Hydrogeology Journal – Electronic Supplementary Material

Groundwater salinity mapping using geophysical log analysis within the Fruitvale and Rosedale Ranch oil fields, Kern County, California, USA

Authors:

Michael J Stephens[1]*, David H Shimabukuro[2], Janice M Gillespie[1], Will Chang[3]

1. U.S. Geological Survey, California Water Science Center, 6000 J Street, Sacramento, CA 95819
2. Department of Geology, California State University, Sacramento, 6000 J Street, Sacramento, CA 95819
3. Hypergradient LLC, Berkeley, CA 94703

*Corresponding author, mjstephens@usgs.gov

S1. *TDS Model Performance*

The TDS model predicts TDS using borehole geophysical log data, the Archie Equation, and kriging. The model parameters are Archie's *a* and *m*. Different *a* and *m* values yield significantly different results (Fig. 5 of the main article). To achieve the most accurate TDS predictions possible, an objective function was constructed to take two sets of *a* and *m* parameters (one set per oil field), make TDS predictions, compare those predictions to measured TDS data, and return the sum of the squared residuals. This function was minimized to find optimal *a* and *m* values. Fig. 5 shows optimization produces better results than using standard parameterizations.

It is also of interest to quantify how well a model can predict previously unseen data to evaluate model performance or check for problems (i.e. overfitting) (James et al. 2013). This can be accomplished by splitting observations into two sets, a training set and a test set. The TDS model can be derived from just the training set, and made to predict TDS for locations in the test set. Its performance is summarized as the mean of the square of prediction error. To accurately estimate prediction error, the TDS model was run 100 times, each time randomly splitting the 64 observations into a training set of 51 data points and a test set of 13 data points. The median and standard deviation of the 100 mean-square-error values are 0.0407 and 0.0312 (for log TDS), respectively. The median implies that the relative error of TDS prediction is around $e^{\sqrt{0.0407}} - 1$, or 22%. This demonstrates that the TDS model can accurately and consistently predict TDS values that were not used in setting model parameters.

S2. *TDS through time*

The TDS model in this report uses resistivity and produced water sample data that span approximately four decades (Fig. S1). If the deep groundwater TDS was dynamic over that time the data used in the model may have a temporal bias. To investigate whether TDS changed significantly with time the TDS model errors were examined. Presumably, if deep groundwater quality was heavily impacted by oil field processes the TDS model would not predict the older and the newer TDS measurements with equal accuracy. However, as shown in Fig. S1c the TDS model prediction error and dates of samples are not correlated (Pearson r = -0.13). This suggests the deep groundwater system has not been heavily impacted by oil field activities, at least not to a large spatial extent. This also agrees well with the injection history of the fields. Most of the injected fluids were produced from the same zones, consequently, the TDS would not be expected to change over time (DOGGR 2018). The remaining injection into different zones than it was sourced may have impacted water quality but is likely limited to small areas of the fields and at deeper depths outside the volume of analysis where much water disposal is located. Therefore, localized changes in groundwater quality over time may have taken place, but apparently do not significantly affect the TDS model data sets.

ESM References

Betts PW (1955) Rosedale Ranch oil field, California Div. Oil and Gas, California Oil Fields - Summary of Operations 41:35-39

California Department of Water Resources (2016) Water data library, http://wdl.water.ca.gov/waterdatalibrary/ Accessed 2016

California Division of Oil, Gas, and Geothermal Resources (2018) Injection/Production data, https://secure.conservation.ca.gov/WellSearch Accessed 2015-2018

Hluza AG (1961) Calloway area of Fruitvale oil field, California Div. Oil and Gas, California Oil Fields - Summary of Operations 47:5-13

Hluza AG (1965) Main area of Fruitvale oil field, California Div. Oil and Gas, California Oil Fields - Summary of Operations 51:31-39

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning (Vol. 112). New York: Springer

