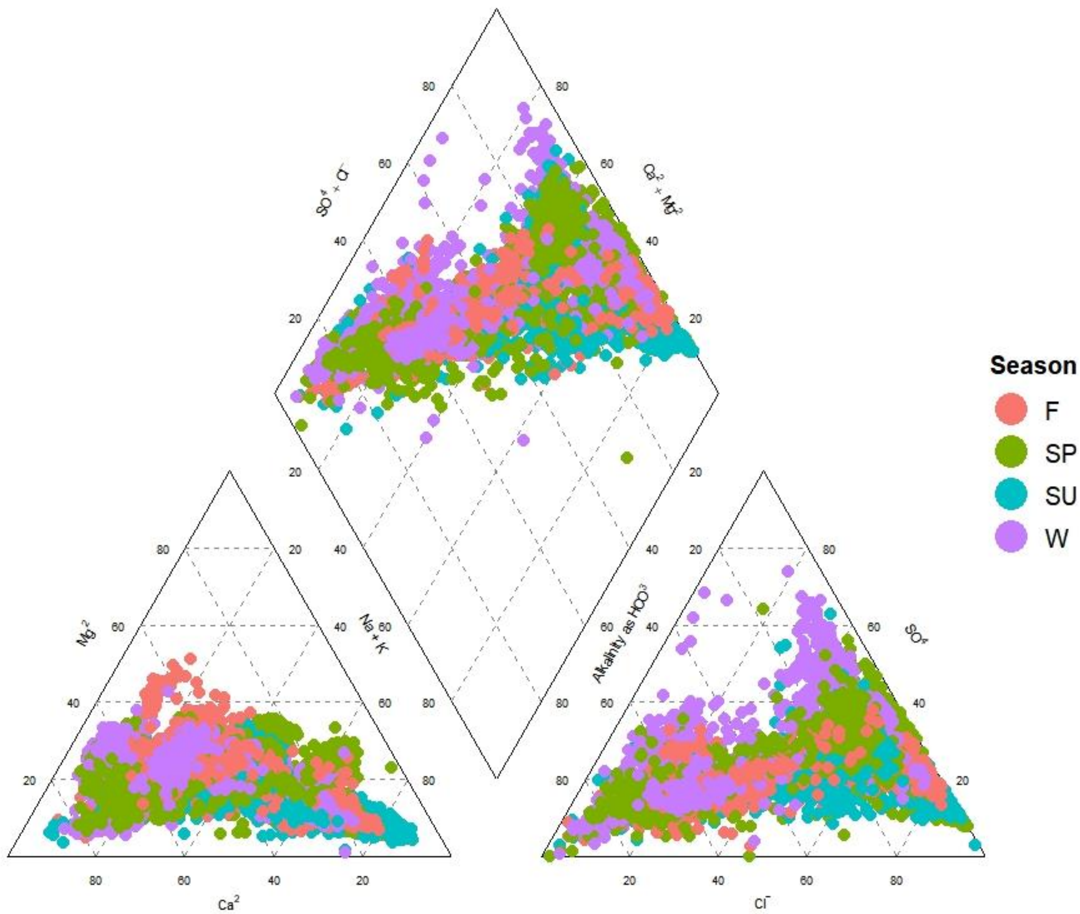

Figure S3 Piper diagram from the Colorado River showing observations colored by season (F=Fall (September, October, November), SP=Spring (March, April, May), SU=Summer (June, July, August), W=Winter (December, January, February))

The composition of stream water across seasons (Figure S3) did not show overarching variations in composition by season which suggests that variations are dominantly dictated by watershed position.

