

1 **Regional Drivers of Stream Chemical Behavior: Leveraging Lithology, Land Use,** 2 **and Climate Gradients across the Colorado River, Texas USA**

3 **Authors:**

4 G. M. Goldrich-Middaugh¹, goldricg@oregonstate.edu, ORCID: 0000-0002-5899-8016

5 Ma, L.², email: lma@utep.edu, ORCID: 0000-0001-6240-2564

6 Engle, M.A.², email: maengle@utep.edu, ORCID: 0000-0001-5258-7374

7 Ricketts, J.W.², email: jricketts@utep.edu

8 Soto-Montero, P.², email: pisotomontero@miners.utep.edu, ORCID: 0000-0003-1047-5192

9 Sullivan, P. L.¹, email: pamela.sullivan@oregonstate.edu, ORCID: 0000-0001-8780-8501

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

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

12 Corresponding author: Pamela Sullivan (pamela.sullivan@oregonstate.edu)

13 Key Points (140 characters)

- 14 • The distribution of reactive minerals (e.g., evaporites & carbonates) dictates the chemical
15 behavior of the Colorado River
- 16 • Climate factors act as a secondary control, with increasing precipitation leading to
17 decreased overall concentrations and dilution behavior
- 18 • Agriculture, urban development, reservoirs, and oil and gas wells lead to noise in
19 chemical measurements and obscures natural signals

20 **Abstract (250 words)**

21 Understanding relationships between stream chemistry and watershed factors: land use/land
22 cover, climate, and lithology are crucial to improving our knowledge of critical zone processes
23 that influence water quality. We compiled major ion data from more than 100 monitoring
24 stations collected over 60 years (1958-2018) across the Colorado River Watershed in Texas
25 (103,000 km²). We paired this river chemistry data with complementary lithology, land use,
26 climate and stream discharge information. A combination of graphical geochemistry and
27 machine learning techniques were used to produce new insights on controls of stream water
28 chemical behavior. Studies on stream flow and chemistry in the American west and globally
29 have shown strong relationships between major ion chemical composition and lithology, which
30 hold true for the Colorado River basin in this study. Reactive minerals, including carbonates and
31 evaporites, dominate major ion chemistry across the upper watershed. Upstream and central

32 reaches of the Colorado River showed shifts from Na-Cl-SO₄ dominated water from multiple
33 sources including dissolution of gypsum and halite in shallow groundwater, agricultural
34 activities, and oil and gas development, to Ca-HCO₃ water types controlled by carbonate
35 dissolution. In the lower portion of the watershed multiple analyses demonstrate that stream
36 chemistry is more influenced by greater precipitation and the presence of relatively fewer
37 reactive silicate minerals than middle and upstream reaches. This study demonstrates the power
38 of applying machine learning approaches to publicly available long term water chemistry
39 datasets to improve the understanding of water and nutrient cycling, salinity sources, and water
40 use.

41 **Plain Language Summary (200 words)**

42 Across the United States public and private users rely on large rivers for access to potable water,
43 making water quality of crucial concern. Water quality measurements are widely available but
44 require intensive pre-treatment due to their irregular collection across space and time. Here,
45 public water quality measurements from the Colorado River basin were used to understand the
46 influence of different land use, geologic, and climate factors on water quality. The Colorado
47 River runs across a range of rock types, land uses, and precipitation regimes and therefore
48 displays complex interactions with the land surface that produce changes in water quality. We
49 found that the upper Colorado River is dominated by Na⁺, Cl⁻, and SO₄²⁻ which are derived from
50 multiple sources including agriculture, oil and gas activity, and rock salt dissolution in shallow
51 groundwater. The middle reaches of the Colorado are dominated by Ca²⁺ and HCO₃⁻ which are
52 mainly contributed by the large areas of limestone bedrock. Downstream reaches of the river
53 show more inputs from precipitation as well as potential seawater mixing in coastal areas.
54 Overall, these techniques were effective in demonstrating large scale trends across this watershed
55 and could be improved with more detailed datasets.

56 **Keywords:** critical zone, machine learning, historical data, water quality, regional hydrology

57

58 **1. Introduction**

59 River chemical composition and behavior reflects geochemical, hydrological, and
60 anthropogenic factors of its upstream contributing areas (Ameli et al., 2017; Gaillardet et al.,
61 1999; Godsey et al., 2009). Understanding how spatially varying watershed factors influence
62 stream chemistry at large scales is crucial for illuminating how hydrologic processes influence
63 global biogeochemical cycles. Particularly, synthetic analyses of coexisting river chemistry
64 datasets with other datasets such as discharge, bedrock, soil types, and land uses are needed to
65 improve understanding of how spatial heterogeneity in lithology, climate, and land use/land
66 cover (LULC) impacts overall chemical behavior and transport of materials within large
67 watersheds (i.e., critical zone function).

68 Water flow paths, residence times, and water-rock interactions within the critical zone
69 govern weathering reactions, the rate of solute transport, and stream water chemical composition
70 and behavior (Ameli et al., 2017; Baronas et al., 2017; Bouchez et al., 2017; Chen et al., 2014;

71 Chorover et al., 2017; Dupré et al., 2003; Gaillardet et al., 1999; Stewart et al., 2022; Torres et
72 al., 2017). Flow paths and residence times are controlled by a combination of topography,
73 lithology, climate, and anthropogenic factors. The degree of influence of each of these factors
74 varies both spatially and temporally. Thus, predicting stream chemistry at any one point within a
75 watershed is challenging. Many studies have attempted to capture the degree of influence of
76 different watershed factors through concentration-discharge (C-Q) analysis at both event, and
77 watershed scales (Abbott et al., 2018; Ameli et al., 2017; Evans & Davies, 1998; Godsey et al.,
78 2009; Herndon et al., 2015; F. Liu et al., 2017; Minaudo et al., 2019; Torres et al., 2017).
79 Significant research has also been conducted focusing on multivariate analyses that capture
80 landscape influences of stream chemistry at a range of scales (Kaushal et al., 2013; X. Liu et al.,
81 2021; Park & Lee, 2020; Sliva & Williams, 2001; Tiwari et al., 2017; Xu et al., 2021). More
82 recently, machine learning techniques, such as random forest algorithms, have been employed to
83 understand the degree to which topography, lithology, and climate govern stream discharge
84 behavior (Addor et al., 2018; Hammond et al., 2021; Konapala & Mishra, 2020; Kratzert, Klotz,
85 Herrnegger, et al., 2019; Kratzert, Klotz, Shalev, et al., 2019; Oppel & Schumann, 2020; Rice et
86 al., 2016), but to the best of our knowledge their use in explaining stream water chemistry has
87 been limited (Nearing et al., 2021). Indeed, machine learning techniques offer a promising and
88 flexible approach to quantify and represent relationships in complex systems and to identify
89 important drivers across a range of environments.

90 Arising from many studies is the consensus that lithology is one of the most influential
91 watershed factors controlling the composition and behavior of geogenic solutes in streams
92 (Dupré et al., 2003; Gaillardet et al., 1999). Widespread homogeneity and availability of reactive
93 minerals, in combination with long transit times relative to dissolution rates, have been proposed
94 to account for chemostatic behavior across watersheds, meaning concentrations are relatively
95 constant across a wide range of flows (Chorover et al., 2017; Godsey et al., 2009; Kirchner,
96 2003; Knapp et al., 2020; Maher, 2011). For example, studies in the Andes-Amazon basin have
97 shown that in large river basins, longer residence times are associated with more chemostatic
98 behavior, and that relationships between solute concentrations and flow change in response to
99 spatial variations in watershed factors as well as contributions from major tributaries (Baronas et
100 al., 2017; Bouchez et al., 2017; Torres et al., 2017). Chemical composition of river waters also
101 generally reflects subsurface characteristics including lithology and geological structures
102 (Stewart et al., 2022). However, large river systems have complex spatiotemporal interactions
103 within the critical zone which makes quantitative models of direct lithologic (or structural)
104 influence on stream chemical composition and behavior difficult to develop.

105 A second highly influential factor for stream solute behavior is climate (Shen et al., 2021;
106 White & Blum, 1997). Precipitation and evapotranspiration (or the difference of the two, i.e.,
107 *effective precipitation*) dictate water availability and residence time in river channels and the
108 shallow subsurface and hence control the depth distribution of reactive mineral fronts over long
109 time periods (Ameli et al., 2017; Bouchez et al., 2017; Brantley & Lebedeva, 2020; Maher,

110 2011; Torres et al., 2017). Effective precipitation (defined as [precipitation – evapotranspiration]
111 / precipitation) represents the fraction of precipitation that remains as available water. Negative
112 values indicate actual evapotranspiration exceeds precipitation while positive values indicate that
113 precipitation exceeds actual evapotranspiration and will infiltrate or run-off. Increased available
114 precipitation leads to decreased water residence times, increased solute dilution behavior, and
115 decreased coupling of river chemical behavior with lithologic influences (Maher, 2011; Torres et
116 al., 2017). Hence, under low effective precipitation lithology and land use are more prominent
117 factors influencing stream water chemistry compared to areas of high effective precipitation.

118 Finally, anthropogenic factors influence riverine chemistry and behavior through
119 processes such as agricultural runoff, waste water discharge, regulation of flows by reservoirs,
120 extraction of stream water, and alterations to hydrologic function via land use changes (Kaushal
121 et al., 2013; Liu et al., 2021). Anthropogenic factors encompass myriad potential effects as they
122 are broad and varied. For example, stream chemistry and discharge can be significantly different
123 from natural conditions where urban and agricultural land represent a large fraction of
124 watershed's area due to inputs of fertilizers, industrial additions, agricultural withdrawals,
125 industrial and municipal withdrawals, regulation via reservoirs (Aitkenhead-Peterson et al.,
126 2011; Chen et al., 2014; Musolff et al., 2015). Outside of direct impacts to flow regimes, studies
127 have shown that special C-Q behaviors can occur in anthropogenically modified systems.
128 Specifically, addition behavior (increase in concentration over large increase in discharge) has
129 been observed for some nutrient-based solutes, including NO₃-N and PO₄ (Musolff et al., 2015).
130 At a sub-watershed scale, agricultural areas have been shown to influence surface and shallow
131 groundwater quality (Liu et al., 2021; Park & Lee, 2020). Understanding the magnitude of
132 human impacts on chemical behavior in large watersheds that cross many land use classes
133 remains elusive.

134 To improve our understanding of how spatially heterogenous catchment characteristics,
135 including lithology, climate, and anthropogenic (e.g., LULC) factors, influence general stream
136 chemical behavior at a regional scale, the objective of this research is to address our main
137 research question: What are the most influential watershed factors governing variations in stream
138 water chemistry in large rivers that cross a range of lithology, climate, and LULC types? More
139 specifically, we apply a top-down approach to test the following three hypotheses through
140 examination of coexisting water chemistry, discharge, lithology, climate, and LULC datasets in
141 the Colorado River watershed, Texas:

142 Hypothesis 1: Lithology is the strongest control on stream chemistry where there is an abundance
143 of reactive minerals in the near-surface. This setting leads to high concentrations of
144 geogenic solutes, and chemostatic behavior.

145 Hypothesis 2: Climatic condition, measured as effective precipitation, governs flow paths, water
146 residence times, and solute generation. In areas of high effective precipitation ($P \gg ET$),
147 low topographic relief, and/or high permeability, reactive mineral weathering fronts occur

148 at greater depths, water residence times are shorter, and systems are source limited and
149 exhibit more dilution behavior.

150 Hypothesis 3: Anthropogenic influences obscure natural controls on stream water chemistry but
151 the intensity and spatial spread of influence depends on the type and degree of
152 development or contamination.

153 To test these hypotheses, we analyzed a 60 year-long (1958-2018), publicly available major ion
154 chemical and discharge dataset using graphical techniques (Piper and Gibb's diagrams), C-Q
155 relationships, geochemical modeling, and machine learning methods. We analyzed these data
156 with spatially explicit land cover, climate, and lithology data from various public sources. We
157 then used random forest algorithms to assess the importance of these factors in explaining the
158 stream water chemistry. Overall, the data suggested that lithology was the first order control of
159 stream water chemistry at regional scales, while climate exerted a secondary influence.

160 Anthropogenic factors (particularly agricultural activities and oil and gas development) were
161 found to be important for some geogenic species but their influence was difficult to separate
162 from lithologic sources using only major ion chemical data.

