Environmental Science & Technology

Article

pubs.acs.org/est

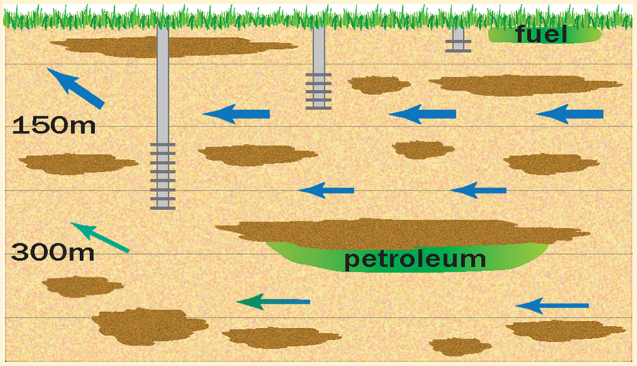
Geogenic Sources of Benzene in Aquifers Used for Public Supply, California

Matthew K. Landon*,† and Kenneth Belitz†

†U.S. Geological Survey, San Diego, California 92101, United States

Supporting Information

ABSTRACT: Statistical evaluation of two large statewide data sets from the California State Water Board's Groundwater Ambient Monitoring and Assessment Program (1973 wells) and the California Department of Public Health (12 417 wells) reveals that benzene occurs infrequently (1.7%) and at generally low concentrations (median detected concentration of 0.024 μg/L) in groundwater used for public supply in California. When detected, benzene is more often related to geogenic (45% of detections) than anthropogenic sources (27% of detections). Similar relations are evident for the sum of 17 hydrocarbons analyzed. Benzene occurs most frequently and at the highest concentrations in old, brackish, and reducing groundwater; the detection frequency was 13.0% in groundwater with tritium <1 pCi/L, specific conductance >1600 μS/cm, and anoxic conditions. This groundwater is typically deep (>180 m). Benzene occurs somewhat less frequently in recent, shallow, and reducing groundwater; the detection frequency was 2.6% in groundwater with tritium ≥1 pCi/L, depth <30 m, and anoxic conditions. Evidence for geogenic sources of benzene include: higher concentrations and detection frequencies with increasing well depth, groundwater age, and proximity to oil and gas fields; and higher salinity and lower chloride/iodide ratios in old groundwater with detections of benzene, consistent with interactions with oil-field brines.



1. INTRODUCTION

Spills from fuel storage and distribution systems have long been investigated as sources of groundwater contamination.[1−3] These spills have been of concern because of the high mutagenicity and potential threat to human health of soluble aromatic hydrocarbons derived from fuels and the widespread distribution of fuel sources across the landscape.[4−6] These concerns have led to widespread regulatory and industry efforts to limit spills from fuel sources, to identify and monitor spills as early as possible, and to implement remediation efforts where necessary to limit fuel plume migration toward wells used for drinking-water supply.[6] In California, a cluster of birth defects and cancer near the community of Kettleman City has served as a recent example of the importance of investigating the links between multiple sources of environmental toxins, including hydrocarbons in groundwater, and human health.[7]

The distribution of fuel sources across the landscape is nearly ubiquitous, so it has been less commonly noted that the presence of hydrocarbons in groundwater can result from naturally occurring subsurface petroleum reservoirs. These natural occurrences and the accompanying anthropogenic effects of petroleum extraction activities on the subsurface environment are defined as geogenic sources in this article. Petroleum reservoirs that can be geogenic sources of hydrocarbons to groundwater are widespread across the globe.[8] Some basin-scale studies have described the presence of hydrocarbons from geogenic sources in aquifers overlying oil or gas fields.[9−13] Extensive contact with groundwater can deplete the most soluble phases of crude oils, particularly aromatics, through a process called water washing.[14] Brines are often associated with oil and gas fields,[12,15] so geochemical signatures such as increased salinity, chloride, and sodium, and diagnostic changes in minor ion and trace element signatures[10,16] have been used to identify mixing between oil-field brines and groundwater in aquifers used for supply. However, this is the first regional assessment of the relative importance of geogenic sources of hydrocarbons to groundwater aquifers used for public supply in petroleum producing regions. Because California exhibits large urban populations dependent on groundwater supplies superimposed with a high density of fuel sources and abundant hydrocarbon reservoirs,[17,18] this region is well suited for assessing the importance of geogenic hydrocarbons to groundwater quality.

The objective of this article is to identify the extent to which hydrocarbons in public-supply aquifers in California are due to geogenic sources. As part of this objective, the article reports the results of statistical analysis of relations of hydrocarbon concentrations to explanatory variables.

Received: May 21, 2012
Revised: July 17, 2012
Accepted: July 23, 2012
Published: July 23, 2012

ACS Publications

This article not subject to U.S. Copyright. Published 2012 by the American Chemical Society

2. METHODS

2.1. Data Sources. Data for this article comes from two sources: (1) data collected by the U.S. Geological Survey (USGS) as part of the California State Water Board's Groundwater Ambient Monitoring and Assessment program [19] Priority Basin Project (GAMA),[20] and (2) data from the California Department of Public Health (CDPH), assembled for drinking water regulatory monitoring purposes.[21] These data sets offer complementary information. The GAMA data includes age tracers, geochemical indicators, and has low laboratory reporting levels (LRL), a median for volatile organic compounds (VOCs) of <0.06 μg/L, at a smaller number of wells (~2000), whereas the CDPH data has less extensive ancillary data and higher laboratory reporting levels, a median for VOCs of 0.5 μg/L, at a larger number of wells (~12000) (Supporting Information, Additional Methods Description). This study, a partnership of the USGS Priority Basin Project and U.S. Environmental Protection Agency, includes GAMA data collected from 1973 wells during May 2004 through August 2010 (median, November 2006). CDPH data included in the study from 12 441 wells were collected between July 1984 and August 2010 (median, March 2005); > 75% of the data was from the 2000s.

GAMA data were collected using protocols[22,23] to ensure that representative samples are collected while minimizing the potential for sample contamination. Detailed descriptions of sample collection, analysis, and quality-assurance (QA) methods and results are available.[20] Samples were analyzed for VOCs using purge and trap capillary-column gas chromatography/mass spectrometry (P&T GC/MS) at the USGS National Water Quality Laboratory.[24] Approximately 80% of the wells in the GAMA data were public-supply or other similarly constructed production wells, including those used for commercial or irrigation purposes, 10% were domestic wells, and 10% were observation wells. The CDPH database lists wells used for municipal and community drinking water supplies, and includes wells from systems classified as nontransient (such as cities, towns, and mobile-home parks) and transient (such as schools, campgrounds, and restaurants).

GAMA data were used for evaluating occurrence and distribution of hydrocarbons in California groundwater because these data were collected using spatially distributed randomized well selection in grid cells covering priority groundwater basins, low LRLs, and QA. The more extensive CDPH data, with higher method detection levels (MDLs), were used for assessing the distribution at higher concentrations and to evaluate explanatory relations identified using the GAMA data. Analytical results for VOCs reported in the CDPH data were determined using EPA methods 502.2 (P&T GC) or 524.2 (P&T GC/MS).[21] The differences in the methodologies, LRLs, and QA for VOC data reflect that the GAMA data were collected by a single study for the purpose of scientific assessment while the CDPH data were compiled from many different institutions and laboratories for the purpose of monitoring of public water supplies for compliance with drinking water standards. The detection frequencies based on CDPH data need to be regarded with caution because the MDLs were not always recorded in the database; the detection frequencies are reported relative to the most common MDLs.