Kitanidis PK (1997) Introduction to geostatistics: applications in hydrogeology. Cambridge University Press

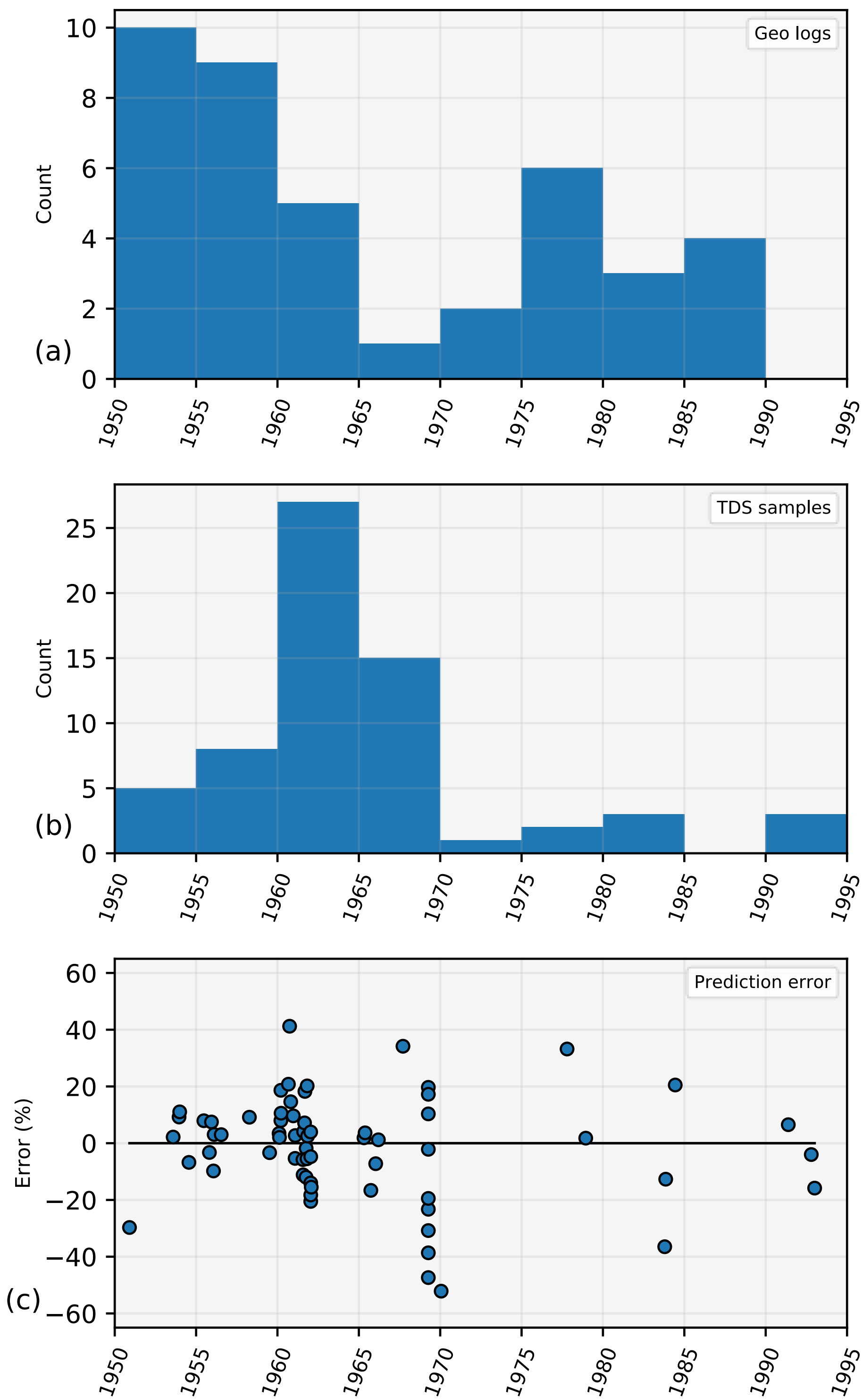

Fig. S1) a Date distributions of the spud dates (when well drilling began) of wells used to obtain formation resistivity and **b** the sample dates of the TDS data. **c** The relative prediction error of the TDS model. The errors do not correlate with time (Pearson r = -0.13) suggesting that the TDS within the oil fields has not significantly changed over time. The black line indicates zero error.

