UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IMPACTS OF WELL PUMPING ON TRACE METALS IN THE CENTRAL OKLAHOMA AQUIFER, NORMAN, OK

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

ZACHARY D. TOMLINSON

Norman, Oklahoma

2021

IMPACTS OF WELL PUMPING ON TRACE METALS IN THE CENTRAL OKLAHOMA AQUIFER, NORMAN, OK

A THESIS APPROVED FOR THE
SCHOOL OF GEOSCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Andy Elwood Madden, Chair

Dr. Megan Elwood Madden

Dr. Kato Dee

Acknowledgement

Firstly, I would like to thank Rachel Croft, Andrea Laws, Neal Engleman, Chris Adams, Scott Aynes, and Geri Wellborn for their help collecting and sending the data that I used in this study and for answering all questions I had related to it. Scott and Geri were also instrumental in coordinating our field work with sampling City of Norman wells. Next, I would like to thank my committee members Dr. Megan Elwood Madden and Dr. Kato Dee for their advice on my thesis and for teaching nearly all of the foundational geochemistry and hydrology classes I had over the course of my master's program at OU. Dr. Kato Dee also loaned equipment, experience, and time to help with the field work discussed in this study. My field partner, Alexa Muntz, also helped tremendously with the field work. Brandon Maples's work with USGS percent data allowed me to get percent sand estimates for wells of interest. I would also like to thank my advisor, Dr. Andy Elwood Madden, for all of his support, patience, and flexibility in allowing me to pursue all of the twists and turns of this project and really make it my own.

Table of Contents

Table of Figures

Table of Tables

Abstract

Due to increasing global demand for fresh water, it is increasingly necessary to understand how aquifer pumping affects groundwater chemistry. Water quality may be affected by mobilizing naturally occurring salts or trace metals via water-rock interactions, mixing with deeper saline groundwaters, increased recharge containing anthropogenic inputs, and other site-specific considerations. Comprehensive predictive relationships between pumping and groundwater quality have yet to be developed. Pumping in the Central Oklahoma Aquifer (COA) has increased drastically from the 1980's, especially around Norman, OK. Studying the changing geochemistry around Norman, OK over the last 40 years in response to changing pumping conditions provides unique insights into how the COA will respond to increased pumping over the next few decades. Data from the City of Norman, the Oklahoma Water Resources Board (OWRB) and the U. S. Geological Survey National Water Information System (NWIS) suggest that changes in pumping rates coincide with sometimes drastic changes in trace metal concentrations, which vary by well, time interval, and how heavily the well is used. The observed trends are most consistent with low pH groundwater flow down wells during inactive periods, a phenomenon already seen in the High Plains Aquifer. This phenomenon can only occur soon after a well has been pumped, while hydraulic gradients facilitate downward flow. If the downward flow (and pH decline) initiated by pumping outweigh the pH increase over the duration of pumping, pH drops overall, leading to trace metal sorption and a decline in trace metal concentrations. If pumping outweighs the effects of the downward flow, pH rises overall and trace metals mobilize. Water volume and pumping frequency also dictate whether inactive periods with downward groundwater flow (producing more negative correlations with pumping) outweigh the active pumping periods (producing more positive correlations with pumping). Inverse modeling in PHREEQC shows that pumping does not significantly affect groundwater chemistry by inducing water mixing/ dilution with a deep Na-Cl brine; it also predicts dolomite dissolution only during modeled time intervals with trace metal decreases and dolomite precipitation only during modeled time intervals with trace metal increases. Lastly, this thesis uses the identified correlations and PHREEQC trends to construct hazard maps of locations most and least likely to see degrading water quality in response to pumping.

Introduction

As global population and living standards increase, global demand for fresh water increases (Mekonnen and Hoekstra, 2016). Groundwater specifically supplies the domestic water for about half of the global population (Rodell et al., 2018). Increased groundwater use takes water out of aquifer storage, increasing the size of cones of depression around pumping wells until either the system reaches a new equilibrium or the aquifer is depleted (de Graaf et al., 2019). Because municipal supply wells have higher pumping rates than domestic wells, they play an especially important role in shaping regional water tables. Municipal supply wells also affect more people. Due to increasing water demand and the importance of groundwater to meet this demand, cones of depression, especially around municipal wells, play a substantial role in many aspects of groundwater studies, a role that will increase with future increases in water demand.

Groundwater pumping (and associated changes in aquifer conditions) has been clearly shown to affect both major ion and trace element geochemistry in an array of aquifer systems (Ayotte et al., 2015, 2011; Brown et al., 2002; Xing et al., 2013). Groundwater quality may influence suitability for public supply due to the presence of anthropogenic contaminants, but also due to naturally occurring salinity and trace metals. Understanding how trace elements such as As, B, Cr, Mn, Se, and U specifically are affected by groundwater pumping is vital due to their effects on human health. Naturally occurring trace metal ingestion (especially arsenic) due to increased groundwater use in the twentieth century has already been linked to outbreaks of keratosis, blackfoot disease, and various cancers (Mitchell et al., 2011; Morales et al., 2000). Thus, understanding how trace metal concentrations in groundwater respond to changing pumping conditions are critical and will continue to be so.

Previous literature

Despite being well-documented in many aquifer systems, the interactions between pumping rates and groundwater geochemistry are very aquifer-specific; an array of results have been reported. Studies of aquifers in the North China Plain, New England, California, Wisconsin, New Jersey, and New York have shown that pumping induces mixing and downward flow, which changes geochemical parameters (such as redox conditions, dissolved oxygen, pH, etc.) and then increases trace element concentrations through dissolution of aquifer minerals (Ayotte et al., 2015, 2011; Brown et al., 2002; Xing et al., 2013). Generally, these studies demonstrate that increased pumping lowers the mixing zone between low pH meteoric water and high pH groundwater, causing previously equilibrated aquifer minerals to dissolve and potentially release trace metals (Fig. 1). Additionally, samples taken from a fixed, shallow depth in the aquifer would have lower pH and increased dissolved oxygen under higher pumping conditions due to the lowered mixing zone.