2.2. Analytes. Because of expected low detection frequencies of fuel hydrocarbons in public-supply wells,[3−5,25−28] this study compiled data for 17 hydrocarbons included on standard USGS VOC analyses [29] (Table SI-1 of the Supporting Information) in an attempt to have sufficient data to evaluate variables explaining detections. Benzene was selected as the primary dependent variable because it is: (1) the most commonly occurring hydrocarbon in the GAMA data, (2) expected to be the most commonly detected hydrocarbon because of its physiochemical properties including a high solubility among petroleum hydrocarbons, (3) hazardous to human health, and (4) the most reliably measured hydrocarbon based on QA analysis. The sum of hydrocarbon concentrations was selected as a secondary dependent variable (Table 1) because of the common co-occurrence of multiple hydrocarbons.

Table 1. Detection Frequencies for Benzene and Sum of Hydrocarbons in GAMA and CDPH Data[a]

	benzene	sum of hydrocarbons
CAS number	71−43−2	
threshold type[1]	MCL-CA	
threshold value (μg/L)	1	
GAMA		
wells	1972	1973
maximum LT-MDL (μg/L)	0.013	
detections at or above LT-MDL	33	53
detection frequency (LT-MDL), %	1.67	2.69
detections at or above CDPH MDL	4	
detection frequency (CDPH MDL), %	0.20	
detected concentrations, maximum, μg/L	78.9	79.8
detected concentrations, median, μg/L	0.024	0.09
detections above threshold	3	
detection frequency above threshold, %	0.15	
CDPH		
wells	12 417	12 441
mode MDL, μg/L	0.5	0.5
detections at or above CDPH MDL	16	71
detection frequency (CDPH MDL), %	0.13	0.57
detected concentrations, maximum, μg/L	140	200
detected concentrations, median, μg/L	3.45	0.60
detections above threshold	11	
detection frequency above threshold, %	0.09	

[a]CAS, Chemical Abstract Service; LT-MDL, long-term method detection level; MCL-CA, California Department of Public Health maximum contaminant level; μg/L, micrograms per liter.

2.3. Data Processing. A single analysis at each well was selected to avoid biasing the data sets to those wells having many analyses over time. About 90% of the wells sampled by GAMA were visited once. For CDPH data, the most recent concentration of each VOC and other target constituents was selected and evaluated using the QA steps described below.

Study-specific reporting levels (SRLs) for toluene, ethylbenzene, *o*-xylene, *m*- and *p*-xylene, and 1,2,4-trimethylbenzene were developed by the GAMA program on the basis of evaluation of blank samples that included detections of these compounds at concentrations and frequencies overlapping with environmental samples.[30] These SRLs were used in this study to exclude analytical results with a reasonable likelihood of being false positives; detections of SRL compounds in 290 samples were treated as nondetections (Supporting Information). Hydrocarbons other than the 5 listed above were not detected in blank samples and did not have SRLs different from

LRLs (28 samples). In 25 GAMA samples, data for the 5 constituents were not screened to SRLs because of independent evidence that detections were likely to be environmental rather than sampling artifacts, including: detection of co-occurring hydrocarbons that were not censored based on detections in blanks, historical detections of hydrocarbons in that well in the CDPH database, or concentrations of the 5 constituents above SRLs.

The CDPH data are collected for regulatory purposes at thousands of wells by different individuals and laboratories, so this data cannot be rigorously evaluated for quality as can be done for the GAMA data. Following CDPH recommendations, a single detection of a constituent is not considered to represent an actual occurrence until the detection is verified by subsequent sample results. For CDPH wells with hydrocarbon detections in the most recent sample, the historical data were evaluated for previous detections. If there were previous detections, the most recent result was used to represent that well. Detections of a single constituent with no other historical detections were treated as nondetections in the most recent CDPH data to avoid false positives; results from wells with only a single analysis were retained in the CDPH data analyzed. The screening procedures resulted in apparent detections being treated as nondetections in 55 samples; 29 of these samples had a single detection of toluene only in the most recent sample. This screening was done to remove detections from the data that were not likely to represent aquifer conditions.

2.4. Explanatory Factors. Continuous potential explanatory variables (Table 2) discussed in this article include variables used in previous studies in California[35−39] and previous studies of VOCs.[40] Geographic information system analysis was done to attribute GAMA and CDPH wells with ancillary spatial data. Analyses of additional continuous and categorical variable data, including water balance, geology, depth to water table, well type, land surface elevation, and soil properties, compiled for this study did not indicate consistent relations with hydrocarbon distribution and are not discussed in this article.

GAMA data were classified on the basis of salinity, age, depth, reduction−oxidation (redox) conditions, and horizontal proximity to an oil and gas field. Detection frequencies were compared between the resulting data bins. For tritium (^{3}H), a tracer of groundwater that was in contact with the atmosphere from above-ground nuclear testing since 1950, a threshold activity of 1 pCi/L was selected for distinguishing water that was likely to have recharged aquifers prior to 1950 (old groundwater, < 1 pCi/L) from post-1950 water (recent groundwater, ≥ 1 pCi/L).[41−45] For salinity, specific conductance (SC), an electrical proxy measure of total dissolved solids (TDS), was available at all GAMA sites, whereas TDS was not. The CDPH upper recommended secondary maximum contaminant level for SC of drinking water was used as a criteria to distinguish brackish (>1600 μS/cm) from nonsaline (≤1600 μS/cm). Well depths were classified into three bins, < 30 m, 30−180 m, and >180 m based on a bimodal relation of hydrocarbon occurrence to well depth and visual identification of possible well depth classification thresholds; these 3 bins were further simplified to a binary classification of <180 m (deep) and >180 m (shallow and intermediate) for use with other variables. Redox conditions were classified based on measured concentrations of dissolved oxygen (DO), nitrate, manganese, iron, and sulfate.[46,47] The redox categories of oxic, suboxic, mixed, and anoxic [46] were further lumped into a binary redox classification of oxic and reducing. The reducing classification includes suboxic, mixed, and anoxic categories (Table SI-2 of the Supporting Information). On the basis of analysis of the relation between hydrocarbon occurrence and minimum distance to the nearest oil and gas field, a threshold value of 5 km was identified for binary classification.