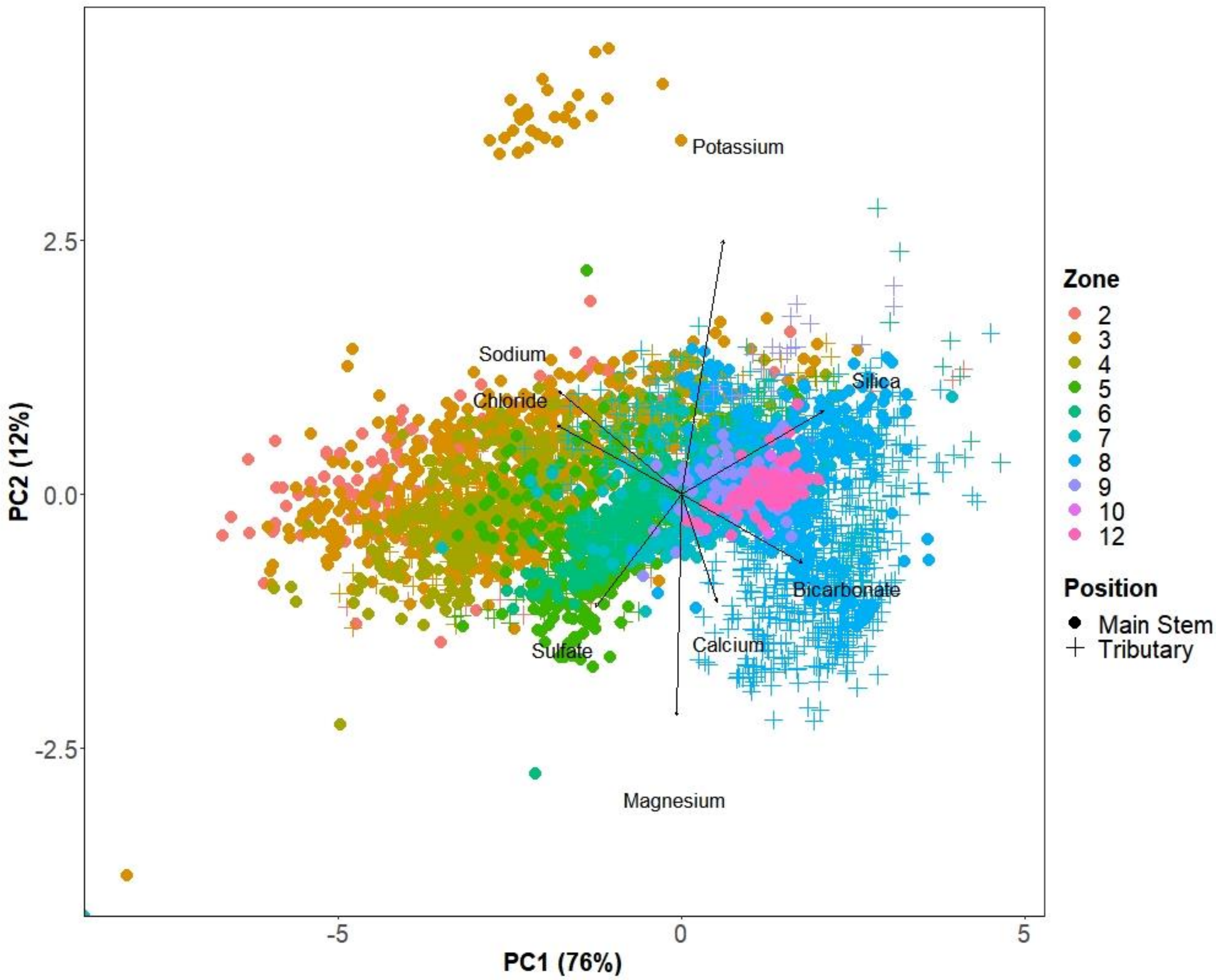

Figure S4 PCA Biplot for the Colorado River showing distribution of observations across PC1 and PC2 colored by river zone (2-12) and watershed position (tributary or main stem)

The compositional PCA analysis also shows that stream type (tributary vs main stem) does not play a dominant role on the distribution of variance observed across PC1 and PC2 (Figure S4). The area where most variation between main stem and tributary points occurs is in zone 8 which is dominated by measurements collected on tributaries near the main stem of the Colorado River.