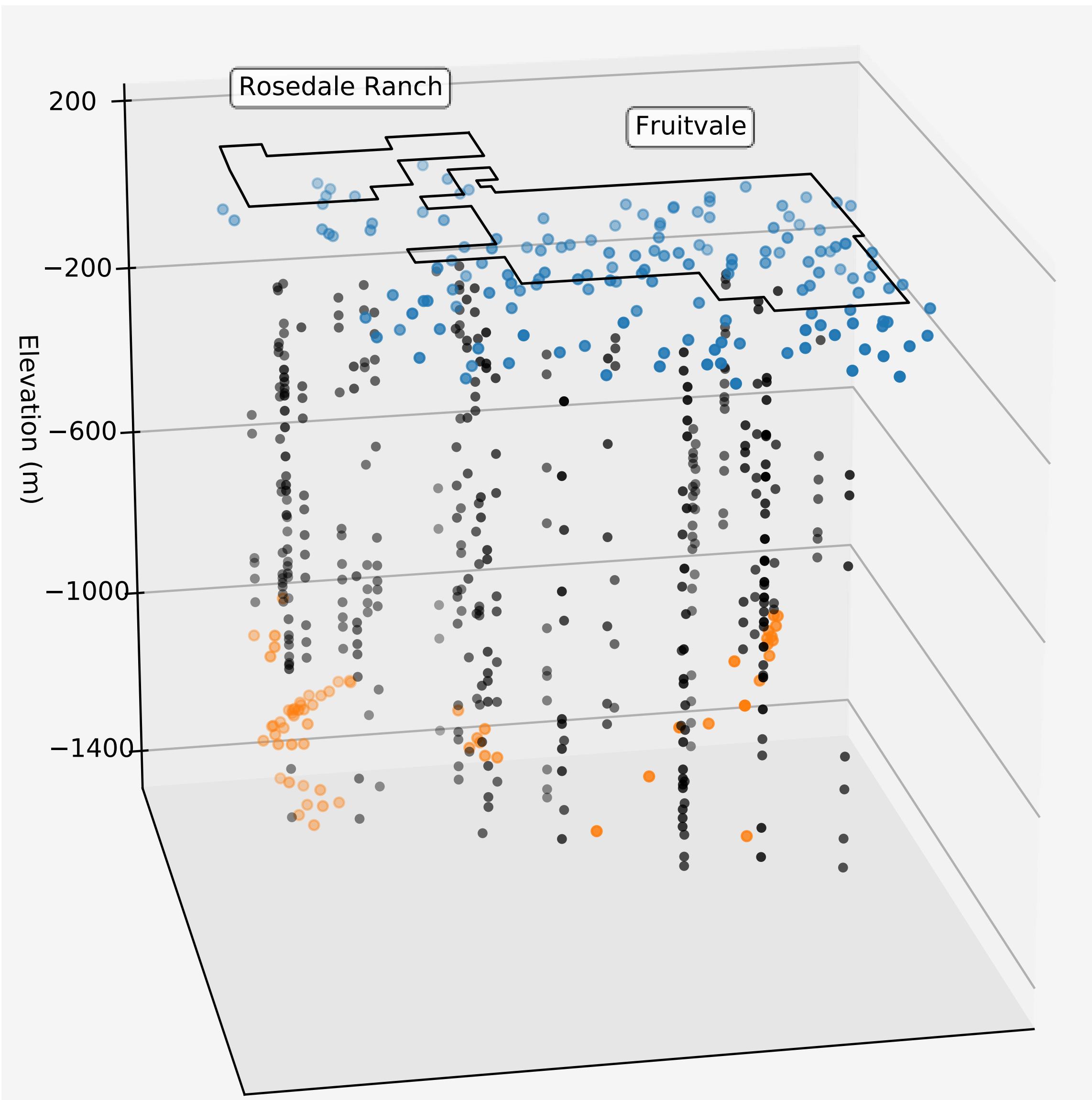

Fig. S2) Three-dimensional projection of the available data distributions. The water well data are located at shallow depths (blue). The produced water TDS measurements are derived from deeper depths (orange). The Archie-derived TDS