Table 2. Spearman's ρ Correlation Test Results between Primary Continuous Explanatory Variables and Benzene[a]

variable, units (reference)	GAMA benzene	GAMA old benzene	GAMA recent benzene	CDPH benzene
source				
Anthropogenic				
LUFT density, no./km^2 [31]	ns	ns	ns	ns
distance to nearest LUFT, m [31]	ns	ns	ns	ns
population density, people/km^2 [32]	ns	ns	ns	ns
commercial land use, % within 500 m of well [33]	+[b]	ns	+[b]	ns
high intensity residential land use, % within 500 m of well [33]	ns	ns	ns	ns
urban land use, % within 500 m of well [33]	ns	ns	ns	ns
Geogenic				
distance to nearest oil and gas field, m [18]	−[c]	ns	ns	ns
receptor				
depth to top of perforations, m bls[20,21]	ns	ns	ns	ID
well depth, m bls[20,21]	+[c]	ns	ns	ID
transport				
Age/depth indicators				
pH, standard pH units[20,21]	ns	ns	ns	+[c]
tritium, pCi/L [20]	−[c]	ns	ns	ID
carbon-14, pmc [20]	−[c]	ns	ns	ID
Aquifer stress/aquifer characteristics				
CDPH well density, wells/km^2 (calculated from 21)	ns	ns	ns	ns
soil permeability, cm/hour [34]	ns	ns	ns	ns
Oxidation−reduction conditions				
dissolved oxygen, mg/L [20]	-[d]	-[d]	ns	ID
nitrate, mg/L as N[20,21]	-[d]	-[d]	ns	-[d]
manganese, μg/L[20,21]	+[d]	+[d]	ns	ns
iron, μg/L[20,21]	+[d]	+[d]	+[d]	ns

[a]+, significant positive correlation; -, significant inverse correlation; ns, correlation not significant; ID, insufficient data to evaluate correlation; significant correlations were determined on the basis of p values (significance level of the Spearman's test) less than the threshold value (α) of 0.05; km^2, square kilometers; no., number; m, meters; %, percent; LUFT, Leaky Underground Fuel Tank; pCi/L, picocuries per liter; pmc, percent modern carbon; cm, centimeters; mg/L, milligrams per liter; μg/L, micrograms per liter; bls, below land surface; N, nitrogen. [b]Correlations with anthropogenic source variables consistent with increasing concentrations with increasing source density and proximity. [c]Correlations of benzene with variables indicating increasing concentrations with proximity to oil and gas fields and greater depth and age. [d]Correlations of benzene with redox conditions indicating increasing concentrations with reducing groundwater.

2.5. Statistical Methods. Nonparametric rank-based methods were used for statistical analysis. These techniques are generally not affected by outliers and do not require that the data follow a normal distribution.[48] The significance level (p)

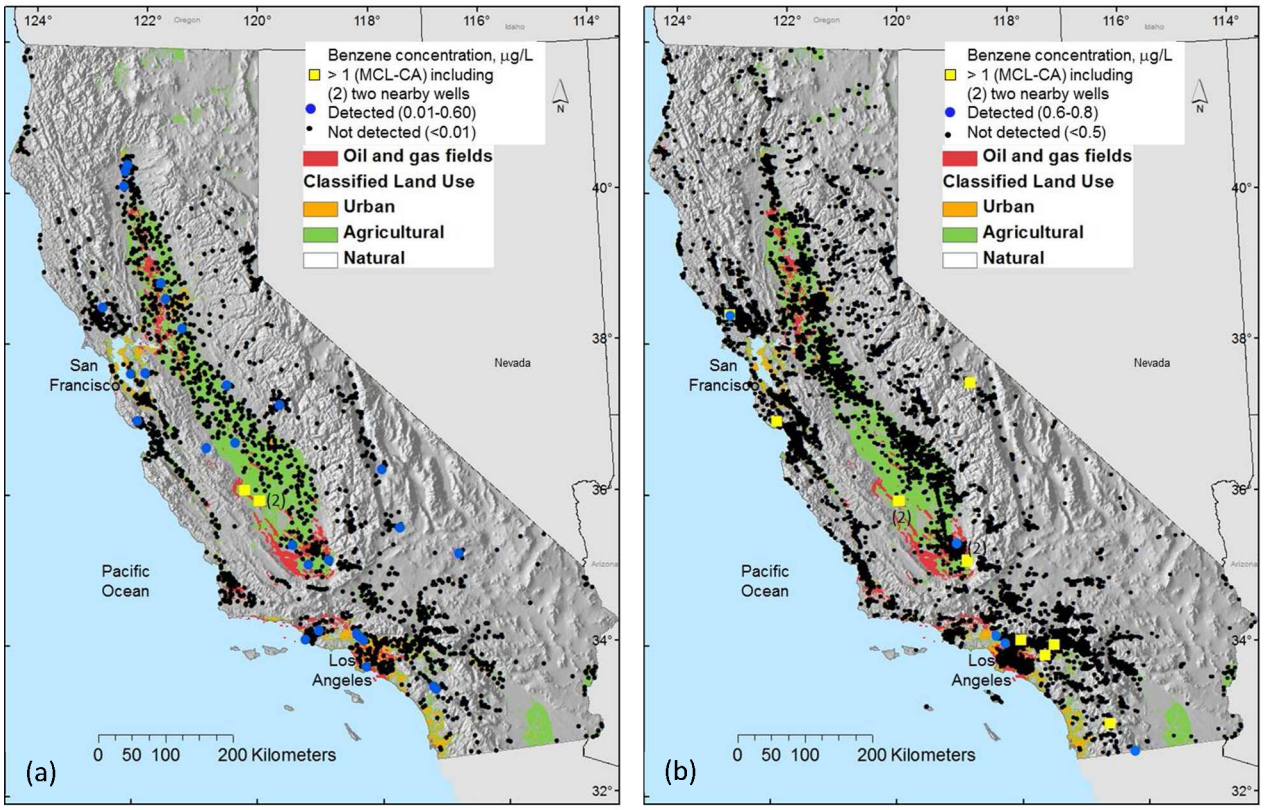


Figure 1. Map showing results of analyses for benzene overlain on oil and gas fields and land use for (a) GAMA data and (b) CDPH data.

used for hypothesis testing was compared to a threshold value (α) of 5% (α = 0.05) to evaluate whether the relation was statistically significant ($p < \alpha$). Correlations were investigated using Spearman's method to calculate the rank-order correlation coefficient (ρ) between concentrations and continuous explanatory variables. The Wilcoxon rank-sum test was used to evaluate the differences between two groups of data.[48] The Kruskal–Wallis test was used to test differences among more than two groups.[49] A Pearson's chi-square (χ^2) contingency table test was used to evaluate whether detection frequencies differed between classified variable bins. All statistical analysis was done using *S-PLUS for Windows*, version 8.1, Professional Edition (TIBCO Software Inc.).

3. RESULTS

Both the GAMA and CDPH data primarily represent depth zones of aquifers used for public supply across California. The median well depth was 120 m for the GAMA data and 140 m for CDPH data (Figure SI-1 of the Supporting Information). Public-supply wells are primarily at depths sufficient to protect the wells from surface contamination and are located in or near the communities that they serve. The GAMA data are primarily collected from randomized spatially distributed wells across entire groundwater basins and are not always located in or near communities, as CDPH wells generally are. These differences in the spatial distribution of wells result in significant differences in well depth and urban land use between the GAMA and CDPH data. Well depths were significantly greater for CDPH data than for GAMA data (Wilcoxon test statistic $Z = -36.3428$, $p < 0.0001$); similarly urban land use was significantly greater for CDPH data than for GAMA data (Wilcoxon test statistic $Z = 16.6783$, $p < 0.0001$; Figure SI-2 of the Supporting Information).