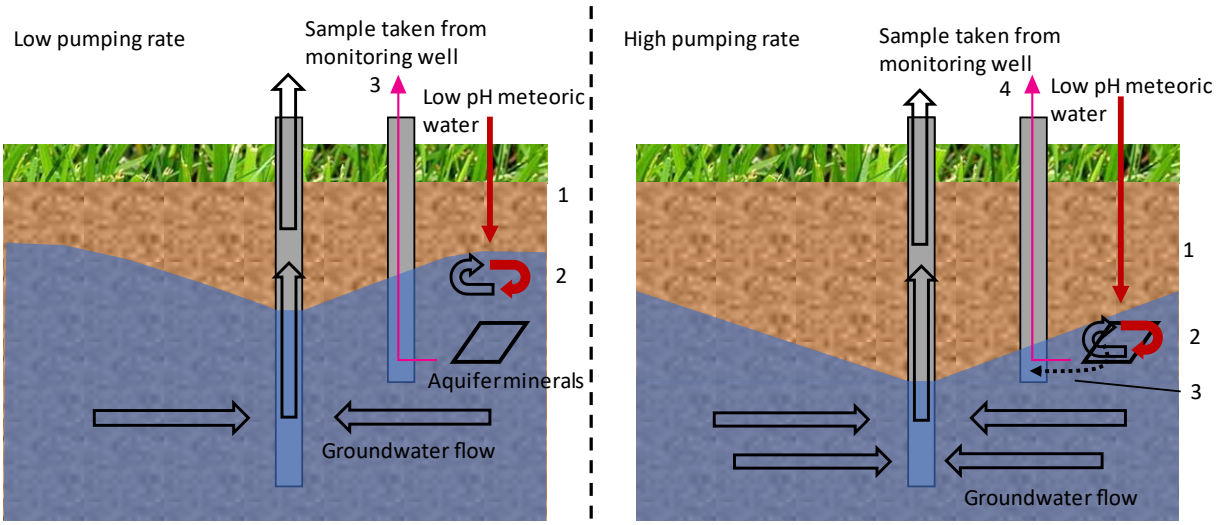


Figure 1: Generalized conceptual diagram for trace element mobilization and changing sample chemistry under increased pumping rates

1. Oxygenated, low pH meteoric water infiltrates
2. The meteoric water mixes with higher pH groundwater near the water table
3. Groundwater is sampled for analysis from a specified depth

1. Oxygenated, low pH meteoric water infiltrates
2. The meteoric water mixes with higher pH groundwater near the water table
3. Minerals that were previously equilibrated with the high pH groundwater dissolve under a lower pH environment, mobilizing any trace metals
4. Groundwater is sampled for analysis from the same specified depth

Other sources have found connections between pumping and trace metal mobilization even in deep confined aquifers, such as the High Plains Aquifer. Ayotte et al. (2011) compared trace metal concentrations in shallow monitoring wells, deep monitoring wells, and deep, depth dependent samples from public supply wells in the High Plains aquifer near York, NE; they found that public supply wells had the highest concentrations of Mo, Co, Fe, Mn, and U, rather than the intermediate concentrations that would be expected with groundwater mixing. They reasoned that pumping facilitates the downward flow of low pH, oxygenated water through the well and into the otherwise sub-oxic deeper aquifer layers during non-pumping conditions, which helps mobilize some trace elements by oxidizing them and/or dissolving material normally stable under sub-oxic to anoxic conditions (Fig. 2, Ayotte et al., 2011). Although it seems counterintuitive that shallow, lower TDS groundwater could flow downward through a well, large downward hydraulic gradients around pumping wells can facilitate downward flow (Clark et al., 2008).

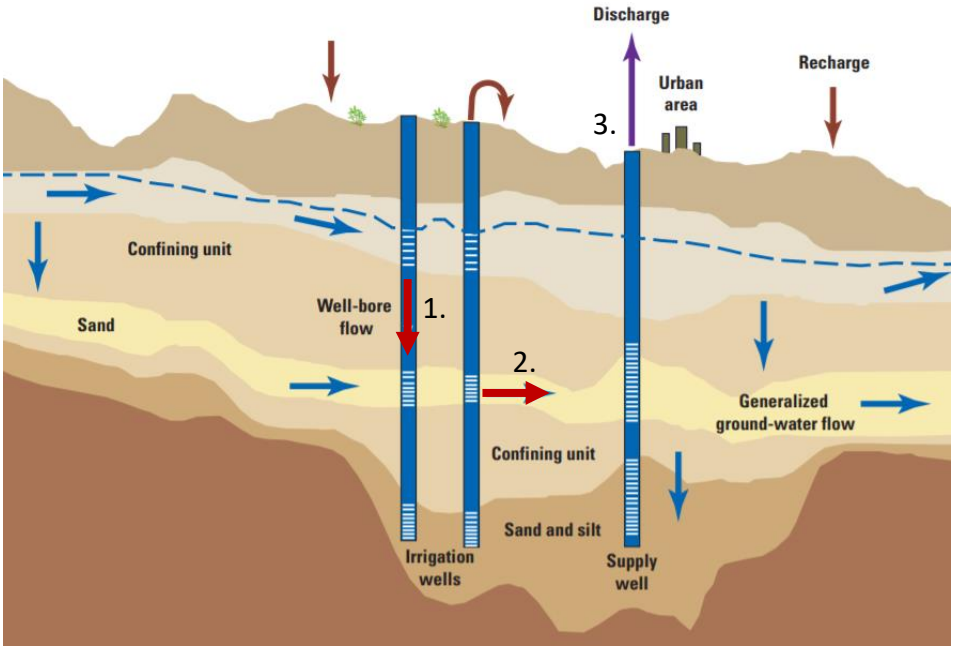


Figure 2: Groundwater flow model for the High Pains Aquifer, York, NE. Adapted from Clark et al. (2008). Red arrows indicate groundwater flow-paths that only exist with recent pumping conditions and include the following steps: (1) high hydraulic gradients near the pumping well facilitate downward flow in the well during inactive periods, (2) oxygenated, low pH groundwater makes its way along the normal groundwater flow path, oxidizing and/or mobilizing trace metals, and (3) the newly modified groundwater enters the public supply well.