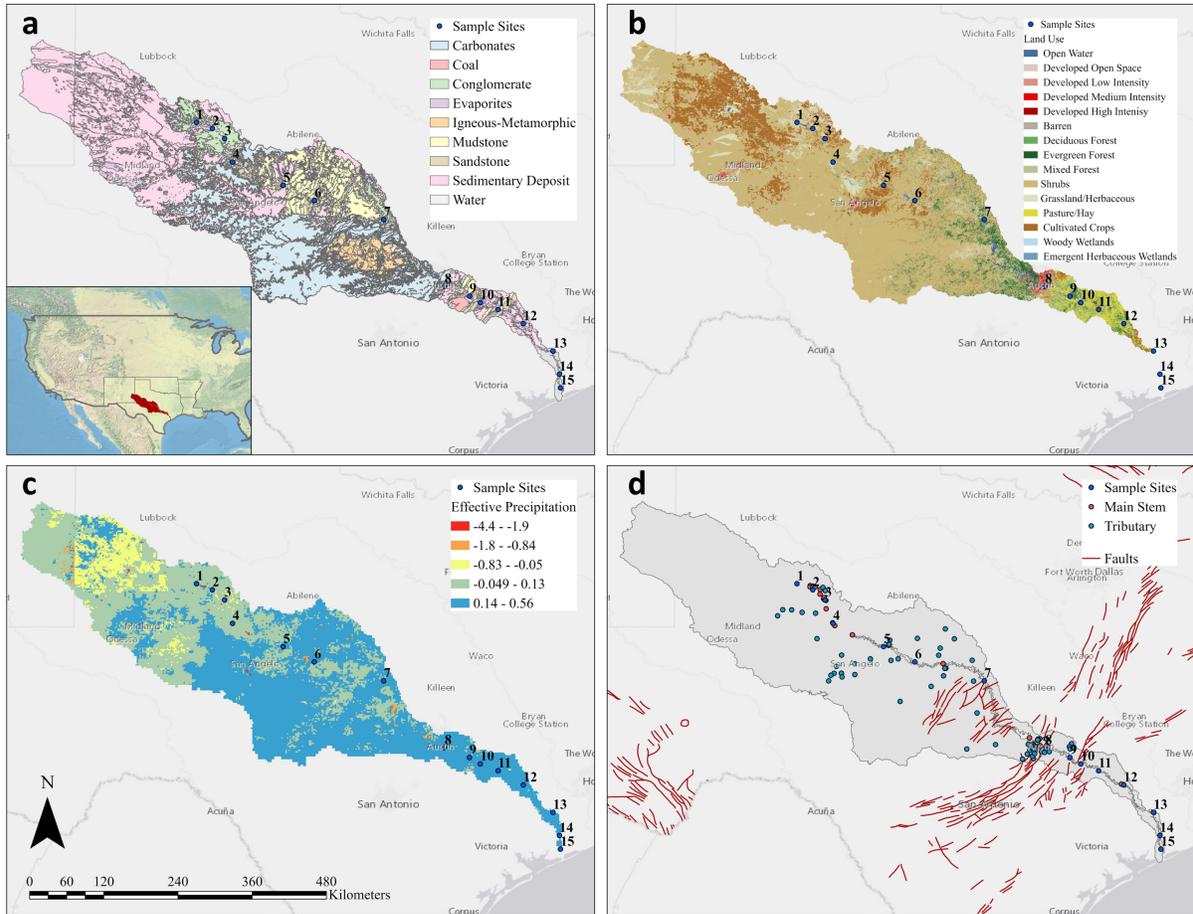
163 **2 Materials and Methods**164 **2.1 Study area: Contrasting lithologic and land use patterns across the Colorado River**
165 **Basin, Texas**

Figure 1 Study area: Colorado River a. simplified lithology, b. distribution of LULC classes with the Austin metropolitan area shown in red underlying sampling point 8 c. distribution of effective precipitation, and d. location of water chemistry sampling points on tributaries (blue) and main stem (red) with 15 flow sampling sites labelled (1-15) and faults shown in light red

166 We focus on the Colorado River of New Mexico and Texas because climate, LULC, and
 167 lithology changes across the watershed set the stage for spatially contrasting chemical behaviors.
 168 The river is approximately 1,288 km long, with 5 major tributaries and a watershed area that
 169 covers 15% of Texas (Colorado River Alliance, 2021, Texas Natural Resource Conservation
 170 Commission, 1999). It is also a crucial municipal water source for large and small communities
 171 in the state. Lithologic units across the watershed are roughly oriented NE to SW, perpendicular
 172 to flow in the main stem (Figure 1a). This systematic change in lithology and thus mineralogy
 173 across the watershed may control solute contributions to the Colorado River. Quaternary
 174 sedimentary deposits (conglomerates, sandstones, and mudstones), Cretaceous sedimentary
 175 carbonates, and gypsum and halite units are common rock types across the Colorado River basin
 176 (Clark et al., 2020). Karst features are common and host several large and productive aquifers

177 that intersect the Colorado River watershed near the Gulf Coast. The major structural feature
178 affecting the Colorado River is the inactive late Oligocene to early Miocene Balcones fault zone,
179 which is a network of NW-striking high-angle normal faults (Figure 1d) that predominantly dip
180 southeast (Clark et al., 2020). The upper Colorado River is dominated by carbonate outcrops
181 with gypsum inclusions (Leifeste & Lansford, 1968; Richter et al., 1991). The lower portions of
182 the watershed are dominated by sedimentary deposits of varying lithologies. Overall, the
183 lithologic shifts lead to general trends from more (evaporite and carbonates) to less (silicate
184 minerals) reactive mineralogy moving downstream in the watershed. Thus, we expect that the
185 upstream reaches of the Colorado River will be strongly influenced by the abundance of reactive
186 mineral assemblages including evaporites and carbonates, the middle reaches will be influenced
187 by a combination of reactive mineralogy (carbonates) and increased agricultural and urban
188 influences, and the outlet of the river will be influenced dominantly by other factors as the
189 lithology (silicate sediments) is more resistant to chemical weathering.

190 Spatial changes in climate across the Colorado River watershed strongly influence the
191 hydrodynamics of the system, and thus can exert control on the chemical composition of the
192 river. Similar to lithologic shifts perpendicular to the river course across the watershed,
193 precipitation increases from NW to SE with the semi-arid western portion receiving an average
194 of 250 mm of rain per year, and the sub-humid east receiving more than 1,500 mm (Harwell et
195 al., 2020; Texas Natural Resource Conservation Commission, 1999; Texas Water Development
196 Board, 2012). Across this area, the annual average temperature changes but not as significantly
197 as the precipitation. Average annual evapotranspiration also increases in accordance with the
198 increase in available water (NW to SE), except in human-modified areas where management of
199 reservoirs and irrigated agriculture occur. A survey of trends in mean annual precipitation, air
200 temperature, and streamflow shows moderate increases in precipitation in the eastern portion of
201 the state and increases in air temperature of 0.6 °C per 50 years across the state leading to
202 increased potential evapotranspiration (Harwell et al., 2020). Overall, the combination of
203 precipitation and evapotranspiration leads to a spatial trend of increasing effective precipitation
204 (-4.4 to 0.56), likely decreasing water residence times from northwest to southeast across the
205 watershed (Figure 1c).

206 Anthropogenic influences are variable in distribution and impact across the watershed
207 and encompass oil and gas development activities in the northwest and southeast, agriculture in
208 the center and southeast, reservoir construction in the upper and middle reaches, and intensive
209 urban development in the southeast, each of which can have distinctive controls on stream solute
210 load and behavior. The northwestern portions of the watershed are generally dominated by
211 shrubs, grassland, and cultivated crops with the presence of diffuse, moderately developed areas.
212 Evergreen and deciduous forest, high intensity development, pasture and hay roughly follow
213 lithologic trends and appear near the Balcones fault zone (Figure 1b and d). Population gradients
214 are present with many major cities centered in the southeast portion of the watershed including
215 the Austin metropolitan area (population of 2.3 million) (US Census Bureau, 2019) which is
216 shown in red around flow sampling site 8 (Figure 1c). Population growth over the next 50 years

217 is also projected at a much higher rate in the eastern half of the watershed than the western half
218 (Scanlon et al., 2005; Texas Water Development Board, 2017). Agricultural production is central
219 to the state and irrigation made up about 25.8% of total water use in the state in 2010 totaling
220 20.78 million m³ per day (Dieter et al., 2018). Many dams are present in the watershed with six
221 major reservoirs on the main stem of the Colorado River including E.V. Spence, O.H. Ivie, and
222 J.B. Thomas. These reservoirs are dominantly for flood control, municipal, and agricultural
223 water use. Trace metal concentrations along the river are also higher in areas of increased
224 development (Chen et al., 2014). It is projected that over the next 50 years the population of the
225 state will increase by 70% while the demand for water will increase by 17% and available water
226 resources will decrease by 11% (Texas Water Development Board, 2017). These changes are
227 likely to impact not only available water and flow regimes, but also the chemical composition of
228 waters across the state.

229 2.2 Data collection and processing

230 2.2.1 Temporal Data

231 Daily flow measurements were extracted from the USGS NWIS database using R
232 package `dataRetrieval` for the earliest recorded date to the current date (or the full period of
233 record) whenever available (De Cicco et al., 2018). The Colorado River has 15 USGS-
234 maintained flow measurement sites (Sites 1-15 on Figure 1), 14 of which (Sites 1-14) have daily
235 flow measurements on the river main stem. In total, 360,670 flow measurements are available
236 across the 14 sites between 1916 and 2020.

237 All water chemistry data were obtained using the water quality portal (WQP) R tool
238 `dataRetrieval` (De Cicco et al., 2018). In the Colorado River basin, which includes the main stem
239 and several major tributaries, we used a total of 108 sites with water chemistry data ranging from
240 1958 to 2018 for our analysis. The data were spatially and temporally irregular with, on average,
241 at least one and at maximum 39 samples collected at each site per year. Here water chemistry
242 measurements of interest included: alkalinity as HCO₃⁻, K⁺, Na⁺, Mg²⁺, P, SO₄²⁻, Ca²⁺, Cl⁻, Si,
243 NO₃⁻, pH, and water temperature. The 108 chemical parameter sampling sites (Figure 1d) were
244 grouped by proximity to the 15 USGS-flow sampling sites and labelled zones 1-15 (upstream to
245 downstream) for further referencing continuity (Figure 1a-d). Measurements were filtered to
246 include sites where all solutes were measured on the same date. Samples included in analysis
247 were assumed to be representative of major constituents and free of significant analytical error if
248 they had a low charge balance error ($\leq 10\%$; Godsey et al., 2009; Güler et al., 2002)

249 2.2.2 Spatial data

250 Sub-watersheds were delineated using the NASA Advanced Spaceborne Thermal
251 Emission and Reflection Radiometer (ASTER) digital elevation model (30 m resolution)
252 accessed using the EarthData Portal. Watershed delineation was conducted using ArcGIS Pro 2.9
253 functions. Flow directions were generated using the D8 method in which flow is routed to one of
254 8 neighboring cells selected by calculating downhill steepness (Qin et al., 2007). Next, flow
255 accumulation was calculated based on the flow direction raster with zones of high accumulation

256 representing stream channels. Water chemistry sampling points were used as pour points and
257 were snapped to nearest zones of flow accumulation. Contributing areas for each sampling point
258 were then delineated using the watershed tool. Next, each watershed factor dataset (lithology,
259 LULC, and effective precipitation) was clipped to reflect the upstream contributing area for each
260 of the sampling sites. A variety of spatial datasets were employed to examine factors influencing
261 stream chemistry at each point within the watershed. Lithologic data is from the USGS via the
262 Texas Water Development Board (TWDB) spatial data services portal. Rock types were merged
263 to convey major classes (i.e., carbonates, sandstone, mudstone, etc.). Comprehensive land use
264 data are from the National Land Cover Database (NLCD) for the most recent year released
265 (2016). Annual evapotranspiration (2000-2013) at 800 m resolution is estimated using an
266 empirical regression equation for long-term water balance ET data at 679 gaged watersheds as a
267 function of land cover, precipitation, and daily temperature (Reitz et al., 2017). Both
268 evapotranspiration and precipitation data (1971-2000) as annual average values (Crawford et al.,
269 2006) were included to represent climate variations across the watershed. Lithology and land use
270 classes were expressed as percent cover for each sub-basin. An areal weighted average of
271 effective precipitation was calculated for the upstream contributing area of each sampling point.
272

273 2.3 Data Analyses

274 2.3.1 Piper diagrams and geochemical modeling of the chemical composition of the Colorado 275 River

276 Piper diagrams were used to compare spatial trends in sample chemical composition to
277 underlying lithology, land use, and climate metrics. Piper diagrams show the distribution of the
278 relative proportion of major cations and anions in two trilinear diagrams, and one quadrilinear
279 diagram which shows overall chemical composition (Piper, 1944).