from resistivity measurements from well logs are used to fill in the vertical and spatial gaps (black).

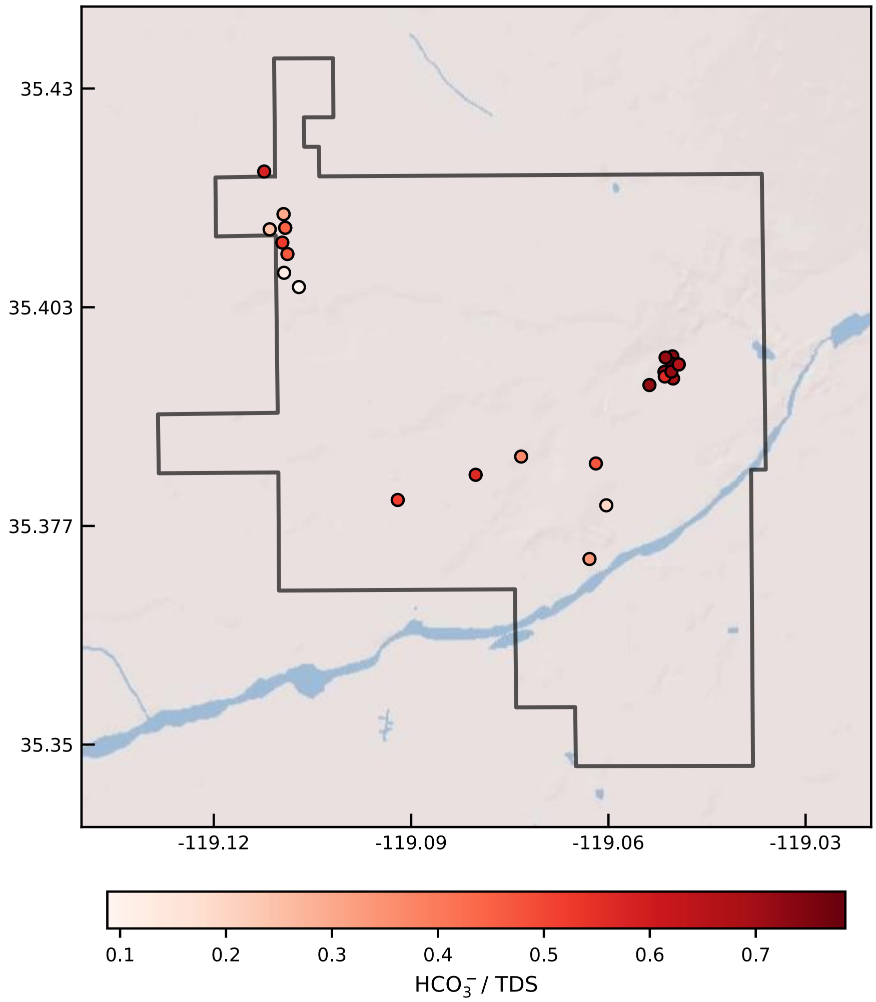

Fig. S3) Map showing the location of elevated bicarbonate levels. The highest ratios are co-located with recharge (see Fig. S6).