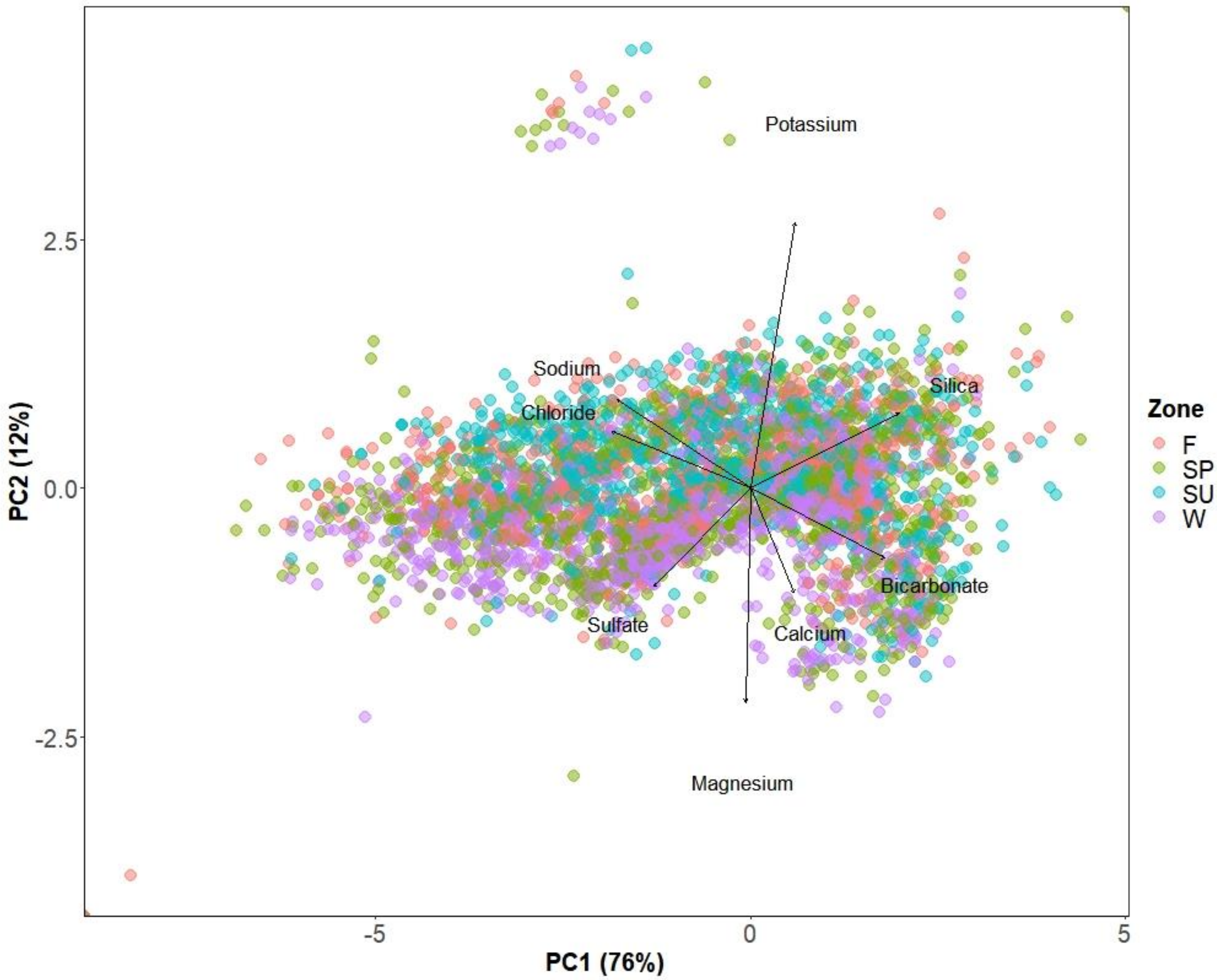

Figure S5 PCA biplot for the Colorado River showing distribution of observations across PC1 and PC2 colored by season (F=Fall (September, October, November), SP=Spring (March, April, May), SU=Summer (June, July, August), W=Winter (December, January, February))

There is not a strong seasonal trend across PC1 which explains 76% of the variation. The distribution of points across PC2 shows some division between measurements collected in summer vs winter indicating that seasonal influence may contribute a very small proportion of the overall variance in the dataset.

Random Forest and Verification

Lithology, LULC (Land use/land cover, and stream chemical data was split into training (80%) and testing (20%) subsets. Individual random forest models were constructed using the training dataset for each watershed factor compared to all solutes using the randomForest package in R (Liaw & Wiener, 2002). The strength of the relationship between each solute and land use factor was assessed using the percent increase in mean square error (%incMSE) if that solute was removed from consideration. A larger %IncMSE indicates a stronger relationship between the two variables Each model was constructed with 500 trees and a minimum terminal node size of 5.