280 To further understand the potential lithologic controls on stream solute concentrations, R
281 package phreeqc (which interfaces the PHREEQC modeling code of Parkhurst & Appelo, 2013)
282 was used to determine saturation indices (SI) of possible expected minerals comprised of major
283 ions (gypsum, halite, aragonite, and chalcedony) for each charge balanced sample. pH
284 measurements were also included and temperature was assumed to be constant at 25°C as actual
285 stream temperature measurements were limited to 73% of observations (see S1). A sample was
286 considered to be undersaturated when SI is < 0 (dissolution is likely), saturated when SI = 0
287 (showing that the system is at equilibrium), and supersaturated when SI > 0 (mineral
288 precipitation is likely) (Drever, 1982). This analysis gives an idea of the degree of mineral
289 dissolution contributing to stream chemical composition across lithologic and climate gradients.
290 It also illustrates mineral presence not captured at the scale of USGS lithologic datasets, such as
291 gypsum and halite. Further analysis using a larger number of measured solutes could increase
292 our understanding of silicate dissolution in particular, as aluminosilicate minerals are likely the
293 main contributors of silica rather than chalcedony. However, this dataset includes very limited
294 measurements of Fe and Al to calculate the SI values of aluminosilicate minerals.

295 2.3.2 Concentration-discharge analysis

296 Concentration-discharge (C-Q) relationships were employed to examine spatial patterns
297 in the behavior of individual solutes. Concentration discharge relationships can be related to
298 watershed characteristics and commonly follow power-law relationships, as shown in equation 1
299 (Abbott et al., 2018; Baronas et al., 2017; Bouchez et al., 2017; Chorover et al., 2017; Godsey et
300 al., 2009; Herndon et al., 2015; Moatar et al., 2017; Musolff et al., 2015; Sullivan et al., 2019):

$$C = aQ^b \quad (1)$$

301 where C is the concentration of a given constituent, Q is the discharge, and a and b are
302 model parameters. When the measurements are log-transformed they can therefore be described
303 by a linear model where a slope (b) of 0 indicates chemostatic behavior, a positive slope
304 indicates addition behavior, and a slope of -1 indicates simple dilution behavior. Dilution occurs
305 when the concentration of the stream decreases with increased flow, which occurs at sites where
306 the baseflow of a river carries higher concentrations of each solute. Chemostatic behavior occurs
307 where concentrations are invariant across a large range of flow values. The addition behavior
308 occurs when concentrations increase systematically with increased flow. Concentration
309 measurements from charge balanced samples were matched to flow measurements taken at the
310 same site and date and converted to mmol/L for all solutes. Flow measurements were converted
311 to L/s and the logarithm (base 10) of both concentration and discharge was calculated. The
312 measurements were grouped by site and solute. Of 108 sites with charge balanced chemical
313 measurements, 88 included flow measurements and 51 had at least 5 observations. Only sites
314 with at least 5 measurements were included in this analysis. The mean number of observations
315 per site was 73 and the maximum was 442. A linear model was fit for each site and solute using
316 the “stats” package in R with the slope describing the concentration-discharge (C-Q) behavior (R
317 Core Team, 2021).

318

319 2.3.3 Principal Components Analysis (PCA) of compositional stream chemical data

320 Multivariate techniques were implemented to expand on observations of individual
321 solutes and to further refine sample groupings and relative influence of watershed factors.
322 Specifically, PCA was used to group samples that were biogeochemically similar and to examine
323 sample distribution relative to trends in lithology, LULC, and climate. To accomplish this goal,
324 data from charge balanced samples were analyzed as compositional measurements.
325 Compositional measurements represent the contribution of each solute to the overall
326 concentration of the sample in mmol/L. Compositional measurements show change in solute
327 concentrations relative to each other rather than outside factors, including dilution (Filzmoser et
328 al., 2009; Filzmoser & Hron, 2008; Shelton et al., 2018). This means that further transformations
329 must be applied to convert the variables into an unconstrained space. PCA weights eigenvectors
330 by the square root of the corresponding eigenvalue to transform a large number of variables into
331 orthogonal components that represent as much variation in the dataset as possible
332 (Christophersen & Hooper, 1992; Jolliffe, 2002; Burke, 1997; Walter et al., 2019). In our PCA,

333 matrix Z was comprised of all compositional data Loadings of each variable on the retained PCs
334 was tested for significance using the student's t-test, and the original data was reprojected using
335 the PCs. The PCs were converted from isometric to centered log ratio format to improve
336 interpretability. The latter log-ratio provides a one-to-one correspondence between input
337 variables and log-ratios but produces a singular covariance matrix, which limits its use in PCA.
338 PCs explaining up to 80% of cumulative variance in the dataset were considered significant for
339 interpretation.

340

341 2.4 Machine learning for determination of watershed factor influence on stream chemistry

342 Multivariate models including spatial data were also employed to quantify the behavior
343 of individual solutes in relation to changing watershed factors at a sub-watershed scale. This
344 process allows for improved understanding of differences in behavior across solute groups in
345 response to variation in lithologic, LULC, and climate distribution. A random forest approach
346 was used to predict the proportion of each watershed factor using sample chemical composition
347 as an input, and to quantify the strength of the relationship between these variables. Stream
348 chemical data was chosen as an input variable rather than a predictive output variable in this
349 investigation because of the greater variety in measurement values as compared to the
350 summarized watershed factor data (multiple water samples were collected from each site).

351 Random forests are a supervised machine learning algorithm consisting of ensembles of
352 decision trees generated using bootstrapped subsamples of the dataset. Random forests are a
353 flexible multivariate technique that is appropriate for analyzing multiple predictor variables and
354 nonlinear relationships (Addor et al., 2018). R package randomForest (R Core Team, 2021) was
355 used for algorithm generation on 80% of the data and tested using the remaining 20% (Liaw &
356 Wiener, 2002). The strength of the relationship between each solute and watershed factor was
357 quantified using the percent increase in mean squared error (%IncMSE) if the predictor variable
358 (solute) was removed from consideration. A high %IncMSE value indicates a stronger
359 relationship between the two variables. This gives a relative metric for comparison of the
360 influence of each factor (lithology, LULC, and climate) on the behavior of solutes across the
361 watershed (Addor et al., 2018; Hammond et al., 2021; Konapala & Mishra, 2020; Opiel &
362 Schumann, 2020). Relationships between all variables were also represented using the
363 Spearman's rank correlation coefficient.

364 **3 Results**

365 **3.1 Piper diagrams show shift from upstream Na-Cl dominated to downstream Ca-HCO₃**
 366 **dominated samples.**

367 Surface water chemistry shifts from Na-Cl dominated to Ca-HCO₃ dominated from
 368 northwest to southeast across the watershed (Figure 2). From zone 2 to zone 6, relative

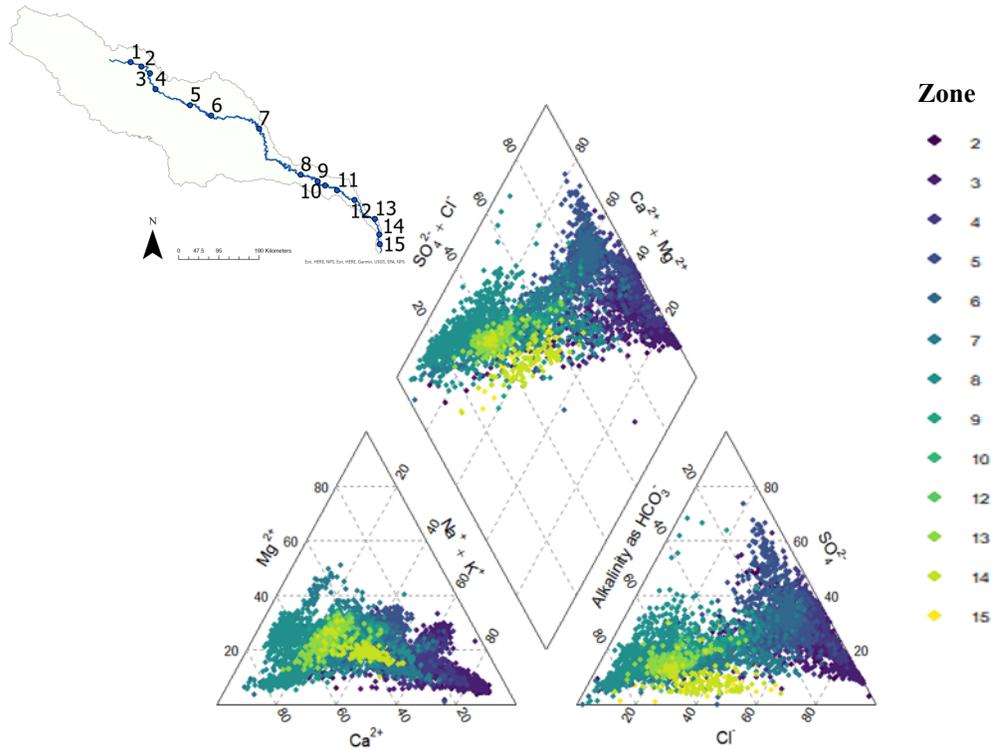


Figure 2 Piper diagrams represent the 14 zones along the Colorado river (inset map upper left). Darker blue are sites located upstream while lighter yellow indicate sites located near the coast.

369 concentration of Na⁺+K⁺ and Cl⁻ are very high. Additionally, the relative concentration of SO₄²⁻
 370 is highly variable. Relative proportions of Ca²⁺ and Mg²⁺, increase from zones 2 to 6 as Cl⁻
 371 decreases corresponding to an increase in HCO₃⁻. Measurements from zones 7-14 generally show
 372 the highest relative proportion of Ca²⁺ and HCO₃⁻. However, after zone 8 there is a slight
 373 decrease in Ca²⁺ and a corresponding increase in Na⁺+K⁺ and Cl⁻ (Figure 2).

374 **3.2 Geochemical modeling to refine conceptual model of landscape factor influences**

375

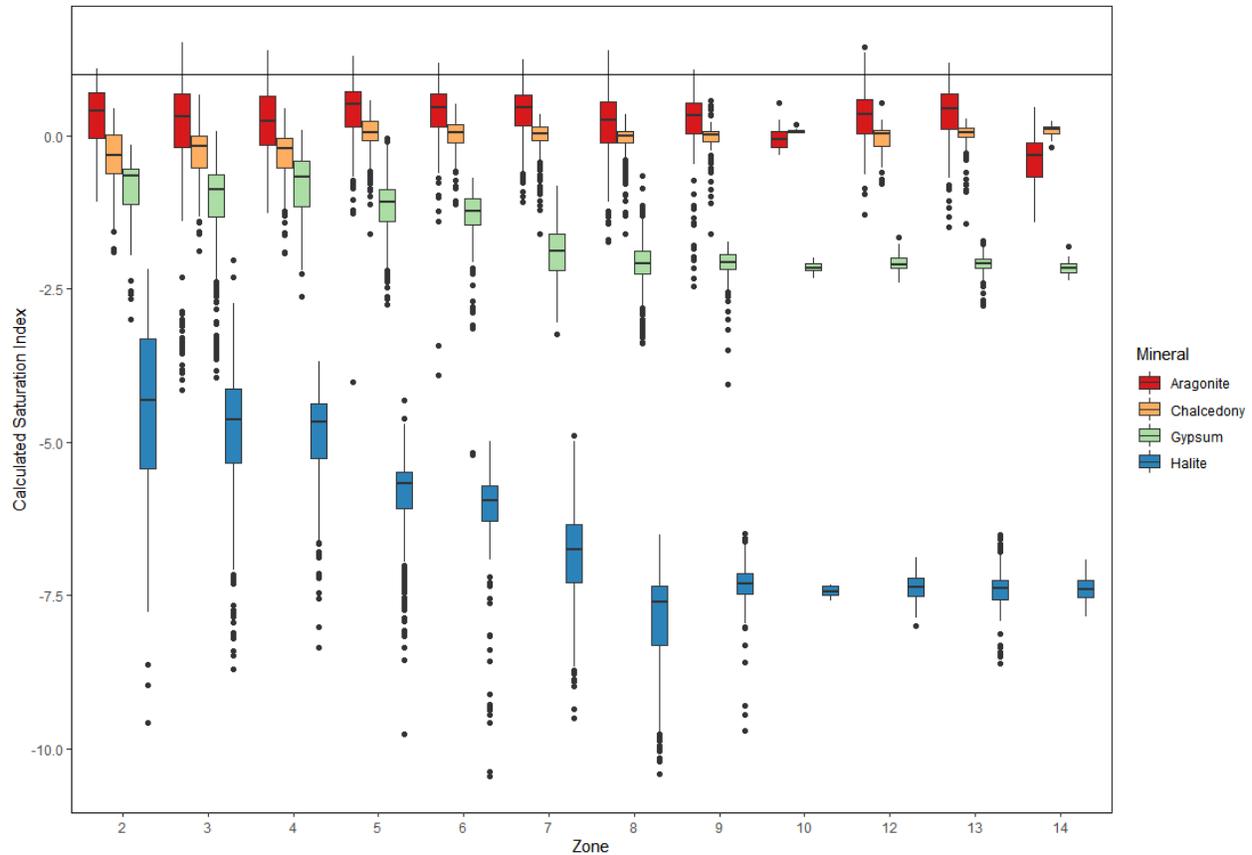


Figure 3 Box plots of the calculated saturation index across the 14 river zones for Aragonite (red), Chalcedony (orange), Gypsum (green), and Halite (blue). Box plots illustrating the median and inter quartile range as well as outlying points for calculated SI in each zone. A saturation index >0 indicates the potential for mineral precipitation, while a value <0 indicates the potential for mineral dissolution.