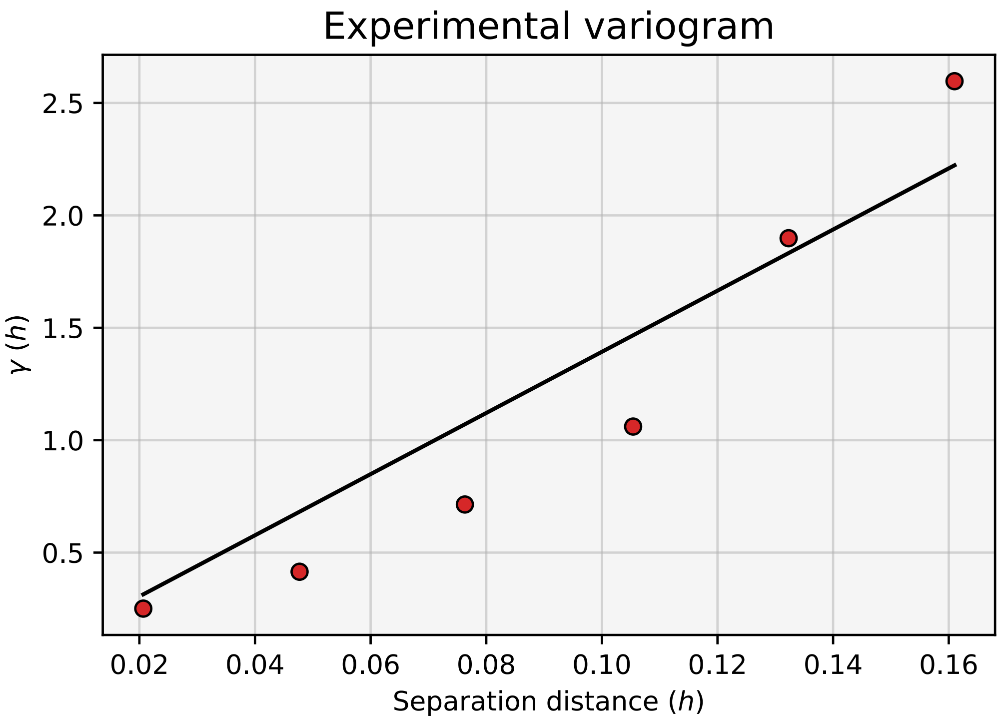

Fig S4) Experimental variogram of the Archie-derived TDS data from Rosedale Ranch and Fruitvale. The experimental variogram describes the covariance structure of the data (Kitanidis 1997). The x-axis units are in decimal degrees, the y-axis is the semivariance.