Shallower aquifer zones above those primarily used for public supply are not directly assessed in this article with the exception of a limited number of GAMA wells within these shallower zones. Hydrocarbons derived from surficial anthropogenic sources are likely to be detected more frequently in these shallower aquifer zones.

3.1. Hydrocarbon Occurrence. Detection frequencies and concentrations of hydrocarbons were low in public-supply aquifers of California. Benzene was the most frequently detected hydrocarbon, at about 1.7%, in the GAMA data (Table 1). Fifteen other hydrocarbons were detected in the GAMA data at frequencies ≤1.0% (Table SI-1 of the Supporting Information). One or more hydrocarbons were detected in 2.7% of GAMA samples. Spatial-weighting of the data could result in slight changes in the raw detection frequencies reported here; however, raw estimates rarely fall outside of the 90% confidence intervals of spatially weighted estimates.[50] The median detected concentration was 0.024 μg/L for benzene and 0.09 μg/L for sum of hydrocarbons (Table 1). Three wells in the GAMA data had concentrations of any hydrocarbon above a regulatory threshold, a frequency of detection above thresholds of 0.15%; these wells had benzene concentrations above the California Maximum Contaminant Level (MCL) of 1.0 μg/L and were located in the southwestern portion of the Central Valley (part A of Figure 1). Benzene detections were widely distributed across California but were especially prevalent in parts of the Central Valley and Southern California. These areas exhibit both oil and gas fields and extensive urban land use.

CDPH MDLs were 2−100 (median 20) times higher than the GAMA long-term MDLs; consequently, detection frequencies for the CDPH data were lower than for the GAMA data. However, when GAMA data were censored to the most common MDLs of the CDPH data, detection frequencies of hydrocarbons for the GAMA data were generally similar to those in the CDPH data (Tables 1 and SI-1 of the Supporting Information). Detected concentrations in the CDPH data were higher than in the GAMA data reflecting higher MDLs and sample numbers about 10 times larger for CDPH, but the frequency of detections above regulatory thresholds were similar for both data sets (<0.2%, Table 1). In the CDPH data, two hydrocarbons were detected above regulatory thresholds: benzene (0.09% detection frequency above threshold, 11 wells) and toluene (0.01%, 1 well). CDPH wells with benzene concentrations above the MCL were primarily located in inland Southern California, the southern Central Valley, and Central California coastal areas, similar to the distribution of detections in the GAMA data (part B of Figure 1).

3.2. Relation to Explanatory Factors. Concentrations of benzene and sum of hydrocarbons are correlated with each other (ρ = 0.78, p < 0.0001) and with the same explanatory variables in nearly all cases (Tables 2 and SI-3 of the Supporting Information). Two or more hydrocarbons were detected in 70% of GAMA samples in which any hydrocarbon was detected.

3.2.1. GAMA Data. Benzene concentration is positively correlated with well depth and negatively correlated with ^{3}H and carbon-14 (^{14}C) values (Table 2), contrary to expectations if benzene is assumed to be derived from anthropogenic sources. Moreover, samples of old groundwater (^{3}H < 1 pCi/L) had a significantly (χ^2 = 6.6834, p = 0.0097) higher detection frequency, 2.8%, than samples of recent groundwater, with 1.1% (part A of Figure 2). Benzene detections exhibited a bimodal relation with respect to well depth, with significantly (χ^2 =13.6933, p = 0.0011) higher detection frequencies for wells <30 m of 3.13%, and >180 m of 3.92%, than for wells with intermediate depths of 30−180 m, with 1.1% (part B of Figure 2). This bimodal relation is consistent with dual benzene sources located in proximity to both deep and shallow groundwater .

Detection frequencies of benzene are highest in groundwater that is brackish, old, and reducing (13.0%) or nonsaline, old, deep, and reducing (8.2%)(Figures SI-3 and SI-4 of the Supporting Information); similar results occurred for sum of hydrocarbons (Table SI-4 of the Supporting Information). Among brackish groundwater, benzene was only detected in samples that are old and reducing, not samples that are recent or oxic. The detection of benzene in brackish, old, and reducing groundwater spanned a range of depths. In contrast, among nonsaline groundwater, including depth as a classification criterion helped to organize the data into bins having higher detection frequencies than for the entire GAMA data (1.7%). Detection frequencies are always higher for samples that are deep rather than shallow or intermediate. This is also true for samples that are reducing rather than oxic, independent of all combinations of age and other variables.

Higher concentrations and detection frequencies of benzene (and hydrocarbons) in old groundwater are associated with increasing proximity to oil and gas fields. Among old groundwater samples, this relation can best be observed as an inverse correlation between concentrations of hydrocarbons and the minimum distance to the nearest oil and gas field

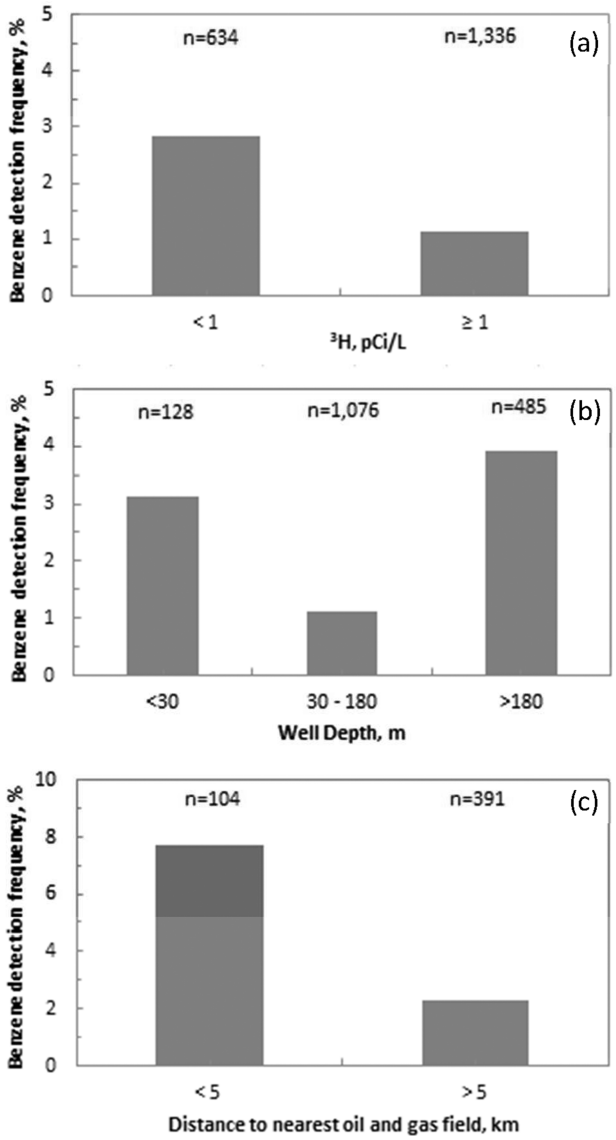


Figure 2. Detection frequency of benzene in GAMA data categorized by: (a) ^{3}H, (b) well depth, and (c) in old groundwater samples, proximity to nearest oil and gas field.