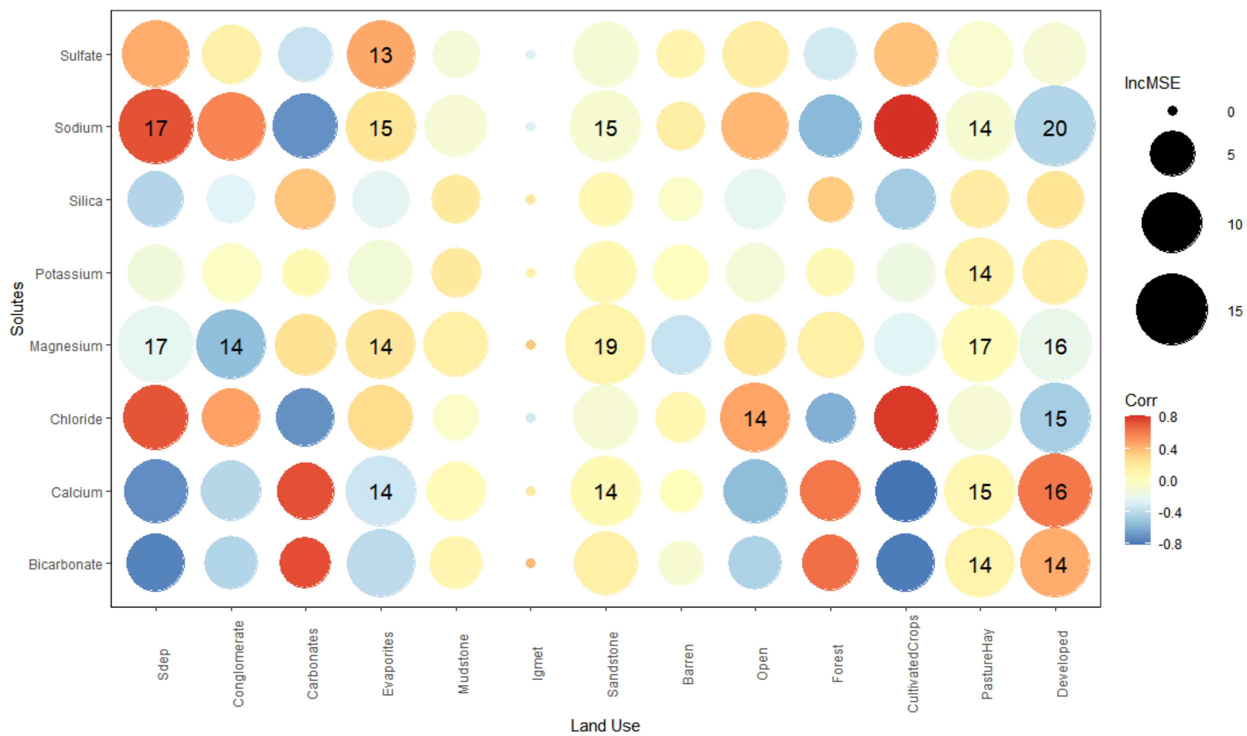

Figure S6 Random forest output for upstream zones (2-5) of the Colorado River including 25 total sites and 2,951 observations. X axis shows lithology and LULC classes and y axis shows solutes. Point size represents magnitude of %IncMSE and point color represents correlation between the two variables. Points with the top 85th percentile and above of %IncMSE are labelled

For the upstream zones (2-5 only) no igneous-metamorphic rocks are present in this region and evaporites are most abundant as compared to the rest of the watershed. Additionally, cultivated crops are common and several small towns represent the largest developed areas. The random forest revealed (Figure S6) that lithologic relationships to stream water chemistry are mostly as-expected with evaporites having strong relationships with Na^+, Ca^{2+}, Mg^{2+}, and SO_4^{2-}. Correlations are low between evaporites and all solutes with the highest correlation between evaporites and sulfate. Sedimentary deposits have the strongest relationships of all lithology classes with %incMSE of 17 for both Na^+ and Mg^{2+}. Sedimentary deposits also have high positive correlations to Na^+ and Cl^- and negative correlations to Ca^{2+} and HCO_3^- .