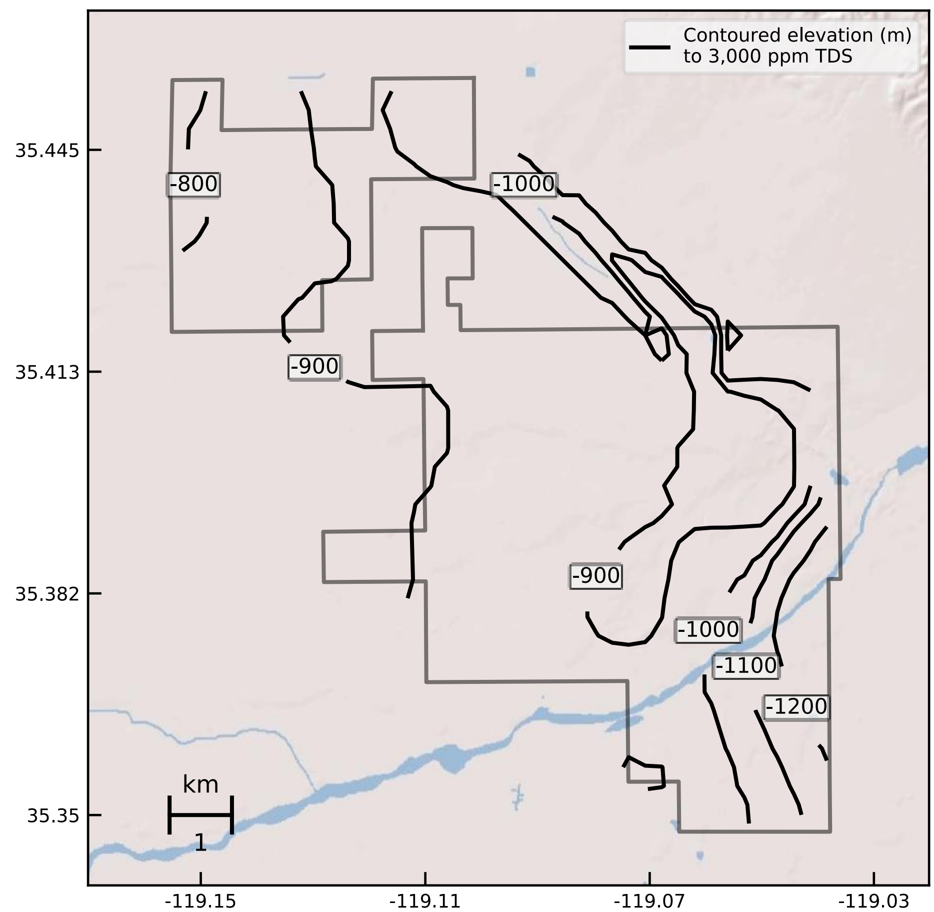

Fig S5) Contoured elevation of the 3,000 ppm TDS surface.