(Table SI-3 of the Supporting Information). Among old GAMA samples, benzene detection frequency is significantly higher (χ^2 = 5.6641, p = 0.0173) in wells <5 km from oil and gas fields (7.7%) than in wells ≥5 km away (2.3%) (part C of Figure 2). For nearly all depth and redox combinations in GAMA old data, benzene and hydrocarbon detection frequencies were higher for wells <5 km from oil or gas fields than for wells ≥5 km away (Table SI-4 of the Supporting Information). Significant correlations of hydrocarbon concentrations with ^{14}C (inverse) and depth to top of perforations (positive) indicate higher concentrations in older and deeper water. These correlations are consistent with interactions with geogenic sources of petroleum.

Independent geochemical evidence supports the interpretation that detections of benzene and hydrocarbons in old groundwater are related to interactions with oil-field brines associated with geogenic petroleum sources. SC was significantly higher for old groundwater with benzene detections than for old groundwater without benzene detections and recent

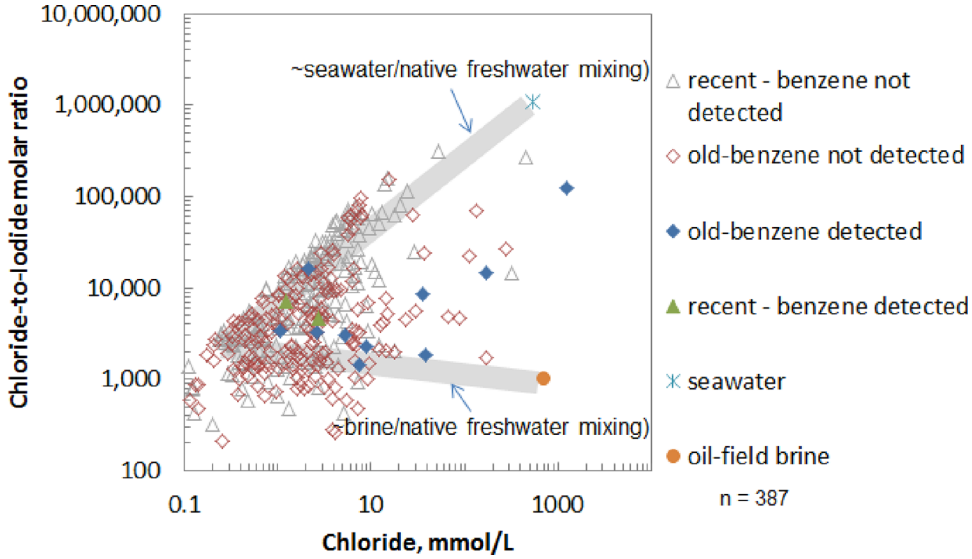


Figure 3. Graph of chloride and chloride/iodide for GAMA data relative to seawater and oil-field brine end-members[10,51] and approximate mixing curves with native freshwater.

groundwater with and without benzene detections (Kruskal–Wallis $\chi^2 = 11.6493$, $p = 0.0087$) (Figure SI-5 of the Supporting Information). Use of geochemical approaches to distinguish sources of salinity in near-coastal aquifers[10,51] and parts of the Central Valley[16] indicate that groundwater impacted by oil-field brines generally have lower molar ratios of chloride to iodide (Cl/I) than groundwater impacted by mixing with seawater. Most old groundwater with benzene detections has Cl/I ratios more similar to southern California oil-field brines than seawater (Figure 3). Old groundwater samples with benzene detections that plot along hypothetical mixing lines between oil-field brine and native nonsaline groundwater samples are located <5 km from an oil or gas field. In contrast, recent groundwater samples with benzene detections have lower Cl and higher Cl/I, consistent with native groundwater (Figure 3). A few samples of old groundwater with benzene detections had Cl and Cl/I that are consistent with mixing of three end-members: oil-field brines, seawater, and native fresh groundwater. These samples were located in a variety of environments. The processes explaining hydrocarbon presence in these samples may be complex but could include mobilization of hydrocarbons from organic-rich sediments[52−54] and inhibition of hydrocarbon degradation in brackish or saline environments[55,56] that may be regional groundwater discharge areas.

The three highest benzene concentrations of 2.8, 7.0, and 79.8 μg/L detected in the GAMA data occurred in old and reducing groundwater from wells located in the southwestern San Joaquin Valley. These wells were located <5 km from mapped oil and gas fields.

3.2.2. CDPH Data. Relations between benzene concentrations and explanatory variables identified from analysis of the CDPH data are congruent with the relations evident from the GAMA data. The higher LRLs and low frequency of occurrence of hydrocarbons in the CDPH data diminishes their use to discern relations to explanatory variables. Moreover, some important variables such as age tracers and perforation depth were infrequent or absent from the CDPH data. Nevertheless, significant relations of benzene to pH, nitrate, and oil and gas field proximity in the CDPH data were consistent with relations of hydrocarbons to geogenic petroleum sources identified from analysis of the GAMA data.

The positive correlation of pH and benzene in the CDPH data (Table 2) is consistent with greater occurrence of benzene in old and deep water than in other groundwater. Further analysis indicated that the relation between benzene and pH is bimodal, with a significantly different detection frequency for waters with pH $\geq$ 8 of 0.27% and pH < 7 of 0.17%, and pH of 7−7.9 of 0.06% (contingency table test $\chi^2 = 7.5794$, $p = 0.0226$, part A of Figure SI-6 of the Supporting Information). Usually pH increases from recharge at the water table to greater depths in a groundwater flow system with increasing residence time as a result of interactions of groundwater with aquifer materials, particularly mineral dissolution.[57,58] These expected relations were evident in the GAMA data, with pH increasing with groundwater age (negative correlation with ^{3}H and ^{14}C) and well depth (positive correlation) (Table SI-5 of the Supporting Information). These relations suggest that increasing pH is a surrogate for increasing groundwater age and depth in the CDPH database. Benzene is also significantly negatively correlated with nitrate concentration in the CDPH data, consistent with greater occurrence of benzene in reducing groundwater, as identified from GAMA data.

Hydrocarbon concentrations were significantly negatively correlated with distance to oil and gas fields ($\rho = 0.0175$, $p = 0.0477$), and the detection frequency of benzene in wells <5 km from oil and gas fields of 0.30% was significantly higher (contingency table test $\chi^2 = 5.0402$, $p = 0.0248$) than the 0.09% in wells $\geq$5 km away (part B of Figure SI-6 of the Supporting Information). This relation to oil and gas field proximity in the CDPH data is consistent with observations using the GAMA data.

4. DISCUSSION AND CONCLUSIONS

This study assesses the sources of benzene and other hydrocarbons present in aquifers used for public supply in California. Analysis of groundwater data sets yields evidence that geogenic sources likely explain many of the detections of hydrocarbons in public-supply wells in California. Previous large-scale assessments of hydrocarbons in groundwater have

focused on shallow groundwater more likely to be impacted by near-surface anthropogenic sources than deeper aquifers used for public supply.[1−3,40] The results of this study augment these previous studies, which indicated the importance of anthropogenic sources of hydrocarbons as a source of contamination of shallow groundwater. Although the variables explaining their occurrence are not the subject of this article, a minority of detections of hydrocarbons in public-supply wells occur in recent and shallow groundwater, and these detections are likely to reflect effects of surficial anthropogenic sources. The results of this study deepen the understanding of the occurrence of hydrocarbons in groundwater within a 3D context that includes the concept that geogenic petroleum in deep and old groundwater may represent sources of contamination in addition to anthropogenic sources at the land surface.