Ayotte et al. (2011) proposed a similar mechanism for a correlation between pumping and trace metal concentrations in the Upper Floridian Aquifer from a single pumping well. In the Upper Floridian Aquifer, arsenic concentrations first increased and then decreased as pumping increased in the pumping well. They reasoned that as pumping increases from no pumping to moderate pumping, downward hydraulic gradients pull water from the surficial aquifer system down to the Upper Floridian Aquifer, oxidizing/dissolving arsenic-bearing pyrite and mobilizing arsenic (Fig. 3); then, as pumping rates increase more, more low-arsenic water gets pulled into the well, diluting the well-water enough to drop arsenic concentrations back down (Fig. 3, Ayotte et al., 2011). In summary, switching pumping status from off to on increased trace metal concentrations in pumping wells in the High Plains Aquifer and the Upper Floridian aquifer, while switching pumping magnitude from low to high decreased trace metal concentrations in the Upper Floridian aquifer (Ayotte et al., 2011).

Interestingly, the switch from sub-oxic conditions to oxic conditions which mobilized arsenic in the Upper Floridian Aquifer had the opposite effect in a glacial alluvial aquifer in eastern Wisconsin (Ayotte et al., 2011). Fe and As concentrations continually increased in the groundwater over 14 days after pumping conditions ceased and water conditions became

increasingly anoxic; Ayotte et al. (2011) suggest that Fe(III) reduction in As-bearing iron hydroxides dissolve the iron minerals and release both iron and arsenic during non-pumping conditions. Although pumping is generally associated with increasingly oxic conditions across all the sources discussed so far, the effect on trace metals depends highly on how they are bound in the aquifer matrix.

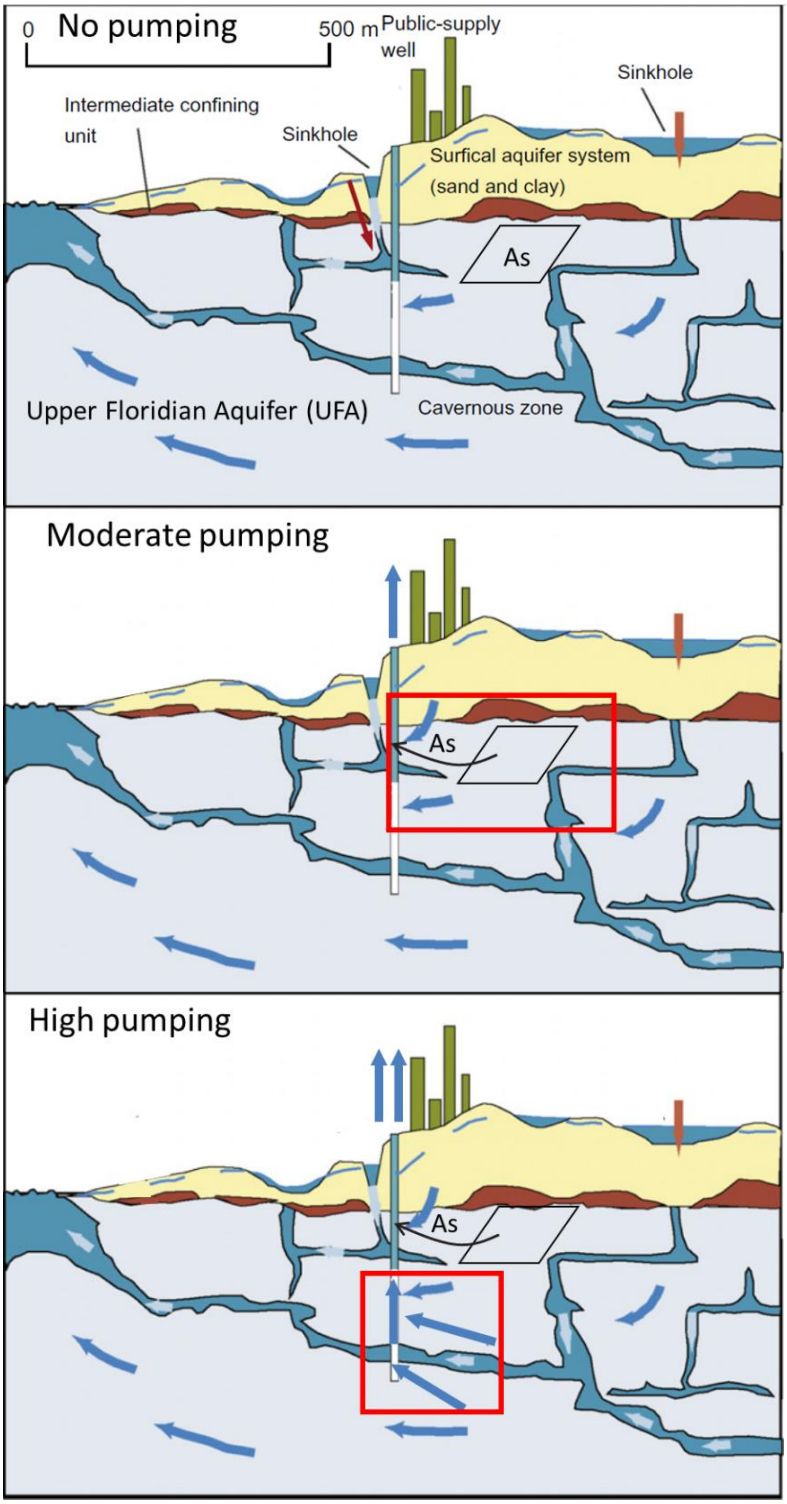


Figure 3: Conceptual diagram for observations in the Upper Floridian Aquifer (UFA), adapted from Ayotte et al. (2011). Red boxes indicate areas of greatest change in groundwater flow conditions and/or geochemistry between pumping conditions and are highlighted below.