376 Saturation indices of aragonite (CaCO_3), chalcedony (SiO_2 used here to evaluate silicate minerals
 377 as SI of aluminosilicates was not available), gypsum (CaSO_4), and halite (NaCl) were calculated
 378 using PHREEQC software with major ion chemistry and pH measurements at a standard
 379 temperature of 25°C (Figure 3; see S1 for a subset of the data where field temperatures are
 380 available, overall trends remain consistent). Carbonate and silicate species are near saturated or
 381 slightly oversaturated while sulfate and evaporite species are dominantly undersaturated.
 382 Saturation indices for aragonite and chalcedony are relatively consistent from upstream to
 383 downstream zones. In contrast, gypsum and halite saturation indices are highest in the upstream
 384 portions of the river and decrease until zone 9 then remain relatively constant.

385 3.3 Concentration-Discharge relationships reveal overarching spatial trends

386 Slopes (parameter b) of the linear regression of $\log(\text{concentration})$ vs $\log(\text{discharge})$ were
 387 calculated for each site with at least 4 observations and are generally within error but still
 388 demonstrate trends in space (Figure 4). In summary, the Colorado River demonstrates 4 trends in
 389 slope of the C-Q relationships. The first trend includes HCO_3^- and Ca^{2+} which have slopes close

390 to 0 (chemostatic behavior) for zones 1-6 followed by increasing dilution behavior (slopes close
 391 to -0.5) from zones 7-14. Trend 2 represents the majority of the geogenic solutes Cl^- , Mg^{2+} , SO_4^{2-}
 392 and Na^+ have negative slopes (~ -0.5 ; dilution behavior) in the northwest zones, with slopes
 393 trending towards 0 in zones 6-7 and returning to dilution behavior for zones 8-14 (slopes < -0.5).
 394 Of these solutes Ca^{2+} is more chemostatic than the others in zone 2 with slopes between 0 and -
 395 0.25. Trend 3 includes K^+ and Si which are very close to chemostatic across all zones. Trend 4
 396 shows NO_3^- and P which are more highly variable than other solutes and have fewer

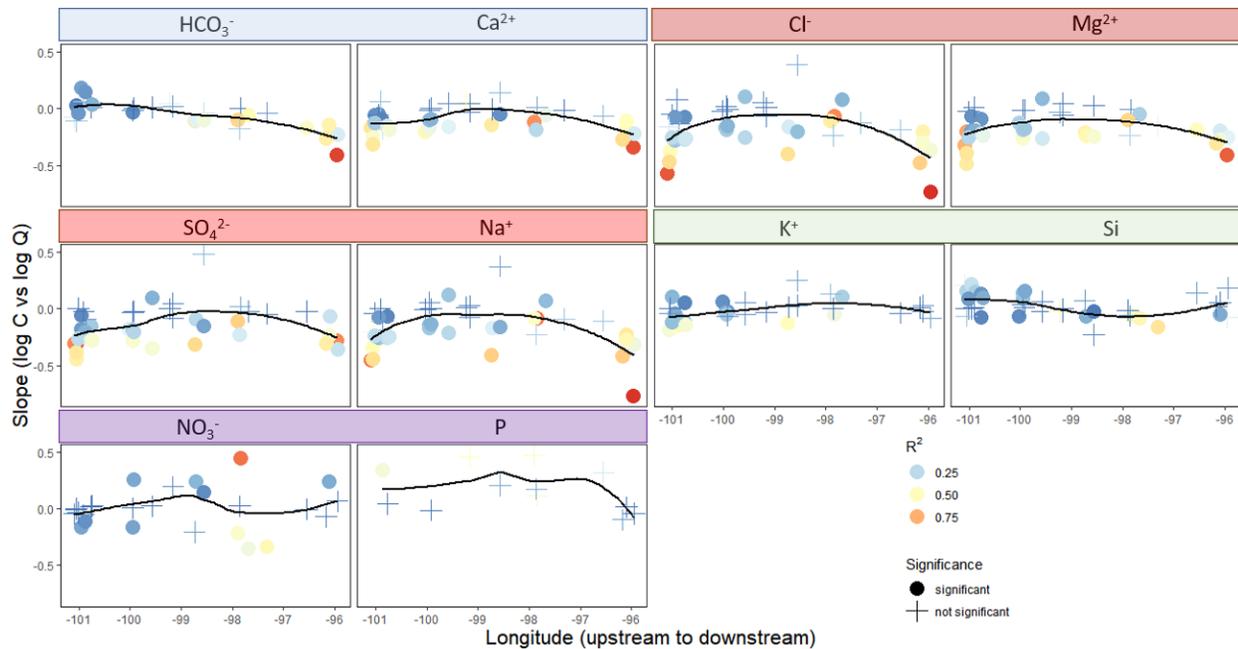


Figure 4 Slope of linear regression of $\log(C)$ versus $\log(Q)$ across the Colorado River Basin. Each point represents the slope calculated for measurements at a single site which are organized from upstream to downstream (left to right). Point shape indicates whether the model coefficient is significantly different from 0 at $\alpha=0.05$. Point color indicates the R^2 of the regression. Black lines indicate the smoothed conditional mean slope to illustrate overall trends for each solute. Title color indicates trends 1-4 with trend 1 in blue, trend 2 in red, trend 3 in green, and trend 4 in purple.

397 measurements included for analysis. NO_3^- shows some addition behavior in zones 4-6 with one
 398 outlying site in zone 8 showing significant addition behavior with a slope of almost +0.5.

399

400 3.4 Compositional PCA mirrors piper diagram trends from Na-Cl to Ca- HCO_3

401 PCA was conducted on all chemical measurements collected within the Colorado River
 402 Watershed for a total of 4,863 points. The first two principal components explain 88.7% of the
 403 variation in the dataset (PC1 = 76.4% and PC2 = 12.3 %) and the third explains an additional
 404 9.76%. The primary loadings of interest for the first principal component are Na^+ , Cl^- , and SO_4^{2-}

405 which are negative, and HCO_3^- , and Si which are positive (Table 1). The primary loadings of
 406 interest for the second principal component are K^+ (positive) and Mg^{2+} (negative).

Table 1 Loadings from compositional PCA for all measurements (n=4,863) in the Colorado River Watershed.

	PC1	PC2	PC3
Bicarbonate	0.448	-0.176	0.485
Calcium	0.146	-0.262	0.208
Chloride	-0.461	0.142	-0.168
Potassium	0.149	0.668	0.338
Magnesium	-0.015	-0.537	-0.116
Sodium	-0.442	0.222	0.039
Sulfate	-0.319	-0.247	-0.038
Silica	0.496	0.190	-0.750

407 Examination of the first 2 principal components reveals a pattern which orders the
 408 samples from upstream (left) to downstream (right and down) across PC1. Corresponding
 409 changes in the ions are also shown, indicating evolution from a Na-Cl dominated composition to
 410 a Ca- HCO_3^- dominated one (Figure 5). Higher relative abundance of Si on the centered log ratio
 411 biplots (Figure 5) generally indicates lower salinity composition due to the dilution of the other
 412 elements relative to Si, whose concentration is relatively constant (Engle et al., 2011). Sites 6-8
 413 are widely distributed and include many tributary measurements. Sites 9-15 include fewer
 414 measurements from each zone, are dominantly collected on the main stem, and are more closely
 415 clustered. The outlying cluster of 30 points on PC2 is from 2 USGS sites 08120700 and
 416 08121000 charge balanced samples collected in 1966 and 1967 and represents anomalously high
 417 values for K^+ . A total of 274 observations from site 08120700 collected from 1965-2002, and 298
 418 measurements collected from site 08121000 from 1963-2003 are included and measurements
 419 from other years are not outliers. Analysis of the spread of points in each zone based on season

420 and tributary position revealed no significant trends supporting the idea that this long-term
 421 dataset is more reflective of spatial variations than temporal trends (S1).

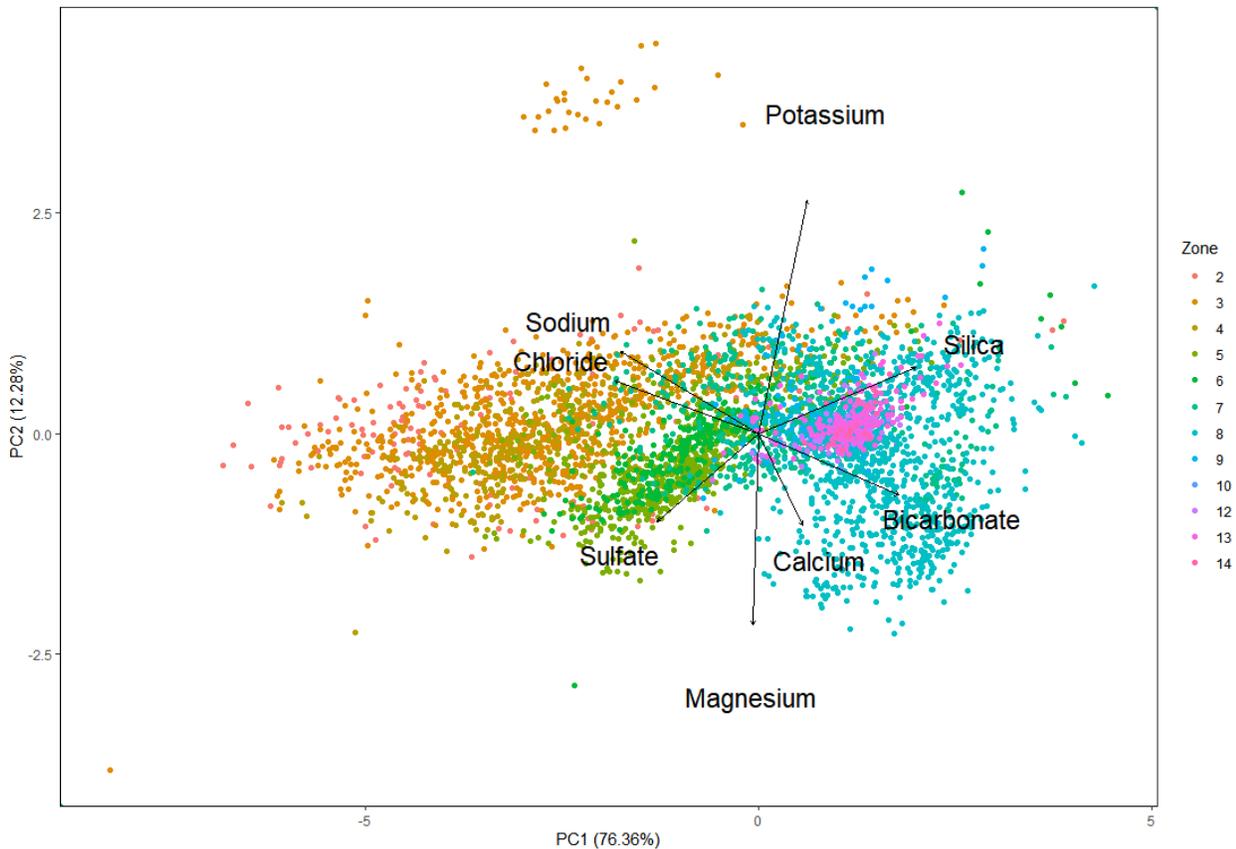


Figure 5 Colorado river water chemistry projected onto first two principal components shows spatial variations, points are colored by proximity to sampling sites 2-14.

422

423 3.5 Random forest analysis gives insight for LULC influences on stream water 424 composition

425 The random forest regression algorithms were used to assess relationships between
 426 solutes and watershed factors through comparison of percent increase in mean squared error
 427 (%IncMSE) for each solute/watershed factor combination (Figure 6). Here we focused on three
 428 groups of watershed factors: lithology, LULC, and climate. Of all LULC classes, cultivated
 429 crops were the strongest predictor across all solutes, while sedimentary deposits and evaporites
 430 had the strongest relationships to solute behavior in terms of lithology classes. Specifically, the
 431 results show a strong relationship between cultivated crops and K^+ , Mg^{2+} , and SO_4^{2-} and to a
 432 lesser degree Cl^- , Ca^{2+} , and HCO_3^- . Cultivated crops are also positively correlated with Na^+ and
 433 Cl^- but negatively correlated with Mg^{2+} , Ca^{2+} , and HCO_3^- . There are large areas covered by
 434 cultivated crops in the upper portions of the Colorado River Watershed and smaller stretches
 435 covered by pasture near zones 9-14. Additionally, there appears to be a strong relationship
 436 between open areas (including barren, grass, and shrubland) with Na^+ , Mg^{2+} , Ca^{2+} , and to a lesser

437 extent Cl^- and HCO_3^- . Open areas are also most prevalent in zones 2-7. Correlation coefficients
 438 are smaller than for agriculture but are positive for Mg^{2+} , Ca^{2+} , and HCO_3^- and negative for Na^+
 439 and Cl^- (opposite to agriculture). Forest and developed areas were both shown to have generally
 440 smaller %incMSE with most of the solutes.