When we examine all sites (Figure S7) the developed areas and pasture/hay are the most strongly related to all solutes for LULC classes but have weak correlations. Cultivated crops are more

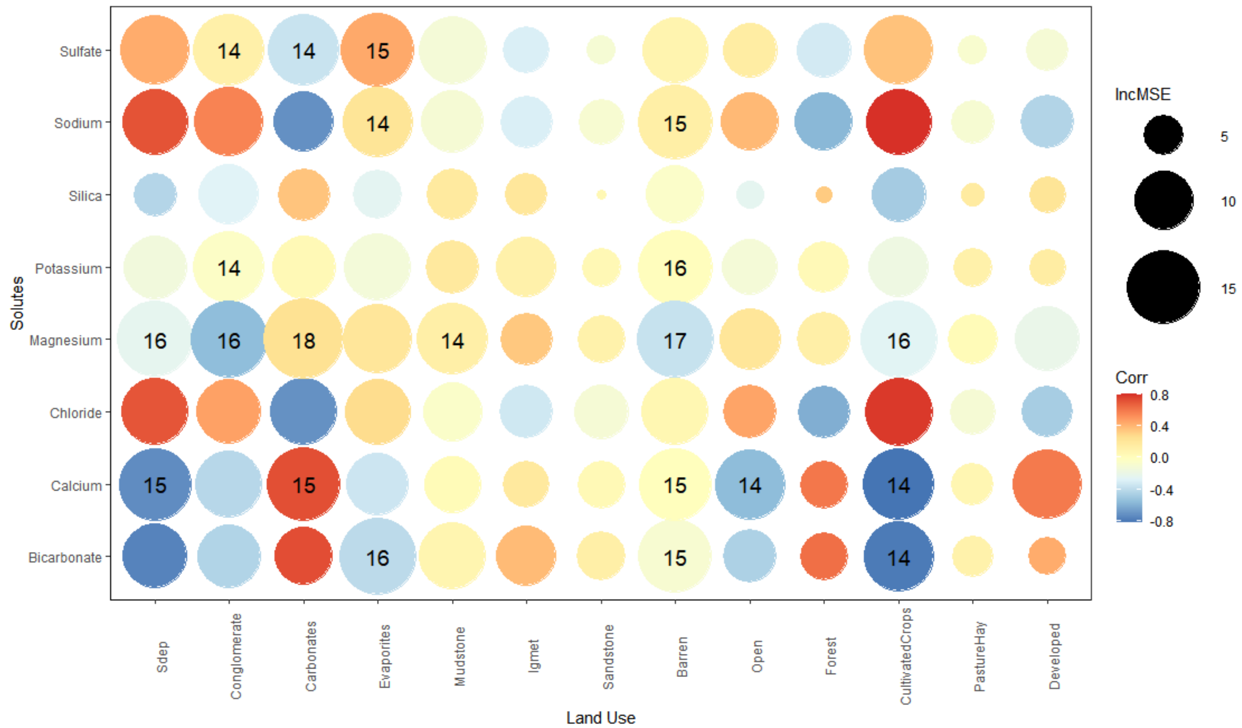

Figure S7 Random forest output for main stem sites on the Colorado River including 36 total sites and 2,686 observations. X axis shows lithology and LULC classes and y axis shows solutes. Point size represents magnitude of %IncMSE and point color represents correlation between the two variables. Points with the top 85th percentile and above of %IncMSE are labelled

highly correlated to Na^+, Cl^-, Ca^{2+} and HCO_3^- but were not shown to have strong relationships via this analysis.