There are multiple lines of evidence suggesting some detections of hydrocarbons in California public-supply aquifers are related to geogenic rather than anthropogenic sources. First, benzene and hydrocarbon concentrations in the GAMA data are positively correlated with well depth and groundwater age and have higher detection frequencies in old (^{3}H < 1 pCi/L) groundwater and deep wells (well depth >180 m) compared with other samples. Second, benzene concentrations and detection frequencies in old groundwater are correlated with proximity to oil or gas fields, and the sum of hydrocarbon concentrations and detection frequencies showed similar correlations. Third, salinity and Cl/I ratios in samples with detections of benzene and hydrocarbons in old groundwater indicate interaction with oil-field brines. Fourth, in the CDPH data, increasing benzene concentrations and detection frequencies are related to increasing pH, which is a surrogate for increasing groundwater age and depth (with limited CDPH data).

The mechanisms by which geogenic benzene occurs in old groundwater are difficult to discern. Hydrocarbons could be transported by advective/dispersive processes from geogenic sources to wells in aquifers used for public supply. Alternatively, wells drilled to the depth of petroleum fields could provide vertical preferential pathways for contaminant movement to public-supply aquifers. Anthropogenic influences could include poorly completed, damaged, or abandoned hydrocarbon or water wells that allow vertical preferential flow of natural gas or hydrocarbon-bearing fluids between stratigraphic layers.[9] In addition, changes in aquifer pressures as a result of oil-field engineering and recovery operations or fresh groundwater withdrawals for water supply could induce movement of brine and geogenic hydrocarbons to supply wells.[10,12] The data available in this study are not sufficient to evaluate the relative importance of these processes.

The presence of hydrocarbons in old groundwater that is not in proximity to oil and gas fields could reflect other sources of geogenic hydrocarbons or contamination of monitoring wells during drilling or well installation. The oil or gas field data map economically productive fields. Noneconomic or unknown geogenic petroleum sources could occur and contribute hydrocarbons to old groundwater. Benzene was detected in some monitoring wells in methanogenic, high salinity, playa, or coastal wetland environments in which mobilization of some hydrocarbons from organic matter in the subsurface is possible.[52−54] Detections of benzene in monitoring wells could also result from contamination introduced during drilling and well installation.

The higher detection frequency of hydrocarbons in brackish groundwater relative to other settings could reflect several factors. Old groundwater containing hydrocarbons may be brackish because brines are commonly associated with petroleum reservoirs.[12,15] High salinity may inhibit microbial activity in some environments;[55,56] however, the effect of salinity on hydrocarbon degradation rates depends upon many variables including whether indigenous microbial communities have adapted to the brackish environment.[59−61]

Occurrence of hydrocarbons in California groundwater used for public supply is relatively infrequent and is more commonly related to geogenic rather than anthropogenic sources. Detection frequencies of benzene are highest in brackish, old, reducing (13.0%) groundwater located <5 km from oil and gas fields (20.0%); detection frequencies are also high in nonsaline, old, deep, and reducing (8.2%) groundwater located <5 km of oil and gas fields (15.4%). The hydrocarbons detected in these environments are likely to be geogenic. The relative importance of geogenic sources of hydrocarbons to the environment in California indicated by this study is consistent with air quality studies in Southern California, which have noted that geogenic emissions from oil-field seeps are important sources of methane and hydrocarbons to the atmosphere.[62,63] In comparison with detections in old, deep, and reducing groundwater, the detection frequency among GAMA samples of recent age was 1.1%; among GAMA samples of recent, shallow, and reducing groundwater, the detection frequency was 2.6% (Table SI-4 of the Supporting Information). Detections in the latter environment are likely to be anthropogenic. As a fraction of all benzene detections in the GAMA data, geogenic sources account for about 45% and anthropogenic sources may account for 27% of benzene detections; the remaining detections (28%) may be either geogenic, anthropogenic, or reflect a mixture of sources.

■ ASSOCIATED CONTENT

Supporting Information

Additional information as noted in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: landon@usgs.gov, phone: 619-225-6109, fax: 619-225-6101.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was funded by the U.S. Environmental Protection Agency, Office of Research and Development under interagency agreement DW-14-95779601-0. This study was coordinated with the California GAMA program, funded by State bonds administered by the California State Water Resources Control Board; we thank the well owners who allowed the USGS to collect samples and personnel who collected and managed the data. The use of brand names in the manuscript is for identification purposes only and does not imply endorsement by the U.S. Government.

■ REFERENCES

(1) Rice, D. W.; Grose, R. D.; Michaelsen, J. C.; Dooher, B.; MacQueen, D. H.; Cullen, S. J.; Kastenburg, W. E.; Everett, L. G.; Marino, M. A. *California Leaking Underground Fuel Tank (LUFT) Historical Case Analyses*; UCRL-AR-122207; Lawrence Livermore

National Laboratory, submitted to the California State Water Resources Control Board, 1995.

(2) Shih, T.; Rong, Y.; Harmon, T.; Suffet, M. Evaluation of the impact of fuel hydrocarbons and oxygenates on groundwater resources. *Environ. Sci. Technol.* **2004**, *38* (1), 42−48.

(3) Zogorski, J. S.; Carter, J. M.; Ivahnenko, Tamara; Lapham, W. W.; Moran, M. J.; Rowe, B. L.; Squillace, P. J.; Toccalino, P. L. *Volatile Organic Compounds in the Nation's Ground Water and Drinking-Water Supply Wells*; U.S. Geological Survey Circular 1292; U.S. Geological Survey, Reston, VA, 2006.

(4) Hadley, P. W.; Armstrong, R. Where's the benzene? − Examining California ground-water quality surveys. *Ground Water* **1991**, *29* (1), 35−40.

(5) Williams, P.; Benton, L.; Warmerdam, J.; Sheehan, P. Comparative risk analysis of six volatile organic compounds in California drinking water. *Environ. Sci. Technol.* **2002**, *36* (1), 74−81.

(6) U.S. Environmental Protection Agency. *Underground Storage Tanks*. 2011. Website accessed December 8, 2011 at: http://www.epa.gov/OUST/index.htm.

(7) Adams, L. S.; Horton, M. B. *Investigation of Birth Defects and Community Exposures in Kettleman City, CA*. California Environmental Protection Agency and California Department of Public Health: Sacramento, CA, December 2010.

(8) U.S. Geological Survey. *World Petroleum Assessment 2000*; Fact Sheet FS-062−03; U.S. Geological Survey: Reston, VA, 2003. Accessed December 1, 2011 at: http://pubs.usgs.gov/fs/fs-062-03/.