441 Lithologic influences are strongest for sedimentary deposits, including the conglomerate,
 442 sandstone, carbonate, and evaporite classes. Sedimentary deposits and carbonates show the
 443 highest correlation coefficients. Sedimentary deposits have the highest %IncMSE for Ca^{2+} and
 444 Mg^{2+} with additional high %IncMSE for HCO_3^- , Na^+ , and Cl^- . Conglomerate has the highest
 445 %IncMSE for Na^+ and Mg^{2+} with additional large values for HCO_3^- and Cl^- .

446

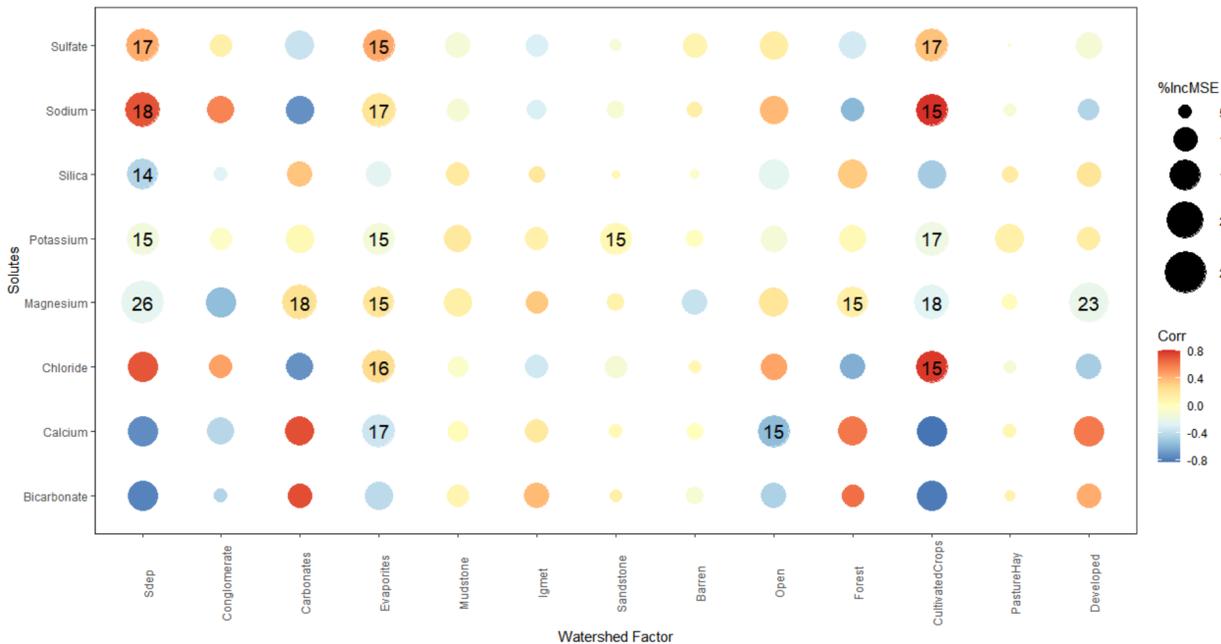


Figure 6 Plot of solutes versus lithology and LULC factors with point size showing %incMSE and color indicating the spearman's correlation coefficient, numeric labels show %incMSE value for top 80th percentile %incMSE scores

447 4 Discussion

448 Long term river chemistry data sets, especially in combination with spatial data, have been
 449 used at a range of scales to reveal underlying processes that control chemical behavior (Godsey
 450 et al., 2009; Park & Lee, 2020; Tiwari et al., 2017; Torres et al., 2017). Yet, the use of publicly
 451 available historical stream chemical measurements has been limited because of temporal and
 452 spatial irregularity. Here, we used PCA, C-Q relationships, geochemical modeling, and
 453 supervised machine learning (random forest) analyses to investigate the influence of watershed
 454 factors in the Texas Colorado River Basin. Our analyses reveal three main regions of behavior:
 455 the upper (zones 2-4), middle (5-8), and lower (9-15) Colorado River. From this analysis we find
 456 that abundant reactive mineralogy, saline shallow groundwater, agricultural and oilfield activities
 457 dominate chemical signatures in low-development, low-precipitation, and low-flow reaches of
 458 the upstream Colorado River watershed. Conversely, the downstream reaches are increasingly

459 controlled by climate factors rather than lithology or anthropogenic signatures. Below these
460 inferences are described in detail.

461 **4.1 Evaporites and brines exert a strong influence on river chemistry in the upper reaches** 462 **(zones 2-4) of the Colorado River while carbonates dominate middle reaches (zones 5-8)**

463 Several lines of evidence show that most of the spatiotemporal variation of solute
464 chemistry in the upper Colorado basin is controlled by the presence of easily weatherable
465 evaporite and carbonate minerals with shifts downstream from zone 2 to 8 in water chemistry
466 related to changes in this lithologic distribution. First, the spread and separation of measurements
467 across piper diagrams (from Na-Cl and Ca-SO₄ to Ca-HCO₃ water types; Figure 2) and PCA
468 (loadings of -6 to 0 across PC1 where Na⁺, Cl⁻, and HCO₃⁻ loaded significantly; Figure 5, Table
469 1) illustrates the control of evaporites (gypsum and halite) and brines across this upstream
470 portion of the Colorado River. Spatially, the proportion of evaporites is high in zones 2-5 as
471 compared to downstream zones while carbonates are ubiquitous across zones 2-8 (Figure 1a),
472 which contributes to the shift from Na-Cl and Ca-SO₄ to Ca-HCO₃ water types shown by piper
473 diagrams and PCA. Geochemical modeling of the saturation state of the stream water also
474 suggests water is near equilibrium with carbonate minerals throughout the upper extent of the
475 Colorado but is consistently undersaturated with respect to evaporite minerals. SI values for
476 evaporite minerals are highest in zones 2-3 and consistently decrease until zone 7 (Figure 3). One
477 interpretation of the decline in SI values for halite and gypsum is that evaporite mineral
478 dissolution acts as a solute source while these minerals consistently decline in proportion moving
479 downstream.

480 This strong lithologic control on stream water chemistry is echoed by other analyses of
481 historic water chemistry in the upper portions of the Colorado River (Scanlon et al., 2005; Slade
482 & Buszka, 1994; Slade et al., 2002), which show that a combination of anthropogenic and
483 natural factors leads to increased salinity in the upstream reaches of the Colorado River.
484 Particularly, abundance of near-surface evaporites and upwelling of brines in this region have
485 been shown to influence shallow groundwater composition and discharge which occurs laterally
486 and through salt-springs to the Colorado River. Analysis of water types using piper diagrams,
487 major ion ratios, and stable isotope analysis of surface and shallow groundwater samples all
488 support these conclusions (Dutton et al., 1989; Leifeste & Lansford, 1968; Reed, 1961). Gibbs
489 diagrams show extremely high TDS, Na⁺, and Cl⁻, particularly in the upstream reaches of the
490 Colorado River (zones 2-5), which suggest evaporation dominated systems (including

491 agricultural influences), brines, and halite/gypsum dissolution, but do not allow for distinction
 492 between these mechanisms (Figure 7).

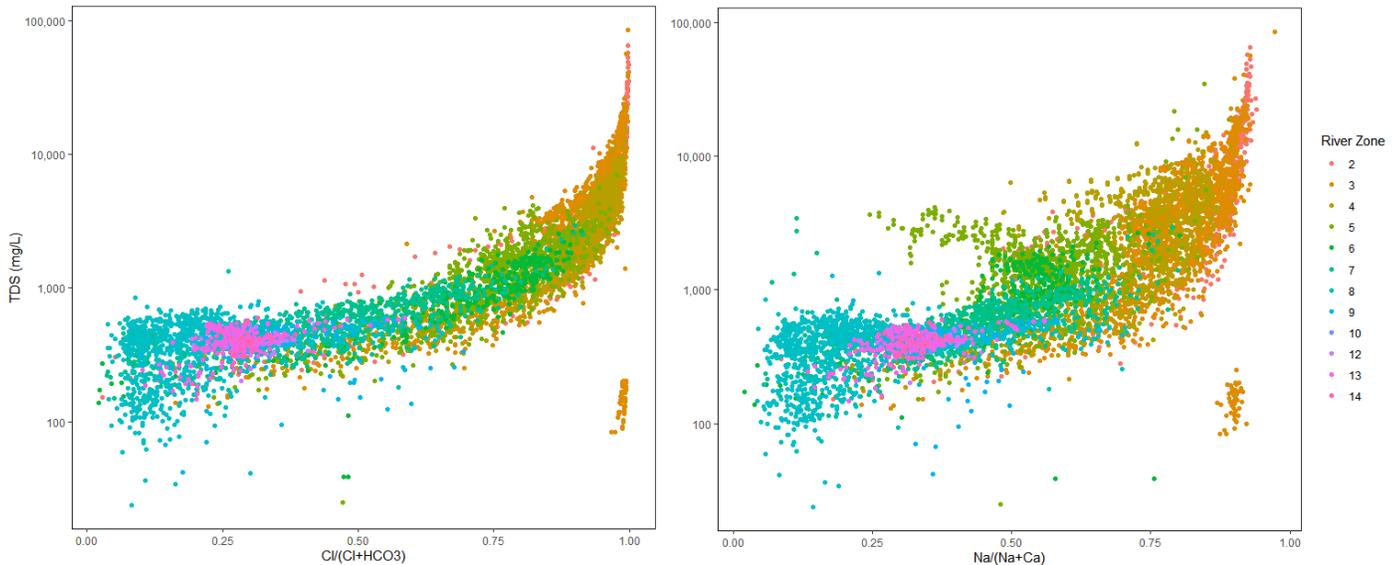


Figure 7 Gibb's diagrams showing the distribution of calculated total dissolved solids compared to Cl, Na, Ca, and HCO₃ concentrations across the Colorado River. Evaporation dominated areas are characterized by high TDS, Cl, and Na. Rock weathering dominated areas are characterized by moderate TDS (100-1,000 mg/L) and low Cl and Na.

493 Along with lithologic controls in this area, a combination of potential leakage from oil
 494 and gas development sites, naturally upwelling brines, and enrichment of solutes in shallow
 495 groundwater contribute to high overall concentrations of Na⁺, Cl⁻, and SO₄²⁻, dilution behavior of
 496 evaporite related solutes, and undersaturated conditions with respect to halite and gypsum
 497 (Dutton et al., 1989; Leifeste & Lansford, 1968; Paine et al., 1999; Richter et al., 1991; Slade &
 498 Buszka, 1994). Particularly, analysis of shallow groundwater and river water major ion and
 499 isotopic chemistry as compared to brines in local formations has shown significant similarity in
 500 composition between surface waters and brines, with other areas showing more direct influence
 501 from evaporite dissolution (Richter & Kreitler, 1986).

502 Concentration discharge behavior analyses also provide evidence of a strong lithologic
 503 control on solute behavior but suggest important differences in the spatial heterogeneity of the
 504 lithology. Specifically, dilution behavior occurs in zones 2-4 for in Na⁺, Cl⁻, and SO₄²⁻ (slopes of
 505 -0.5 to -0.1), while HCO₃⁻ and Ca²⁺ exhibit chemostatic behavior across these zones (Figure 4).
 506 This difference in C-Q behaviors across the upper Colorado River suggests that evaporites in
 507 zones 2-4 are distributed more heterogeneously in space, limiting solute sources and leading to
 508 the observed dilution behavior while the abundance of carbonates supports chemostatic behavior.
 509 Previous studies have shown that heterogeneity in lithologic distribution in combination with
 510 hydrologic factors (activation of different flow pathways and water sources during precipitation)
 511 can lead to non-chemostatic behavior due to mixing of different source waters (Baronas et al.,
 512 2017; Torres et al., 2017).

513 Hydrogeologic conditions in the semi-arid upper Colorado basin are conducive to
514 dominant lithologic influences on stream chemistry. Low precipitation and high potential
515 evapotranspiration can reduce pore flushing and lead to a greater degree of water-rock
516 interaction and elevated concentrations of geogenic solutes in subsurface water (Brantley &
517 Lebedeva, 2020; Stewart et al., 2022; White & Blum, 1997). Additionally, this region's shallow
518 groundwater bodies are subject to elevated solute concentrations as a result of evapotranspiration
519 (Dutton et al., 1989; Paine et al., 1999; Richter et al., 1991). Finally, the depth to reactive
520 minerals may be shallower in areas where effective precipitation is lower (Ameli et al., 2017;
521 Kim et al., 2017).