No pumping: No or little mixing occurs between the Surficial Aquifer System (SAS) and the UFA, so As-bearing minerals in the UFA do not dissolve and As levels are low.

Moderate pumping: Mixing between the SAS and UFA dissolves As-bearing minerals and causes high As concentrations.

High pumping: Although the As is still mobilized through dissolution, increased water flow from other parts of the aquifer dilute the As increase so overall low As levels are observed (Ayotte et al., 2011).

Aquifer matrix also played a large role in the correlations between pumping and trace metal chemistry in another aquifer in Lansing, MI. Haack and Laukkanen (2013) found that changes in chemistry with well pumping in Lansing, MI varied on a well-by-well basis due to extreme heterogeneity in the aquifer. Other sources found no correlation between pumping and trace metal concentrations. Pichler et al. (2017) showed that groundwater flow along screened domestic wells facilitates mixing and subsequent dissolution to release As and Mo in Lithia, FL, but that the regional municipal pumping conditions had no effect on this flow (Fig. 4).

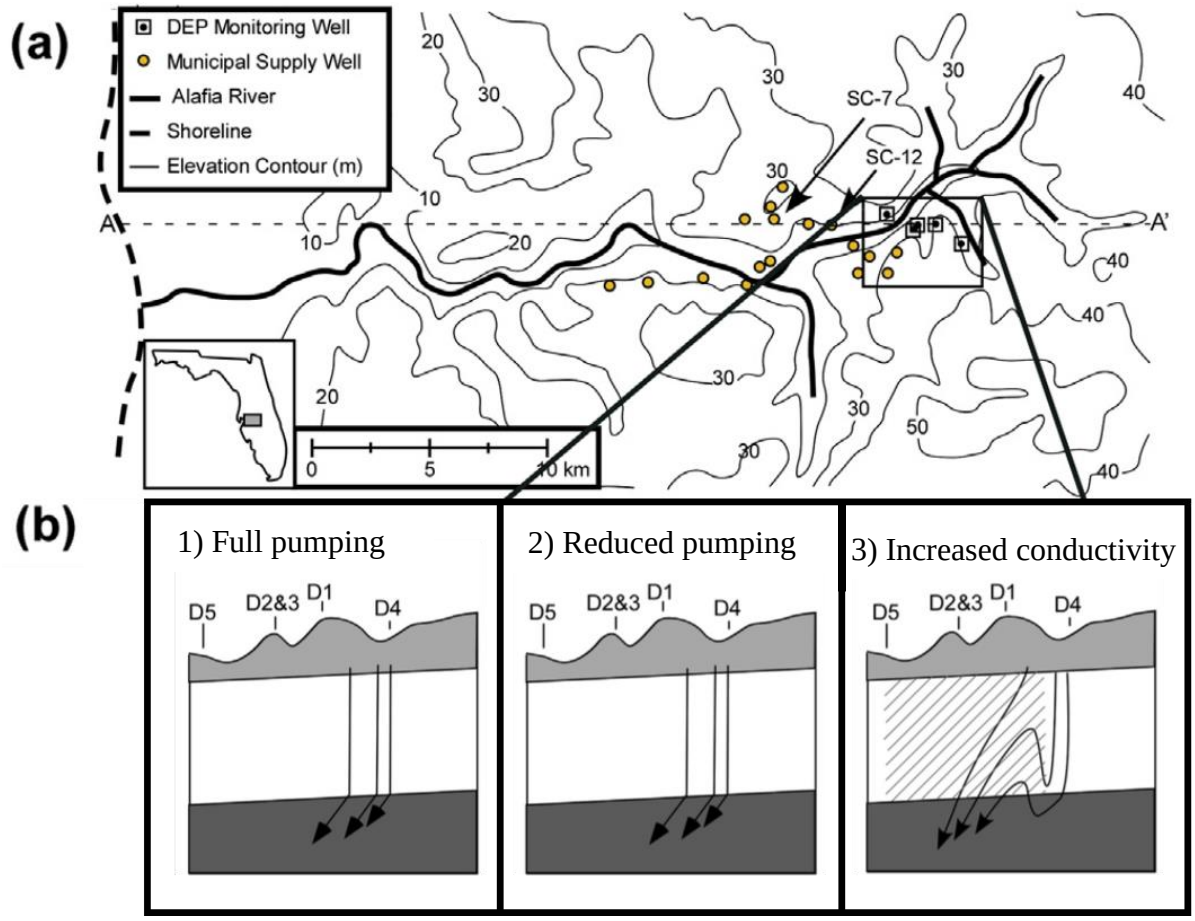


Figure 4: Modeled groundwater flow paths with varied hydrologic parameters. Part (a) shows the overall study area, where the black boxes indicate monitoring wells and the yellow points indicate municipal supply wells. Part (b) shows three different sets of model flow paths under variable model parameters for the region outlined in black from part (a). Inset 1 shows the modeled flow paths when the six nearest municipal wells are pumping at their average rates; inset 2 shows the modeled flow paths when the six nearest municipal wells are NOT pumping. Note that the models produce identical results. Inset 3 shows the modeled flow paths when the conductivity in the grey region is increased by two orders of magnitude. D1-D5 represent monitoring well clusters; each contains 4 wells screened across different depths (Pichler et al., 2017).