Analysis of sites on the main stem of the Colorado river shows disproportionate influence of Barren areas on many solutes including Na^+, K^+, Mg^{2+}, Ca^{2+}, and HCO_3^- even though correlations are low. Cultivated crops represent a second stand out LULC class with high %incMSE for Mg^{2+}, Ca^{2+}, and HCO_3^- and high correlations for Na^+, Cl^-, Ca^{2+}, and HCO_3^-. A variety of lithologic classes have strong relationships for a few solutes. Sedimentary deposits show high %incMSE for Ca^{2+} and Mg^{2+} and high correlations with Na^+, Cl^-, Ca^{2+} and HCO_3^-. Conglomerate shows high %incMSE for Mg, K, and SO_4. With low to moderate correlations and higher correlations to Na^+. Carbonates have high %incMSE for SO_4^{2-}, Mg^{2+}, and Ca^{2+} and have high correlations for Na^+, Cl^-, Ca^{2+}, and HCO_3^-. Evaporites have weaker correlations but high %incMSE for SO_4^{2-}, Na^+, and HCO_3^-.

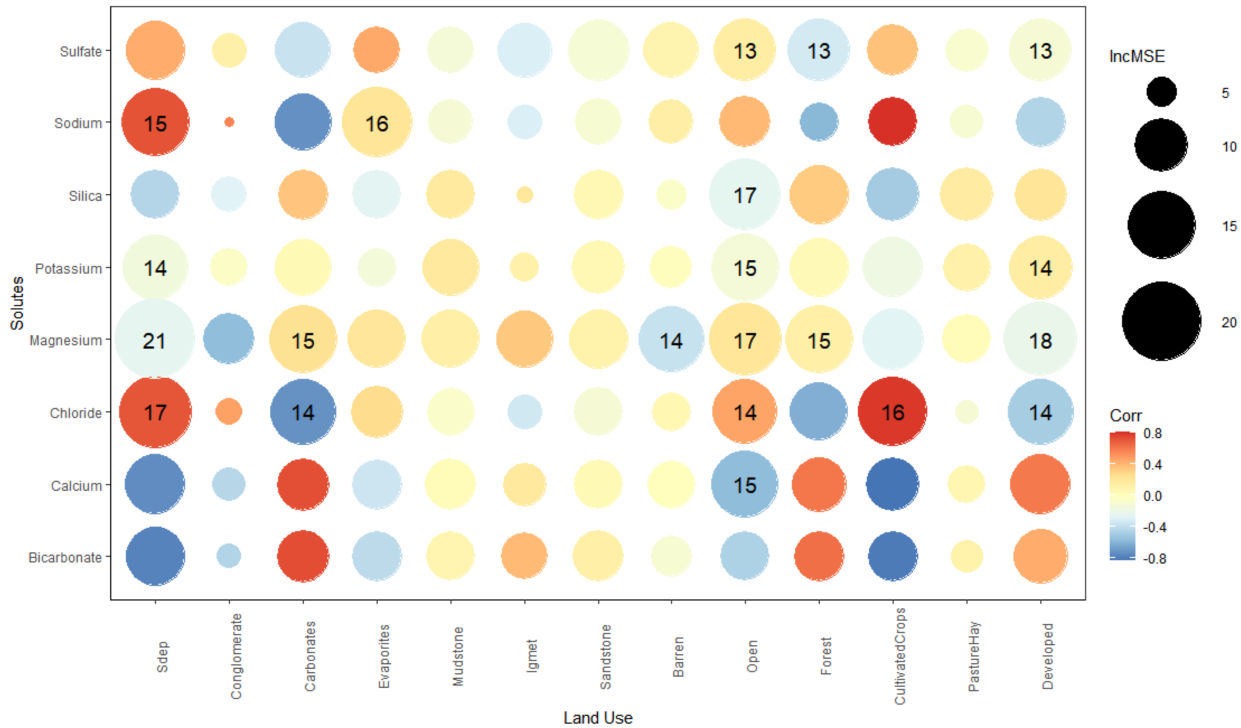

Figure S8 Random forest output for tributary sites on the Colorado River including 44 total sites and 1,766 observations. X axis shows lithology and LULC classes and y axis shows solutes. Point size represents magnitude of %IncMSE and point color represents

It was expected that tributary measurements might have stronger relationships to the underlying watershed factors due to limited mixing and upstream influences as compared to measurements collected from the main stem. The output of the random forest models (Figure S8) shows that some relationships are stronger for tributary sites to a small degree, but the difference is not large. This suggests that water chemical composition between tributaries and the main stem is relatively homogeneous. This is supported by observations using piper diagrams and compositional PCA.