522 **4.2 Lower reaches (zones 9-14) of the Colorado River are more dominated by** 523 **climate and sedimentary deposits**

524 As effective precipitation increases downstream along the Colorado River watershed,
525 studies suggest the increase in available water will yield faster transit times (shorter residence
526 time) that support more dilution behavior (Brantley et al., 2017; Chorover et al., 2017; Maher,
527 2011). All geogenic solutes (such as HCO_3^- , Ca^{2+} , Cl^- , Na^+ , SO_4^{2-} , and Mg^{2+}) included here
528 reflect this climatic shift with increasing dilution behavior (decreasing slope b values towards -1)
529 beginning in zone 8 and continuing until the river outlet (zone 15). In the upper Colorado River
530 watershed HCO_3^- was highly chemostatic until zones 7 and 8 where the slopes declined showing
531 increased dilution. Evaporite and brine related solutes (such as Cl^- , Na^+ , and SO_4^{2-}) show the
532 same trends toward dilution in these zones. In zone 8 the river flow increases markedly, the local
533 watershed receives more precipitation, and the Colorado river passes the last of the major
534 reservoirs. Here, long-term weathering resulting from greater precipitation may have led to
535 deepening of reactive mineral fronts (Brantley et al., 2017; Brantley & Lebedeva, 2020). This,
536 along with increased contributions of meteoric water, directly via precipitation, and indirectly
537 through shallow groundwater contributions to streamflow and decreased water residence time in
538 the shallow subsurface lead to slightly increased dilution behavior. This is consistent with
539 previously developed C-Q and reactive transport conceptual models which show strong
540 dependence on thermodynamic and hydrologic controls by stream chemical behavior (Brantley
541 & Lebedeva, 2020; Maher, 2011; Wlostowski et al., 2021).

542 As the watershed moves through areas of increased effective precipitation, chemical
543 weathering occurs at a faster rate leading to clustering of points on PC1 and in the Piper diagram,
544 changes in calculated SI for evaporites, and increased chemostatic behavior of Na^+ , Cl^- , and SO_4^{2-}
545 (Figures 2-5). Variations in C-Q relationships and PCA results are consistent with previous
546 findings which suggest that thermodynamic (reaction rates of underlying lithology) and
547 hydrologic (flow and climatic) controls exert strong influence on stream chemical composition
548 and behavior except where this influence is obscured by contributions of different source waters,
549 climate factors, and anthropogenic influences which all increase moving downstream in this
550 watershed (Baronas et al., 2017; Bouchez et al., 2017; Knapp et al., 2020; Maher, 2011; Torres et
551 al., 2017). These findings show that, for evaporite and carbonate related solutes in this

552 watershed, lithology is the dominant influencing factor until increased contributions from
553 precipitation and urban development (zone 8) shift river chemistry away from lithologic signals.

554 **4.3 Possible anthropogenic inputs in the upper Colorado watershed**

555 While lithology and climate dominantly control river chemical composition across the
556 Colorado River, anthropogenic factors, particularly agricultural activities and oil and gas
557 development, are also influential in both the upstream and central reaches of the watershed. Oil
558 and gas development has been posited to increase surface water salinity via a variety of
559 mechanisms. Orphaned and improperly abandoned oil and gas wells and poorly managed
560 disposal of produced waters allows for inputs of brines in the shallow subsurface (Dutton et al.,
561 1989). Additionally, salts accumulate below historic brine disposal pits and are leached into
562 shallow groundwater or laterally moved into surface waters (Dutton et al., 1989; Richter et al.,
563 1991). While it is possible that this occurs on the Colorado River it is difficult to distinguish
564 between the impacts of oil and gas activities (either historical or recent) versus natural
565 contributions like evaporite dissolution, concentrated shallow groundwater, natural upwelling of
566 deep brines, or other human activities including agriculture. Gibb's diagrams and analysis of
567 major ion ratios give some support to brine influence (Figure 7) but further analyses (such as U,
568 Sr, and B isotope tracers) need to be conducted using a greater number of chemical metrics to
569 determine the degree and extent of oil and gas versus other salinity sources across this region
570 (Engle et al., 2011; Osborn et al., 2011).

571 Agricultural land (cultivated crops) is most abundant in zones 2-6 of the Colorado River
572 basin and may contribute to increased salinity due to agricultural amendments and increased
573 groundwater use. Groundwater use for irrigation is particularly common in this region. The
574 strong relationship between Na^+ and agricultural land is shown by the random forest analysis
575 (Figure 6) where there is a large %IncMSE for Na^+ , K^+ , Cl^- , and Mg^{2+} all components of
576 subsurface brines and agricultural amendments. A potential mechanism for this is the ET-based
577 accumulation and following dissolution and percolation of naturally occurring salts accumulated
578 under agricultural irrigation areas (Kondash et al., 2020; Yurtseven et al., 2018). Further,
579 agricultural irrigation may lead to increased soil salinity over time due to the use of saline
580 groundwater.

581 Urban areas, including the developed region around the city of Austin, impact both flow
582 and water quality (Aitkenhead-Peterson et al., 2011). Urban regulation of discharge and
583 influence of water chemistry in combination with increased precipitation obscures the underlying
584 lithologic influences and cause clustering of measurements on principal component 1 (Chen et
585 al., 2014; Kaushal et al., 2013; Raymond et al., 2008). Particularly in large cities (i.e., Austin)
586 significant municipal water needs require importing water from multiple sources. Discharges of
587 this water for various uses (irrigation, treated wastewater, road runoff) may be of entirely
588 different composition than the naturally occurring water sources (Aitkenhead-Peterson et al.,
589 2011). However, these impacts are not always well-captured through analysis of major ion
590 chemistry. Analyses including stable isotope ratios (H, O, and B) and/or nutrients that are

591 common indicators of anthropogenic activity (e.g., NO_3^- , NH_4^+ , PO_4^{3-}) could better capture the
592 influence of urban development on stream chemical composition (Lyon et al., 2008; Raymond et
593 al., 2008). While reservoirs are likely impacting the system, they do not appear to be
594 significantly impacting the major ion composition. Reservoir storage has also been shown to be a
595 weak predictor of flow metrics across varying environments (Hammond et al., 2021).
596 Differing LULC types including oil and gas acreage, urban areas, and reservoirs alter natural
597 signals along the Colorado River. The influence of brines associated with oil and gas production
598 activities was discussed earlier (see 4.1 and 4.3) as their chemical signature with the available
599 data is hard to distinguish from evaporite dissolution or naturally upwelling brines (Leifeste &
600 Lansford, 1968). Land use and intensive urban development are the next most influential factors,
601 as illustrated by the shift in data distribution occurring after zone 8 in the PCA biplot,
602 contributing to the spread of measurements occurring in the downstream zones of the Colorado
603 River. From zone 8 (Austin area) to the river outlet the position of measurements on the PCA
604 biplot displays a scattered rather than linear trend. This again suggests that intensive
605 development, while it does not display a dominant signature in major ion chemical composition
606 of surface waters, tends to add variability in other hydrologic signals.

607 Along the central reaches of the Colorado River, influences from tributaries, reservoirs,
608 and differing LULC become increasingly dominant, shifting C-Q behavior away from dilution
609 and towards chemostatic and leading to data clustering on the PCA biplot. Some elements may
610 even display addition behavior such as P and NO_3^- that can be linked to LULC and human
611 impacts. Major tributaries dilute elements transported from upstream and contribute deep
612 groundwater derived carbonates drawn from underlying lithology on both the main stem and
613 tributaries of the Colorado River in connection with the nearby Llano uplift (Bruun et al., 2016;
614 Leifeste & Lansford, 1968; Slade et al., 2002). This is shown through increases in relative
615 concentration of HCO_3^- and chemostatic behavior of C-Q slopes for all solutes in zones 5-8. In
616 the Amazon basin, contributions from relatively homogeneous water sources from major
617 tributaries were also shown to produce chemostatic behavior of major ions (Baronas et al., 2017;
618 Torres et al., 2017). Similar to the behavior of HCO_3^- and Ca^{2+} in the upstream zones of the
619 Colorado River, the influence of geologic features (including the Llano uplift) near major
620 tributaries serves to increase the availability of reactive carbonate minerals through connections
621 to deep groundwater and increases chemostatic behavior and homogeneity of chemical
622 composition.

623

624 **4.4 Broader applications and limitations due to reliance on major ion concentrations**

625 The analyses presented here rely on publicly available major ion chemical data collected
626 from over 100 sites in the Colorado River watershed, collected at irregular spatial and temporal
627 intervals. This valuable resource allows for analysis of many influences of stream water quality
628 across lithology, LULC, and climate gradients but does not provide sufficient information to
629 distinguish among some of these influences with great detail. With additional major, trace
630 element, and isotope data (e.g., $\delta^{18}\text{O}$, δD , $\delta^{11}\text{B}$, $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{234}\text{U}/^{238}\text{U}$), other relationships

631 could be examined and more details could be distinguished concerning existing processes and
632 influences. Salinity in the Colorado River watershed encompasses much of the overall chemical
633 composition and is derived from multiple sources including deep brines, shallow groundwater,
634 evaporite dissolution, and agricultural influences (Leifeste & Lansford, 1968; Paine et al., 1999;
635 Richter et al., 1991). Further measurements of stable isotopes, age-dating methods, and bromide
636 could be used in zones 2-4 of the Colorado river to better define the degree of each of these
637 influences and their variation in space.

638

639 The simplified lithology classes analyzed here provided valuable information about the
640 degree of influence of various reactive minerals on stream chemistry. Strong relationships were
641 evident in the upper Colorado River watershed between evaporites (halite, gypsum), carbonates
642 (aragonite, dolomite), and stream water chemistry. Further studies could leverage the information
643 and analytical techniques presented here and include additional measurements of trace metals
644 (e.g., Fe, Al, etc.) with more detailed lithology maps to better understand the distribution of
645 different lithologies, mineralogic influences on stream chemistry, and sources of solutes across
646 the watershed. These analyses can also focus on the potential structural controls of stream
647 chemistry from possible groundwater inputs from the Balcones fault zone, which crosses the
648 Colorado River at zone 8. Finally, the application of random forest algorithms provided useful
649 insight to watershed factor influences on stream chemical composition and behavior in this
650 catchment. Future work could include larger-scale public datasets in combination with a variety
651 of non-linear data analysis techniques to refine and generalize understanding of influences of
652 stream water chemistry across diverse environments.

653 **5 Conclusions**

654 Over 5,280 publicly available water quality observations were paired with daily discharge
655 measurements and high resolution geospatial (e.g., lithology and land cover) data from the
656 Colorado River (103,000 km²) over a 60 year period (1958-2018) to determine the degree of
657 influence of key watershed factors on stream chemical behavior. Comparison of trends in
658 chemical composition using Piper and Gibb's diagrams and compositional PCA set a baseline for
659 interpretation of changing contributions of evaporite and carbonates vs. silicate minerals across
660 the watershed. Further analyses including C-Q relationships and their changes in space showed
661 that not all solutes behaved similarly at the watershed scale and C-Q behavior changes could be
662 attributed to watershed processes such as shallow groundwater contributions, long term
663 precipitation influences, and the distribution of reactive mineralogy. Finally, random forest
664 analysis provided a metric for interpretation of the relative influence of watershed factors on
665 each solute and showed strong relationships between sedimentary units, contribution of
666 carbonates, and Na⁺, Cl⁻, Ca²⁺, Mg²⁺, and SO₄²⁻. Further, human factors such as LULC types
667 including cultivated crops, grass, and shrublands were shown to have strong relationships with

668 various solutes across the watershed while developed areas and forestland did not appear to have
669 a similarly strong relationship in this watershed.

670 Overall, this research highlighted the importance of reactive mineral abundance and
671 dissolution, the effect of climate (precipitation), and the impacts of multiple sources of salinity in
672 the Colorado River watershed. Broader implications are that regional, historical major ion
673 chemical datasets at the watershed scale can be leveraged to improve understanding of lithologic
674 and climate impacts on stream water chemistry at relatively large spatial and temporal scales.
675 Further analyses should be employed to better-determine the processes occurring at key locations
676 with shorter time scales (“hotspots and hot moments”) using a broader and combined set of
677 chemical, climate, lithologic, and land use parameters. Additionally, these techniques could be
678 applied to examine processes controlling water chemical composition and behavior at larger
679 scales leveraging national databases.