They found high local concentrations of As and Mo in monitoring well clusters D1-3 but low concentrations in D4 and D5 (Fig. 4a). The area around D1-D3 contains a large population of domestic wells screened across three aquifer systems: the Surficial Aquifer System (SAS), the Intermediate Aquifer System (IAS) and the Upper Florida Aquifer (IAS) (Pichler et al., 2017). Using MODFLOW to investigate the effects of increased local conductivity due to screening in the domestic wells around D1-D3 (Fig. 4b), Pichler et al. (2017) showed that the high domestic well density could make groundwater flow between all three aquifer systems, thus changing the water chemistry and releasing As and Mo.

Pichler et al. (2017) also showed that pumping from nearby municipal wells (Fig. 4a) could not account for the As and Mo spike in D1-3. D5 and D4 both contained low concentrations of As and Mo even though D5 is closest to the municipal supply wells and D4 is farthest away. Modeled groundwater flow paths also showed no change when the nearby municipal supply wells were simulated at their normal pumping conditions versus when they were simulated as off (Fig. 4b, panels 1 and 2). Thus, Pichler et al. showed that the pumping volume had no effect on As and Mo in Lithia, FL, but the presence of domestic wells increased regional conductivity enough to facilitate groundwater mixing and trace metal mobilization.

Groundwater in the Central Oklahoma Aquifer

Previous works have shown a wide array of interactions between pumping rates and trace metal concentrations. Thus, even though some relationship between trace elements and pumping rates is likely in the Central Oklahoma Aquifer (COA), previous work alone cannot predict how increased pumping in various regions will affect trace element concentrations. Trace elements are also a major concern for this aquifer. Although the water quality in the COA is generally good for young, shallow groundwater samples, deeper waters are older and more likely to contain arsenic, chromium, and selenium (Wilkins, 2019). These elements do not exceed EPA maximum contaminant levels (MCLs) until (generally) deep in the aquifer, while uranium exceeds its safe concentration sporadically throughout the aquifer (Table 1).

Notably, between 25-50% of arsenic samples and < 5% of chromium samples from deep wells exceeded their MCLs as of 1998, but no samples from shallow or intermediate wells did. Conversely, 5-25% of uranium samples from all three well depth classes exceeded its MCL (Table 1). Less than 5% of selenium samples surpassed safe concentrations for intermediate wells but 5-25% exceeded safe concentrations for deep wells (Christenson et al., 1998). Thus, trace metals in the COA are both a major concern and are largely depth dependent.

Physical Properties and Constituents	Sample Size	Minimum	Percentiles					Maximum	EPA Maximum Contaminant Level (Secondary Maximum Contaminant Level)
			5	25	50	75	95		
Shallow (< 30 m) wells									
Fluoride (mg/L)	25	<.10	.10	.20	.30	.40	2.2	2.9	4 mg/L (2 mg/L)
Uranium (ug/L)	25	<.20	.02	.21	1.1	8.6	60	69	30 ug/L
Intermediate (30 - 91 m) wells									
Fluoride (mg/L)	35	.10	.10	.30	.30	.5	2.9	3.9	4 mg/L (2 mg/L)
Selenium (ug/L)	35	<1	.05	.25	.80	2	29	75	50 ug/L
Uranium (ug/L)	35	<.20	.05	.5	1.0	7.1	130	220	30 ug/L
Deep (> 91 m) wells									
Arsenic (ug/L)	27	<1	.11	.71	2	14	86	110	10 ug/L
Chromium (ug/L)	27	<5	1.36	4.11	8	23	84	100	100 ug/L
Selenium (ug/L)	27	<1	.12	.85	3	20	120	150	50 ug/L
Uranium (ug/L)	27	1.0	1.1	2.2	11	25	120	130	30 ug/L

Table 1: Constituents exceeding EPA maximum contaminant levels in the COA by well depth; adapted from Christenson et al. (1998), and Wilkins (2019).

A more recent Oklahoma Water Resources Board (OWRB) overview of 1835 wells in the COA with over 4,300 trace metal analyses (conduced in 2011) found that 15% of arsenic, 2% of chromium, 7.5% of selenium, and 11% of uranium samples exceeded their EPA maximum contaminant levels (Wilkins, 2019)[1]. As largely shallow domestic wells, these data are more likely to underrepresent the arsenic, chromium, and selenium averages. Despite its high concentrations of naturally occurring trace metals, the COA is widely used for human use; 66% of the 40 million gallons per day of groundwater pumped from the COA is used for public water supply (Tortorelli, 2009). Norman, OK specifically pumps between 1.7 to slightly over 3.3 million gallons per day from the COA for its public water supply (City of Norman, 2018).

In addition to having high concentrations of trace elements and a high potential for human health risks, there is direct evidence that pumping influences the regional groundwater flow regime in Norman, OK. The potentiometric surface for the COA around Norman drastically changed from when it was surveyed by the OWRB in 1986-1987 to when it was surveyed in 2009 (Fig. 5). This data was collected and made public as part of a 2013 U.S. Geological Survey (USGS)/OWRB report (Mashburn et al., 2018). Not only does the center of the cone of depression move slightly SE between the 1980's and 2009, but the magnitude of depression goes from between 100-150 ft maximum to between 200 and 250 ft maximum over the 22-year period. Although the center of the 2009 cone of depression is centered around well 15, wells 20 and 13 likely contributed to its change in position and magnitude because both had generally higher pumping rates between the 1990's and 2009; well 15 went inactive in 2005. Therefore, the potentiometric surface changes reflect regional pumping conditions rather than conditions in a single well.

[1] The full dataset, titled *Assessment of Distribution of Arsenic, Chromium, Selenium, and Uranium, in Groundwater in the Garber-Wellington Aquifer – Central Oklahoma,* can be found at ArcGIS.com.