(9) Chilingar, G. V.; Endres, B. Environmental hazards posed by the Los Angeles Basin urban oilfields: An historical perspective of lessons learned. *Environ. Geol.* **2005**, *47*, 302−317, DOI: 10.1007/s00254-004-1159-0.

(10) Izbicki, J. A.; Christensen, A. H.; Newhouse, M. W.; Aiken, G. R. Inorganic, isotopic, and organic composition of high-chloride water from wells in a coastal southern California aquifer. *Appl. Geochem.* **2005**, *20*, 1496−1517.

(11) Edet, A. Characterization of groundwater contamination using factor analysis in the Niger Delta (Nigeria). *J. Hydrol. Engineering* **2009**, *14* (11), 1255−1261.

(12) Ma, J.; Pan, F.; He, J.; Chen, L.; Fu, S.; Jia, B. Petroleum pollution and evolution of water quality in the Malian River Basin of the Longdong Loess Plateau, Northwestern China. *Environ. Earth Sci.* **2011**, *66*, 7, 1769−1782 .

(13) Wang, Z.; Wu, Q.; Zhang, Y.; Cheng, J. Confined groundwater pollution mechanism and vulnerability assessment in oilfields, North China. *Environ. Earth Sci.* **2011**, *64*, 1547−1553.

(14) Lafargue, E.; Barker, C. Effect of water washing on crude-oil compositions. *AAPG Bull.* **1988**, *72*, 263−276.

(15) McCormack, P.; Jones, P.; Hetheridge, M. J.; Rowland, S. J. Analysis of oilfield produced waters and production chemicals by electrospray ionization multi-stage mass spectrometry. *Water Res.* **2001**, *35* (15), 3567−3578.

(16) Izbicki, J. A.; Metzger, L. F.; McPherson, K. R.; Everett, R. R.; Bennett, G. L. *Sources of High-Chloride Water to Wells, Eastern San Joaquin Ground-Water Subbasin, California*; Open-File Report 2006−1309; U.S. Geological Survey: Reston, VA, 2006. Available online at: http://pubs.er.usgs.gov/usgspubs/ofr/ofr20061309.

(17) Belitz, K.; Dubrovsky, N. M.; Burow, K. R.; Jurgens, B. C.; Johnson, T. D. *Framework for a ground-water quality monitoring and assessment program for California*; Water-Resources Investigations Report 03−4166; U.S. Geological Survey: Reston, VA, 2003.

(18) California Division of Oil, Gas, and Geothermal Resources. *Oil and Gas Fields*; California Department of Conservation: Sacramento, CA, 2006.

(19) California State Water Resources Control Board. *GAMA—Groundwater Ambient Monitoring and Assessment Program*; 2011, Website, accessed July 29, 2011 at: http://www.swrcb.ca.gov/gama.

(20) U.S. Geological Survey. *California Groundwater Ambient Monitoring and Assessment Program, Priority Basin Project, Publications*; 2011, Website accessed December 8, 2011 at: http://ca.water.usgs.gov/gama/.

(21) California Department of Public Health, *Drinking Water Program*; 2011, Website accessed September 7, 2011 at: http://www.cdph.ca.gov/programs/Pages/DWP.aspx.

(22) Koterba, M. T.; Wilde, F. D.; Lapham, W. W. *Ground-water data-collection protocols and procedures for the National Water-Quality Assessment Program—collection and documentation of water-quality samples and related data*; Open-File Report 95−399; U.S. Geological Survey: Reston, VA, 1995.

(23) U.S. Geological Survey. *National field manual for the collection of water-quality data*; Techniques of Water-Resources Investigations Book 9, ch. A1-A9; U.S. Geological Survey, Reston, VA, variously dated. Accessed online 10 Jun 2009 at http://pubs.water.usgs.gov/twri9A.

(24) Connor, B. F.; Rose, D. L.; Noriega, M. C.; Murtagh, L. K.; Abney, S. R. *Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—Determination of 86 volatile organic compounds in water gas chromatography/mass spectrometry, including detections less than reporting limits*; Open-File Report 97−829; U.S. Geological Survey: Reston, VA,1998.

(25) Landon, M. K.; Belitz, K.; Jurgens, B. C.; Kulongoski, J. T.; Johnson, T. D. *Status and understanding of groundwater quality in the Central−Eastside San Joaquin Basin, 2006: California GAMA Priority Basin project*; Scientific Investigations Report 2009−5266; U.S. Geological Survey: Reston, VA, 2010.

(26) Kulongoski, J. T.; Belitz, K.; Landon, M. K.; Farrar, C. *Status and understanding of groundwater quality in the North San Francisco Bay groundwater basins, 2004—California GAMA Priority Basin Project*; Scientific Investigations Report 2010−5089; U.S. Geological Survey: Reston, VA, 2010.

(27) Bennett, G. L.; Fram, M. S.; Belitz, K. *Status and understanding of groundwater quality in the Southern, Middle, and Northern Sacramento Valley study units, 2005−08—California GAMA Priority Basin Project*; Scientific Investigations Report 2011−5002; U.S. Geological Survey: Reston, VA, 2011.

(28) Burton, C. A.; Montrella, J.; Landon, M. K.; Belitz, K. *Status and understanding of groundwater quality in the Santa Clara River Valley, 2007—California GAMA Priority Basin Project*; Scientific Investigations Report 2011−5052; U.S. Geological Survey: Reston, VA, 2011.

(29) Bender, D. A.; Zogorski, J. S.; Halde, M. J.; Rowe, B. L. *Selection procedure and salient information for volatile organic compounds emphasized in the National Water-Quality Assessment Program*; Open-File Report 99−182; U.S. Geological Survey: Reston, VA, 1999.

(30) Fram, M. S.; Olsen, L. D.; Belitz, K. *Review of Volatile Organic Compound Blank Data Collected for the California Groundwater Ambient Monitoring and Assessment (GAMA) Program, May 2004 − October 2009*; Scientific Investigations Report 2012−5139; U.S. Geological Survey: Reston, VA, in press.

(31) California State Water Resources Control Board. *Geographic Environmental Information Management System GeoTracker (GEIMS) Leaking Underground Fuel/Storage Tank database (LUFT) [digital data]*; 2011, Website, accessed November 18, 2011 at: http://geotracker.waterboards.ca.gov/data_download.asp.

(32) California Department of Finance, Demographic Research Unit. *1970−1980−1990−2000 Comparability File, digital dataset of US CENSUS data normalized to the 1990 tract boundaries for the State of California*; 2000, Website accessed April 19, 2004 at: http://www.dof.ca.gov/html/demograp/scdc%5Fproducts.htm.

(33) Nakagaki, N.; Price, C. V.; Falcone, J. A.; Hitt, K. J.; Ruddy, B. C. *Enhanced National Land Cover Data 1992 (NLCDe 92)*; U.S. Geological Survey: Reston, VA, 2007. Raster digital data, available online at: http://water.usgs.gov/lookup/getspatial?nlcde92.

(34) U.S. Department of Agriculture (USDA). *STATSGO, state soil survey geographic data base ARC/INFO 7.0 coverage, UNIX*; USDA: Forth Worth, TX, 1994.[CD-ROM].