680 **Acknowledgments, Samples, and Data**

681 We acknowledge funding support from the National Science Foundation (EAR-1933261; EAR-
682 1933259). Any opinions, findings, and conclusions or recommendations expressed in this article
683 are those of the author(s) and do not necessarily reflect the views of the National Science
684 Foundation. The authors would like to thank Lourdes Moreu for her support in initial site
685 selection and project setup. We are grateful for the thoughtful contributions Sean Fleming, Nuria
686 Andreu Garcia, and Jennifer Herrera to project design and analysis techniques.

687 Stream discharge data can be accessed from the USGS National Water Information System
688 (<https://waterdata.usgs.gov/nwis/rt>), stream water quality data can be accessed from the USGS
689 Water Quality Portal (<https://www.waterqualitydata.us/>), spatial data can be accessed via the
690 TCEQ Spatial Database (<https://gis-tceq.opendata.arcgis.com/>) and Multi-resolution land
691 characteristics consortium (<https://www.mrlc.gov/>).

692 **References**

- 693 Abbott, B. W., Gruau, G., Zarnetske, J. P., Moatar, F., Barbe, L., Thomas, Z., Fovet, O., Kolbe,
694 T., Gu, S., Pierson-Wickmann, A. C., Davy, P., & Pinay, G. (2018). Unexpected spatial
695 stability of water chemistry in headwater stream networks. *Ecology Letters*, *21*(2), 296–308.
696 <https://doi.org/10.1111/ele.12897>
- 697 Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A
698 Ranking of Hydrological Signatures Based on Their Predictability in Space. *Water*
699 *Resources Research*, *i*, 8792–8812. <https://doi.org/10.1029/2018WR022606>
- 700 Aitkenhead-Peterson, J. A., Nahar, N., Harclerode, C. L., & Stanley, N. C. (2011). Effect of
701 urbanization on surface water chemistry in south-central Texas. *Urban Ecosystems*, *14*(2),
702 195–210. <https://doi.org/10.1007/s11252-010-0147-2>
- 703 Ameli, A., Beven, K., Erlandsson, M., Creed, I. F., McDonnell, J. J., & Bishop, K. (2017).
704 Primary weathering rates, water transit times, and concentration-discharge relations: A
705 theoretical analysis for the critical zone. *Water Resources Research*.
706 <https://doi.org/10.1111/j.1752-1688.1969.tb04897.x>

- 707 Baronas, J. J., Torres, M. A., Clark, K. E., & West, A. J. (2017). Mixing as a driver of temporal
708 variations in river hydrochemistry: 2. Major and trace element concentration dynamics in
709 the Andes-Amazon transition. *Water Resources Research*. [https://doi.org/10.1111/j.1752-](https://doi.org/10.1111/j.1752-1688.1969.tb04897.x)
710 [1688.1969.tb04897.x](https://doi.org/10.1111/j.1752-1688.1969.tb04897.x)
- 711 Bouchez, J., Moquet, J. S., Espinoza, J. C., Martinez, J. M., Guyot, J. L., Lagane, C., Filizola, N.,
712 Noriega, L., Hidalgo Sanchez, L., & Pombosa, R. (2017). River Mixing in the Amazon as a
713 Driver of Concentration-Discharge Relationships. *Water Resources Research*, *53*(11),
714 8660–8685. <https://doi.org/10.1002/2017WR020591>
- 715 Brantley, S. L., & Lebedeva, M. I. (2020). Relating land surface, water table, and weathering
716 fronts with a conceptual valve model for headwater catchments. *Hydrological Processes*.
- 717 Brantley, S. L., Lebedeva, M. I., Balashov, V. N., Singha, K., Sullivan, P. L., & Stinchcomb, G.
718 (2017). Toward a conceptual model relating chemical reaction fronts to water flow paths in
719 hills. *Geomorphology*, *277*, 100–117. <https://doi.org/10.1016/j.geomorph.2016.09.027>
- 720 Bruun, B., Jackson, K., Lake, P., & Walker, J. (2016). *Texas Aquifers Study: Groundwater*
721 *Quantity, Quality, Flow, and Contributions to Surface Water*.
- 722 Chen, J.-B., Gaillardet, J., Bouchez, J., Louvat, P., & Wang, Y.-N. (2014). Anthropophile
723 elements in river sediments: Overview from the Seine River, France. *Geochemistry,*
724 *Geophysics, Geosystems*, *4526–4546*. <https://doi.org/10.1002/2014GC005516>.Received
- 725 Chorover, J., Derry, L. A., & McDowell, W. H. (2017). Concentration-Discharge Relations in
726 the Critical Zone: Implications for Resolving Critical Zone Structure, Function, and
727 Evolution. *Water Resources Research*, *53*(11), 8654–8659.
728 <https://doi.org/10.1002/2017WR021111>
- 729 Christophersen, N., & Hooper, R. P. (1992). Multivariate analysis of stream water chemical data:
730 The use of principal components analysis for the end-member mixing problem. *Water*
731 *Resources Research*, *28*(1), 99–107. <https://doi.org/10.1029/91WR02518>
- 732 Clark, A. K., Morris, R. E., & Pedraza, D. E. (2020). Geologic framework and hydrostratigraphy
733 of the edwards and trinity aquifers within northern medina county, Texas. In *U.S.*
734 *Geological Survey Scientific Investigations* (Vol. 2020, Issue 3461).
735 <https://doi.org/10.3133/sim3461>
- 736 Crawford, C., Hamilton, P., & Hoos, A. (2006). *National Water-Quality Assessment Program -*
737 *Modifications to the status and trends network and assessments of streams and rivers*. US
738 Geological Survey.
739 https://water.usgs.gov/GIS/metadata/usgswrd/XML/nhd_ppt30yr.xml#stdorder
- 740 De Cicco, L. A., Hirsch, R. M., Lorenz, D., & Watkins, W. D. (2018). *dataRetrieval: R packages*
741 *for discovering and retrieving water data available from Federal hydrologic web services*
742 (2.7.6). U.S. Geological Survey.
- 743 Dieter, C. A., Maupin, M. A., Caldwell, R. R., Harris, M. A., Ivahnenko, T. I., Lovelace, J. K.,
744 Barber, N. L., & Linsey, K. S. (2018). Estimated Use of Water in the United States in 2015.
745 In *US Geological Survey Circular* (Issue 1441). <https://doi.org/10.3133/cir1441>
- 746 Drever, J. (1982). *The Geochemistry of Natural Waters*. Prentice-Hall.
- 747 Dupré, B., Dessert, C., Oliva, P., Goddérés, Y., Viers, J., François, L., Millot, R., & Gaillardet, J.

- 748 (2003). Rivers, chemical weathering and Earth's climate. *Comptes Rendus - Geoscience*,
749 335(16), 1141–1160. <https://doi.org/10.1016/j.crte.2003.09.015>
- 750 Dutton, A. R., Richter, B. C., & Kreitler, C. (1989). Brine Discharge and Salinization , Concho
751 River Watershed , West Texas. *Ground Water*. <https://doi.org/10.1111/j.1745->
752 6584.1989.tb00461.x
- 753 Engle, M. A., Bern, C. R., Healy, R. W., Sams, J. I., Zupancic, J. W., & Schroeder, K. T. (2011).
754 Tracking solutes and water from subsurface drip irrigation application of coalbed methane-
755 produced waters, Powder River Basin, Wyoming. *Environmental Geosciences*, 18(3), 169–
756 187. <https://doi.org/10.1306/eg.03031111004>
- 757 Evans, C., & Davies, T. D. (1998). Causes of concentration/discharge hysteresis and its potential
758 as a tool for analysis of episode hydrochemistry. *Water Resources Research*, 34(1), 129–
759 137. <https://doi.org/10.1029/97WR01881>
- 760 Filzmoser, P., & Hron, K. (2008). Outlier detection for compositional data using robust methods.
761 *Mathematical Geosciences*, 40(3), 233–248. <https://doi.org/10.1007/s11004-007-9141-5>
- 762 Filzmoser, P., Hron, K., & Reimann, C. (2009). Principal component analysis for compositional
763 data with outliers. *Environmetrics*, 20(6), 621–632. <https://doi.org/10.1002/env.966>
- 764 Gaillardet, J., Dupre, B., Louvat, P., & Allegre, C. J. (1999). Global silicate weathering and CO2
765 consumption rates deduced from the chemistry of large rivers. *Chemical Geology*, 159, 3–
766 30.
- 767 Godsey, S. E., Kirchner, J., & Clow, D. (2009). Concentration-discharge relationships reflect
768 chemostatic characteristics of US Catchments. *Hydrological Processes*, 32(18), 2829–2844.
769 <https://doi.org/10.1002/hyp.13235>
- 770 Güler, C., Thyne, G. D., McCray, J. E., & Turner, A. K. (2002). Evaluation of graphical and
771 multivariate statistical methods for classification of water chemistry data. *Hydrogeology*
772 *Journal*, 10(4), 455–474. <https://doi.org/10.1007/s10040-002-0196-6>
- 773 Hammond, J. C., Zimmer, M., Shanafield, M., Kaiser, K., Godsey, S. E., Mims, M. C., Zipper, S.
774 C., Burrows, R. M., Kampf, S. K., Dodds, W., Jones, C. N., Krabbenhoft, C. A., Boersma,
775 K. S., Datry, T., Olden, J. D., Allen, G. H., Price, A. N., Costigan, K., Hale, R., ... Allen, D.
776 C. (2021). Spatial Patterns and Drivers of Nonperennial Flow Regimes in the Contiguous
777 United States. *Geophysical Research Letters*, 48(2), 1–11.
778 <https://doi.org/10.1029/2020GL090794>
- 779 Harwell, G. R., McDowell, J., Gunn-Rosas, C., & Garrett, B. (2020). Precipitation, temperature,
780 groundwater-level elevation, streamflow, and potential flood storage trends within the
781 Brazos, Colorado, Big Cypress, Guadalupe, Neches, Sulphur, and Trinity River Basins in
782 Texas Through 2017. *Scientific Investigations Report*, April.
783 <http://pubs.er.usgs.gov/publication/sir20195137>
- 784 Herndon, E. M., Dere, A. L., Sullivan, P. L., Norris, D., Reynolds, B., & Brantley, S. L. (2015).
785 Landscape heterogeneity drives contrasting concentration-discharge relationships in shale
786 headwater catchments. *Hydrology and Earth System Sciences*, 19(8), 3333–3347.
787 <https://doi.org/10.5194/hess-19-3333-2015>
- 788 Jolliffe, I. T. (2002). Principal components analysis. In *Springer Series in Statistics*.

- 789 <https://doi.org/10.1016/B978-0-08-044894-7.01358-0>
- 790 Kaushal, S. S., Likens, G. E., Utz, R. M., Pace, M. L., Grese, M., & Yepsen, M. (2013).
791 Increased river alkalization in the eastern U.S. *Environmental Science and Technology*,
792 47(18), 10302–10311. <https://doi.org/10.1021/es401046s>
- 793 Kim, H., Dietrich, W. E., Thurnhoffer, B. M., Bishop, J. K. B., & Fung, I. Y. (2017). Controls on
794 solute concentration-discharge relationships revealed by simultaneous hydrochemistry
795 observations of hillslope runoff and stream flow: The importance of critical zone structure.
796 *Water Resources Research*, 1424–1443. <https://doi.org/10.1002/2016WR019722>. Received
- 797 Kirchner, J. W. (2003). A double paradox in catchment hydrology and geochemistry.
798 *Hydrological Processes*, 17(4), 871–874. <https://doi.org/10.1002/hyp.5108>
- 799 Knapp, J. L. A., Von Freyberg, J., Studer, B., Kiewiet, L., & Kirchner, J. W. (2020).
800 Concentration-discharge relationships vary among hydrological events, reflecting
801 differences in event characteristics. *Hydrology and Earth System Sciences*, 24(5), 2561–
802 2576. <https://doi.org/10.5194/hess-24-2561-2020>
- 803 Konapala, G., & Mishra, A. (2020). Quantifying Climate and Catchment Control on
804 Hydrological Drought in the Continental United States. *Water Resources Research*, 56(1),
805 1–25. <https://doi.org/10.1029/2018WR024620>
- 806 Kondash, A. J., Redmon, J. H., Lambertini, E., Feinstein, L., Weinthal, E., Cabrales, L., &
807 Vengosh, A. (2020). The impact of using low-saline oilfield produced water for irrigation
808 on water and soil quality in California. *Science of the Total Environment*, 733, 139392.
809 <https://doi.org/10.1016/j.scitotenv.2020.139392>
- 810 Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019).
811 Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine
812 Learning. *Water Resources Research*, 55(12), 11344–11354.
813 <https://doi.org/10.1029/2019WR026065>
- 814 Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019).
815 Towards learning universal, regional, and local hydrological behaviors via machine learning
816 applied to large-sample datasets. *Hydrology and Earth System Sciences*, 23(12), 5089–5110.
817 <https://doi.org/10.5194/hess-23-5089-2019>
- 818 Leifeste, D. K., & Lansford, M. W. (1968). *Reconnaissance of the chemical quality of surface*
819 *waters of the Colorado River Basin, Texas* (Issue March).
- 820 Liaw, A., & Wiener, M. (2002). *Classification and regression by random forest* (pp. 2(3), 18–
821 22).
- 822 Liu, F., Conklin, W. H., & Shaw, G. . (2017). Insights into hydrologic and hydrochemical
823 processes based on concentration-discharge and end-member mixing analyses in the mid-
824 Merced River Basin, Sierra Nevada, California. *Journal of the American Water Resources*
825 *Association*. <https://doi.org/10.1111/j.1752-1688.1969.tb04897.x>
- 826 Liu, X., Zhang, Y., Li, Z., Li, P., Xu, G., Cheng, Y., & Zhang, T. (2021). Response of water
827 quality to land use in hydrologic response unit and riparian buffer along the Dan River,
828 China. *Environmental Science and Pollution Research*, 28(22), 28251–28262.
829 <https://doi.org/10.1007/s11356-021-12636-z>