To meet an ever-increasing water demand, municipalities in central Oklahoma may need to pump more deep groundwater from the COA. Due to high concentrations of trace metals in the deep COA groundwater, it is necessary to understand how increased pumping will affect water quality in order to protect public health. Previous studies alone cannot predict how water quality will be affected by a pumping increase due to extremely variable results in different aquifers. Groundwater pumping has likely contributed to the drastic changes in hydrology over the last 40 years in Norman; therefore, this study performs new methods and analyses using historical groundwater data supplemented with additional modern sampling from around Norman to provide insights into how increased pumping will affect COA groundwater quality in the future.

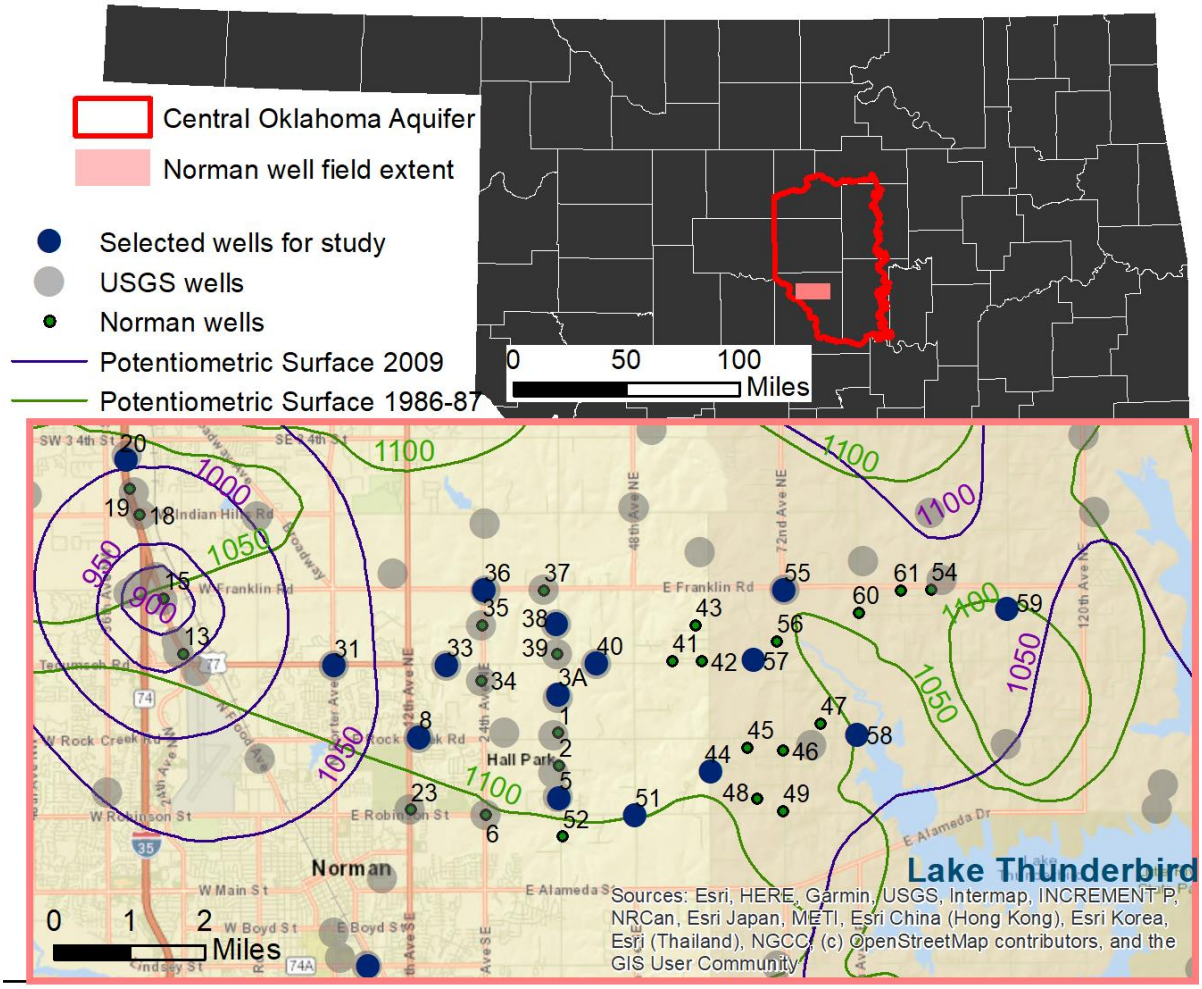


Figure 5: Potentiometric surface and well site locations for the COA around Norman, OK. The 1986-1987 potentiometric surface is featured in Christenson et al. (1998); the 2009 potentiometric surface is featured in Mashburn et al. (2018). Both include OWRB data.

Geologic Context

The COA is contained largely within the Permian-age Garber and Wellington geologic units; approximately 17% of its surface exposure is confined under the Hennessy Group along its western boundary (Fig. 6, Wilkins, 2019). Western-most wells in the City of Norman penetrate the confined portion of the aquifer (Fig. 6). The Hennessy group is dominated by shale and mudstone with some very thin, fine-grained sandstone beds (Breit, 1998). Total thickness of the aquifer units ranges from 1,100 to 1,600 ft (Smith et al., 2009). The Garber and Wellington formations have very similar lithologies and are generally undistinguished when referring to the aquifer system. Both formations contain mostly lenticular, fine-grained quartz litharenite (sandstone) interbedded with mudstone (mostly mixed-layer illite-smectite), siltstone, and small amounts of conglomerate (Breit, 1998). All three units dip approximately 0.5º west (Christenson et al., 1998) and have a red color due to hematite precipitation during early diagenesis (Breit, 1998). Sandstone dominates in the upper third of the COA while mudstone dominates in the lower third, which is shown in the cross section produced by Smith et al. (2009, Fig. 7). A full section of the Garber and Wellington formations contains between 25% and 75% sandstone (Breit, 1998), varying laterally.