(35) Jurgens, B. C.; Fram, M. S.; Belitz, K.; Burow, K. R.; Landon, M. K. Effects of Groundwater Development on Uranium: Central Valley, California, USA. *Ground Water* **2010**, *48*, 913−928, DOI: 10.1111/j.1745-6584.2009.00635.x.

(36) Johnson, T. D.; Belitz, K. Assigning land use to supply wells for the statistical characterization of regional groundwater quality: Correlating urban land use and VOC occurrence. *J. Hydrol.* **2009**, *370*, 100−108.
(37) Wright, M. T.; Belitz, K. Factors Controlling the Regional Distribution of Vanadium in Groundwater. *Ground Water* **2010**, *48*, 515−525.
(38) Fram, M. S.; Belitz, K. Probability of detecting perchlorate under natural conditions in deep groundwater in California and the Southwestern United States. *Environ. Sci. Technol.* **2011**, *45*, 1271−1277, DOI: 10.1021/es103103p.
(39) Landon, M. K.; Green, C. T.; Belitz, K.; Singleton, M. J.; Esser, B. K. Relations of hydrogeologic factors, groundwater reduction-oxidation conditions, and temporal and spatial distributions of nitrate, Central-Eastside San Joaquin Valley, California. *Hydrogeology J.* **2011**, *19*, 1203−1224.
(40) Moran, M. J.; Zogorski, J. S.; Squillace, P. J. MTBE and gasoline hydrocarbons in ground water of the United States. *Ground Water* **2005**, *43* (4), 615−627.
(41) Michel, R. L. *Tritium deposition in the continental United States, 1953−83*; Water-Resources Investigations Report 89−4072; U.S. Geological Survey: Reston, VA, 1989.
(42) Plummer, L. N.; Michel, R. L.; Thurman, E. M.; Glynn, P. D. 1993, Environmental tracers for age-dating young groundwater. In *Regional Groundwater Quality*; Alley, W. M., Ed.; Van Nostrand Reinhold, New York, p 255−294.
(43) Michel, R.; Schroeder, R. Use of long-term tritium records from the Colorado River to determine timescales for hydrologic processes associated with irrigation in the Imperial Valley, California. *Appl. Geochem.* **1994**, *9*, 387−401.
(44) Clark, I. D.; Fritz, P. *Environmental Isotopes in Hydrogeology*; Lewis Publishers: New York, 1997.
(45) Manning, A. H.; Solomon, D. K.; Thiros, S. A. 3H/3He age data in assessing the susceptibility of wells to contamination. *Ground Water* **2005**, *43* (3), 353−367.
(46) McMahon, P. B.; Chapelle, F. H. Redox processes and water quality of selected principal aquifer systems. *Ground Water* **2008**, *46* (2), 259−271.
(47) Jurgens, B. C.; McMahon, P. B.; Chapelle, F. H.; Eberts, S. M. *An Excel Workbook for Identifying Redox Processes in Ground Water*; Open-File Report 2009−1004; U.S. Geological Survey: Reston, VA, 2009.
(48) Helsel, D. R.; Hirsch, R. M. *Statistical Methods in Water Resources*; Techniques of Water-Resources Investigations, book 4, chap. A3; U.S. Geological Survey: Reston, VA, 2002; Available at http://water.usgs.gov/pubs/twri/twri4a3/
(49) Conover, W. J. *Practical Nonparametric Statistics*, 2nd ed.; Wiley, New York, 1980.
(50) Belitz, K.; Jurgens, B. C.; Landon, M. K.; Fram, M. S.; Johnson, T. D. Estimation of aquifer scale proportion using equal area grids: Assessment of regional scale groundwater quality. *Water Resour. Res.* **2010**, *46*, W11550 DOI: 10.1029/2010WR009321.
(51) Hanson, R. T.; Izbicki, J. A.; Reichard, E. G.; Edwards, B. D.; Land, M.; Martin, P. Comparison of groundwater flow in Southern California coastal aquifers. *In Earth Science in the Urban Ocean: The Southern California Continental Borderland*; Lee, H. J., Normark, W. R., Eds.; Geological Society of America Special Paper 454, **2009**; DOI: 10.1130/2009.2454(5.3).
(52) Devai, I.; Delaune, R. D. Light hydrocarbon production in freshwater marsh soil as influenced by soil redox conditions. *Water, Air, and Soil Pollut.* **1996**, *88*, 39−46.
(53) Redeker, K. R.; Meinardi, S.; Blake, D.; Sass, R. Gaseous emissions from flooded rice paddy agriculture. *J. Geophys. Res.* **2003**, *108* (D13), 4386 DOI: 10.1029/2002JD002814.
(54) White, M. L.; Russo, R. S.; Zhou, Y.; Ambrose, J. L.; Haase, K.; Frinak, E. K.; Varner, R. K.; Wingenter, O. W.; Mao, H.; Talbot, R.; Sive, B. C. Are biogenic emissions a significant source of summertime atmospheric toluene in the rural Northeastern United States? *Atmos. Chem. Phys.* **2009**, *9*, 81−92.
(55) Ward, D. M.; Brock, T. D. Hydrocarbon degradation in hypersaline environments. *Appl. Environ. Microbiol.* **1978**, *35*, 353−359.
(56) Mille, G.; Almallah, M.; Bianchi, M.; Wambeke, F.; van, Bertrand, J. C. Effect of salinity on petroleum biodegradation. *Fresenius' J. Anal. Chem.* **1991**, *339*, 788−791.
(57) Koh, D. C.; Plummer, L. N.; Solomon, D. K.; Busenberg, E.; Kim, Y. J.; Chang, H. W. Application of environmental tracers to mixing, evolution, and nitrate contamination of ground water in Jeju Island, Korea. *J. Hydrol.* **2006**, *327* (1−2), 258−275.
(58) Jurgens, B. C.; Burow, K. R.; Dalgish, B. A.; Shelton, J. L. *Hydrogeology, Water Chemistry, and Factors Affecting the Transport of Contaminants in the Zone of Contribution of a Public-Supply Well in Modesto, Eastern San Joaquin Valley, California*; Scientific Investigations Report 2008−5156; U.S. Geological Survey: Reston, VA, 2008. Available at http://pubs.usgs.gov/sir/2008/5156.
(59) Leahy, J. G.; Colwell, R. R. Microbial degradation of hydrocarbons in the environment. *Microbiol. Rev.* **1990**, *54* (3), 305−315.
(60) Margesin, R.; Schinner, F. Biodegradation and bioremediation of hydrocarbons in extreme environments. *Appl. Microbiol. Biotechnol.* **2001**, *56*, 650−663.
(61) Ukpaka, C. P. Biodegradation model on effect of some physiochemical parameters on aromatic compounds in fresh water medium. *J. Bacteriol. Res.* **2011**, *3* (3), 42−55.
(62) Killus, J. P.; Moore, G. E. Factor analysis of hydrocarbon species in the south-central coast air basin. *J. Appl. Meteorol.* **1991**, *30*, 733−743.
(63) Bradley, E.; Leifer, I.; Roberts, D. Long-term monitoring of a marine geologic hydrocarbon source by a coastal air pollution station in Southern California. *Atmos. Environ.* **2010**, *44* (38), 4973−4981.