- 830 Lyon, S. W., Seibert, J., Lembo, A. J., Steenhuis, T. S., & Walter, M. T. (2008). Incorporating
831 landscape characteristics in a distance metric for interpolating between observations of
832 stream water chemistry. *Hydrology and Earth System Sciences*, *12*(5), 1229–1239.
833 <https://doi.org/10.5194/hess-12-1229-2008>
- 834 Maher, K. (2011). The role of fluid residence time and topographic scales in determining
835 chemical fluxes from landscapes. *Earth and Planetary Science Letters*, *312*(1–2), 48–58.
836 <https://doi.org/10.1016/j.epsl.2011.09.040>
- 837 Minaudo, C., Dupas, R., Gascuel-odoux, C., Roubeix, V., Danis, P., & Moatar, F. (2019).
838 Seasonal and event-based concentration-discharge relationships to identify catchment
839 controls on nutrient export regimes. *Advances in Water Resources*, *131*(July 2018), 103379.
840 <https://doi.org/10.1016/j.advwatres.2019.103379>
- 841 Moatar, F., Abbott, B. W., Minaudo, C., Curie, F., & Pinay, G. (2017). Elemental properties,
842 hydrology, and biology interact to shape concentration-discharge curves for carbon,
843 nutrients, sediment, and major ions. *Journal of the American Water Resources Association*.
844 <https://doi.org/10.1111/j.1752-1688.1969.tb04897.x>
- 845 Musolff, A., Schmidt, C., Selle, B., & Fleckenstein, J. H. (2015). Catchment controls on solute
846 export. *Advances in Water Resources*, *86*, 133–146.
847 <https://doi.org/10.1016/j.advwatres.2015.09.026>
- 848 Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C.,
849 & Gupta, H. V. (2021). What Role Does Hydrological Science Play in the Age of Machine
850 Learning? *Water Resources Research*, *57*(3). <https://doi.org/10.1029/2020WR028091>
- 851 Opel, H., & Schumann, A. H. (2020). Machine learning based identification of dominant
852 controls on runoff dynamics. *Hydrological Processes*, *34*(11), 2450–2465.
853 <https://doi.org/10.1002/hyp.13740>
- 854 Osborn, S. G., Vengosh, A., Warner, N. R., & Jackson, R. B. (2011). Methane contamination of
855 drinking water accompanying gas-well drilling and hydraulic fracturing. *Proceedings of the
856 National Academy of Sciences of the United States of America*, *108*(20), 8172–8176.
857 <https://doi.org/10.1073/pnas.1100682108>
- 858 Paine, J. G., Dutton, A. R., & Blum, M. U. (1999). *Using Airborne Geophysics to Identify
859 Salinization in West Texas*. <https://doi.org/10.23867/RI0257D>
- 860 Park, S. R., & Lee, S. W. (2020). Spatially varying and scale-dependent relationships of land use
861 types with stream water quality. *International Journal of Environmental Research and
862 Public Health*, *17*(5). <https://doi.org/10.3390/ijerph17051673>
- 863 Parkhurst, D. L., & Appelo, C. A. J. (2013). *Description of input and examples for PHREEQC
864 version 3 - A computer program for speciation and batch-reaction, one-dimensional
865 transport, and inverse geochemical calculations: U.S. Geological Survey Techniques and
866 Methods, book 6, chap. A43*. U.S. Geological Survey.
- 867 Piper, A. M. (1944). A graphic procedure in the geochemical interpretation of water-analyses.
868 *Eos Transactions, American Geophysical Union*, 914–928.
- 869 Qin, C., Zhu, A. X., Pei, T., Li, B., Zhou, C., & Yang, L. (2007). An adaptive approach to
870 selecting a flow partition exponent for a multiple flow direction algorithm. *International*

- 871 *Journal of Geographical Information Science*.
- 872 Rao, A., Burke, T. (1997). Principal Component Analysis of Hydrologic Data. In *Integrated*
873 *Approach to Environmental Systems* (pp. 275–290). Springer, Dordrecht.
- 874 Raymond, P. A., Oh, N. H., Turner, R. E., & Broussard, W. (2008). Anthropogenically enhanced
875 fluxes of water and carbon from the Mississippi River. *Nature*, 451(7177), 449–452.
876 <https://doi.org/10.1038/nature06505>
- 877 Reed, E. L. (1961). *A study of the salt water pollution of the Colorado River, Scurry and Mitchell*
878 *Counties, Texas: Consultant's report to Colorado River Municipal Water District, Big*
879 *Spring, Texas*.
- 880 Reitz, M., Sanford, W. ., Senay, G. B., & Cazenias, J. (2017). *Annual Estimates of recharge,*
881 *quick-flow runoff and ET for the contiguous US using empirical regression equations*. US
882 Geological Survey Data Release.
883 <https://www.sciencebase.gov/catalog/item/55d3730fe4b0518e35468e1e>
- 884 Rice, J. S., Emanuel, R. E., & Vose, J. M. (2016). The influence of watershed characteristics on
885 spatial patterns of trends in annual scale streamflow variability in the continental U.S.
886 *Journal of Hydrology*, 540, 850–860. <https://doi.org/10.1016/j.jhydrol.2016.07.006>
- 887 Richter, B.C., Dutton, A. R., & Kreitler, C. W. (1991). *Identification of sources and mechanisms*
888 *of salt-water pollution affecting ground-water quality: A case study, west Texas*.
- 889 Richter, Bernd C., & Kreitler, C. W. (1986). *Identification of sources of ground-water*
890 *salinization using geochemical techniques*.
- 891 Scanlon, B. R., Tachovsky, J. A., Keese, K., Slade, R. M., Howard, M. T., Wells, G. L., &
892 Mullins, G. J. (2005). *Groundwater - Surface water interactions: Texas* (Issue August).
- 893 Shelton, J. L., Engle, M. A., Bucciatti, A., & Blondes, M. S. (2018). The isometric log-ratio
894 (ilr)-ion plot: A proposed alternative to the Piper diagram. *Journal of Geochemical*
895 *Exploration*, 190(September 2017), 130–141. <https://doi.org/10.1016/j.gexplo.2018.03.003>
- 896 Shen, B., Wu, J., Zhan, S., Jin, M., Saporov, A. S., & Abuduwaili, J. (2021). Spatial variations
897 and controls on the hydrochemistry of surface waters across the Ili-Balkhash Basin, arid
898 Central Asia. *Journal of Hydrology*, 600(May), 126565.
899 <https://doi.org/10.1016/j.jhydrol.2021.126565>
- 900 Slade, R. M., & Buszka, P. M. (1994). *Characteristics of Streams and Aquifers and Processes*
901 *Affecting the Salinity of Water in the Upper Colorado River Basin , Texas*.
- 902 Slade, R. M. J., Bentley, J. T., & Michaud, D. (2002). Results of stream-flow gain-loss studies in
903 Texas, with emphasis on gains from and losses to major and minor aquifers, Texas, 2000.
904 *U.S. Geol. Surv. Open File Rep.*, 02-068, 136.
- 905 Sliva, L., & Williams, D. D. (2001). Buffer zone versus whole catchment approaches to studying
906 land use impact on river water quality. *Water Research*, 35(14), 3462–3472.
907 [https://doi.org/10.1016/S0043-1354\(01\)00062-8](https://doi.org/10.1016/S0043-1354(01)00062-8)
- 908 Stewart, B., Shanley, J. B., Kirchner, J. W., Norris, D., Adler, T., Bristol, C., Harpold, A. A.,
909 Perdrial, J. N., Rizzo, D. M., Sterle, G., Underwood, K. L., Wen, H., & Li, L. (2022).
910 Streams as Mirrors : Reading Subsurface Water Chemistry From Stream Chemistry Water

- 911 Resources Research. *Water Resources Research*, 1–20.
912 <https://doi.org/10.1029/2021WR029931>
- 913 Sullivan, P. L., Stops, M. W., Macpherson, G. L., Li, L., Hirmas, D. R., & Dodds, W. K. (2019).
914 How landscape heterogeneity governs stream water concentration-discharge behavior in
915 carbonate terrains (Konza Prairie, USA). *Chemical Geology*, 527(April).
916 <https://doi.org/10.1016/j.chemgeo.2018.12.002>
- 917 Team, R. C. (2021). *R: A language environment for statistical computing* (4.0.4). R Foundation
918 for Statistical Computing.
- 919 Texas Natural Resource Conservation Commission. (1999). *Surface Water/Groundwater*
920 *Interaction Evaluation for 22 Texas River Basins*.
921 [http://www.twdb.texas.gov/publications/reports/contracted_reports/doc/Surface-](http://www.twdb.texas.gov/publications/reports/contracted_reports/doc/Surface-Groundwater_Interaction.pdf)
922 [Groundwater_Interaction.pdf](http://www.twdb.texas.gov/publications/reports/contracted_reports/doc/Surface-Groundwater_Interaction.pdf)
- 923 Texas Water Development Board. (2012). *2012 State Water Plan*. 1–314.
924 http://www.twdb.texas.gov/publications/state_water_plan/2012/2012_SWP.pdf
- 925 Texas Water Development Board. (2017). *2017 State Water Plan Water for Texas*.
926 www.twdb.texas.gov
- 927 Tiwari, T., Lidman, F., Laudon, H., Lidberg, W., & Agren, A. (2017). GIS-based prediction of
928 stream chemistry using landscape composition, wet areas, and hydrological flow pathways.
929 *Journal of Geophysical Research: Biogeosciences*.
- 930 Torres, M. A., West, A. J., Clark, K. E., & Feakins, S. J. (2017). Mixing as a driver of temporal
931 variations in river hydrochemistry: 1. Insights from conservative tracers in the Andes-
932 Amazon transition. *Water Resources Research*. [https://doi.org/10.1111/j.1752-](https://doi.org/10.1111/j.1752-1688.1969.tb04897.x)
933 [1688.1969.tb04897.x](https://doi.org/10.1111/j.1752-1688.1969.tb04897.x)
- 934 Walter, J., Chesnaux, R., Gaboury, D., & Cloutier, V. (2019). Subsampling of regional-scale
935 database for improving multivariate analysis interpretation of groundwater chemical
936 evolution and ion sources. *Geosciences (Switzerland)*, 9(3).
937 <https://doi.org/10.3390/geosciences9030139>
- 938 White, A. F., & Blum, A. E. (1997). Effects of climate on chemical weathering in watersheds.
939 *Geochimica et Cosmochimica Acta*, 59(9), 1729–1747.
940 [https://doi.org/https://doi.org/10.1016/0016-7037\(95\)00078-E](https://doi.org/10.1016/0016-7037(95)00078-E).
- 941 Wlostowski, A. N., Molotch, N., Anderson, S. P., Brantley, S. L., Chorover, J., Dralle, D.,
942 Kumar, P., Li, L., Lohse, K. A., Mallard, J. M., McIntosh, J. C., Murphy, S. F., Parrish, E.,
943 Safeeq, M., Seyfried, M., Shi, Y., & Harman, C. (2021). Signatures of Hydrologic Function
944 Across the Critical Zone Observatory Network. *Water Resources Research*, 57(3).
945 <https://doi.org/10.1029/2019WR026635>
- 946 Xu, H., Cai, C., Du, H., & Guo, Y. (2021). Responses of water quality to land use in riparian
947 buffers: a case study of Huangpu River, China. *GeoJournal*, 86(4), 1657–1669.
948 <https://doi.org/10.1007/s10708-020-10150-2>
- 949 Yurtseven, E., Çolak, M. S., Öztürk, A., & Öztürk, H. S. (2018). Drainage water salt load
950 variations related to the salinity and leaching ratios of irrigation water. *Tarım Bilimleri*
951 *Dergisi*, 24(3), 394–402. <https://doi.org/10.15832/ankutbd.456667>

952

953