The accessory minerals in the aquifer provide a more interesting story. Pits and voids in minerals from approximately 200 rock samples from eight test holes in the COA show that dolomite, calcite, chert, feldspar, and possibly chlorite have dissolved since the Garber and Wellington formations were first exposed to meteoric water (Breit, 1998). Possible barite dissolution was also noted due to embayment on crystals present in void spaces (Breit, 1998). Breit (1998) also presented evidence for considerable precipitation since exposure to meteoric water: goethite as overgrowths on hematite and as void lining; manganese oxides coating dolomite grains and with pseudo-rhombohedral structures; kaolinite as pore-fill; quartz overgrowths and calcite were all found in COA rock thin-sections. Dolomite occurrences in the COA are nearly 1 weight percent manganese (Parkhurst et al., 1996). Notably, the thin conglomerate layers contain higher concentrations of iron, arsenic, and dolomite than the other layers (Smith et al. 2009). Although the lithology of the COA is fairly consistent, the accessory minerals and textures show that a variety of chemical reactions occur in the COA.

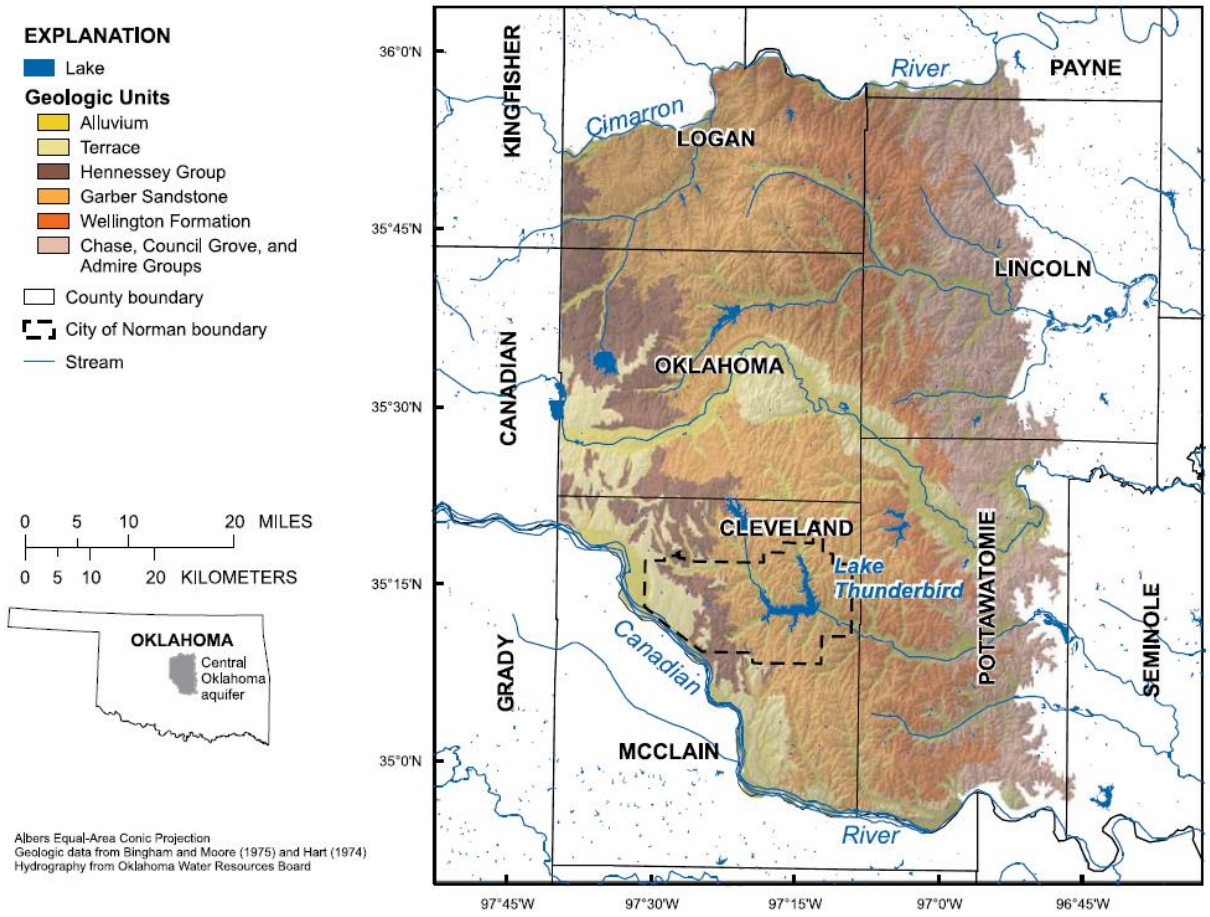


Figure 6: Map of COA surficial geology and City of Norman boundaries. Adapted from Smith et al. (2009).

Conceptual Models

Observed relationships between pumping and trace metal concentrations in other aquifers depend heavily on geochemical conditions and how trace metals are bound in the aquifer matrixes. Conceptual models for trace metal occurrences therefore provide essential frameworks for understanding changes in response to pumping. An EPA/USGS study showed that, of all the reactions occurring in the COA, three water-rock interactions dominate changes in groundwater chemistry: dolomite dissolution, ion exchange with clay, and interactions between the water column and the Na-Cl rich brine below the freshwater limit of the COA (Smith et al., 2009). These reactions also have effects on trace metal mobilization and distribution which have been discussed in literature.

Arsenic

Smith et al. (2009) discussed a conceptual model for how As mobilizes in the deep COA through ion exchange and dolomite dissolution. Ion exchange undersaturates groundwater with respect to dolomite, and subsequent dolomite dissolution drives the pH above 8.5. Arsenic in the COA is sorbed onto iron oxide mineral coatings, but it desorbs above a pH of 8.5 and enters the

groundwater. At depth there is more ion exchange due to a greater proportion of mudstone (Fig. 7) and less carbon dioxide, so pH can reach higher levels than in the shallow COA (Smith et al., 2009). Thus, there is a clear trend between elevated As levels and pH values above 8.5 in the COA, a trend which can only occur at depth.