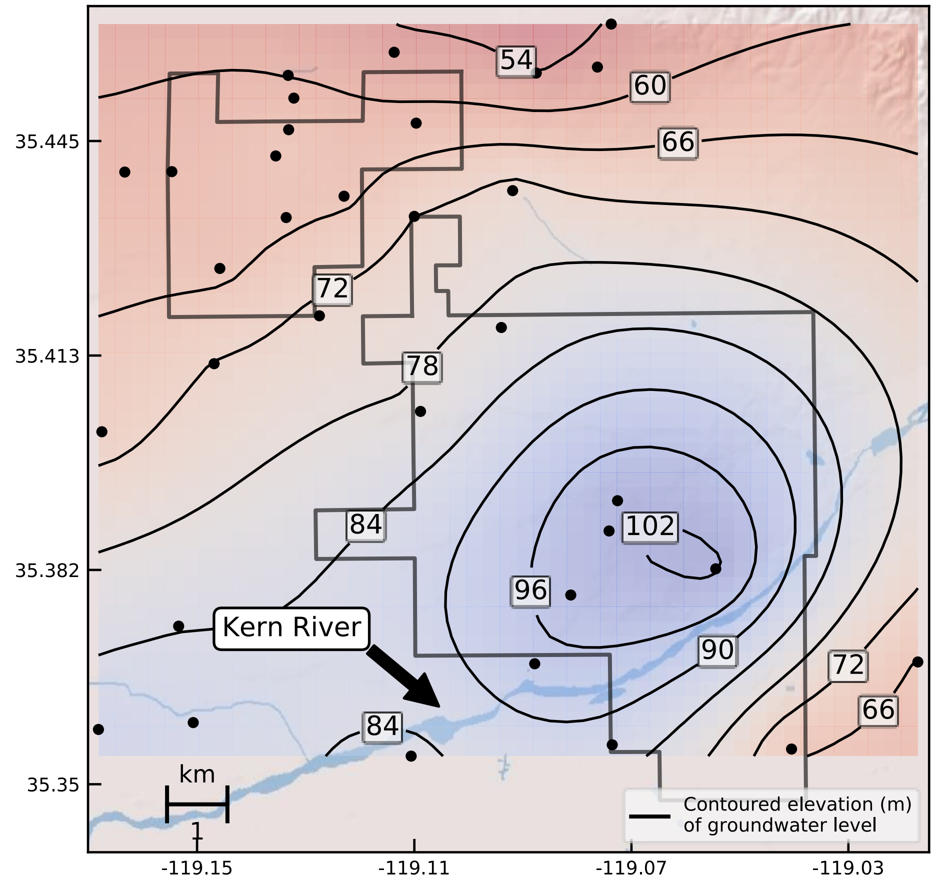

Fig S6) Contoured groundwater elevation data from wells in the study area for the 2000 water year. Shallow groundwater flow originates from the Kern River, indicating it provides fresh water recharge to the area (data from California Department of Water Resources 2016).

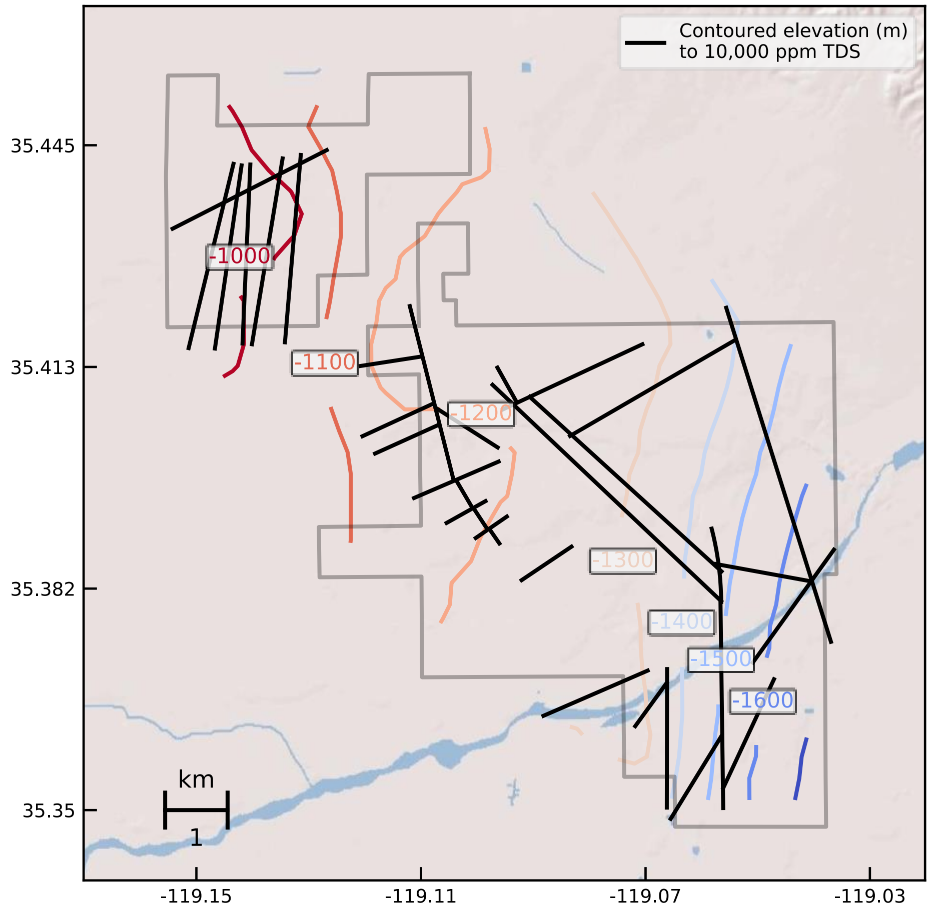

Fig S7) Contoured elevation of the 10,000 ppm TDS surface with faults. Fault data were derived from contour maps with depths of ~792 m (2,600 ft) below sea level in Fruitvale main area, ~1,036 m (3,400 ft) in Fruitvale Calloway area (NW Fruitvale), and ~1,036 m (3,400 ft) in Rosedale Ranch (Betts 1955; Hluza 1961, 1965).