Chromium

Cr does not show such a clear pH pattern; groundwaters with elevated chromium occur over a range of pH values (Tomlinson, 2019). Investigations of Cr behavior in soils, sediments, and groundwater have widely demonstrated that manganese oxides are required to oxidize chromium (III), which is relatively insoluble, to chromium (VI), which is much more toxic and mobile (Ji-Zheng et al., 2010; Fendorf and Zasoski, 1992; Silvester et al., 1995; Nico and Zasoski, 2000). As discussed above, dolomite in the COA contains 1 weight percent total manganese (Parkhurst et al., 1996). Westrop et al. (2016) have proposed that ion exchange and the resulting dolomite dissolution releases Mn(II), which oxidizes to Mn(III/IV) and then oxidizes Cr(III) to Cr(VI) (Fig. 7). Again, this can only occur at depth where mudstone is more prevalent and where longer flow paths have allowed more ion exchange, dolomite dissolution, and chromium oxidation to occur.

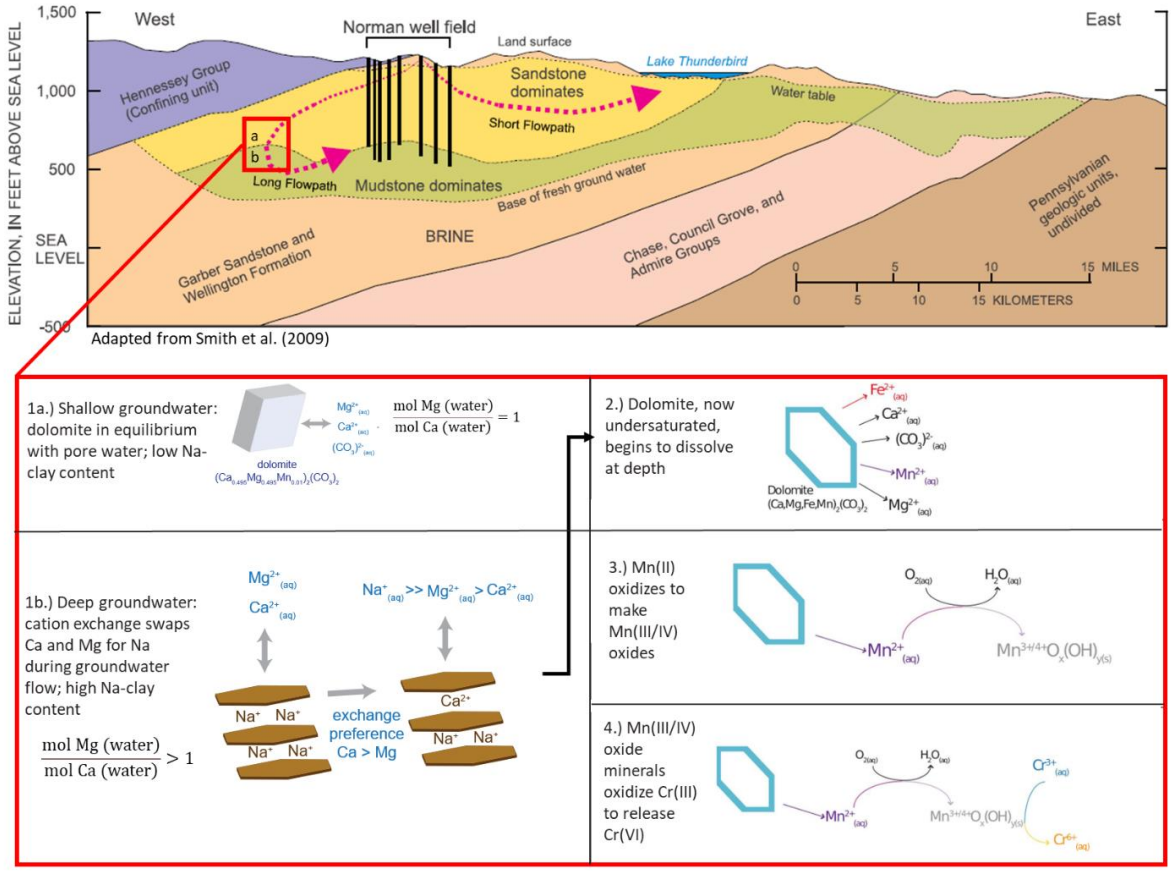


Figure 7: Hypothesized conceptual map for chromium mobilization occurring in the COA. Top: Generalized cross-section (with an exaggerated vertical scale) from Smith et al. (2009). Bottom: Conceptual model for Cr(VI) release in the COA, adapted from Westrop et al. (2016).

Uranium

There has also been substantial work on uranium mobility in aquifer systems; however, a comprehensive understanding of uranium occurrence in the COA is lacking. Carbonate ions can complex with U(VI) (which mobilizes under the oxidizing conditions found in the COA) and reduce the likelihood of uranium sorption (Liesch et al., 2015; Riedel and Kübeck, 2018). Interestingly, Hudak did not observe a statistically significant correlation between uranium and alkalinity in the Edwards-Trinity Aquifer; they attributed this finding to lack of variation in alkalinity (2018). Calcium-carbonate-uranium complexes are even more soluble than uranium-carbonate complexes, and positive correlations between calcium and uranium have been observed as well (Hudak, 2018; Riedel and Kübeck, 2018). In over 8,000 groundwater samples from Germany, 4.7x more samples exceeded 1 μg/L uranium when calcium concentrations were greater than 200 mg/L (Riedel and Kübeck, 2018). Although less widely studied, sulfate-uranium complexes have been observed as well (Langmuir, 1997). Extremely reducing conditions have also been shown to sequester uranium (Riedel and Kübeck, 2018). Reduction is not likely an important consideration in the COA, however, as dissolved oxygen is prevalent throughout COA samples. In the COA, samples with highest uranium concentrations had bicarbonate concentrations above 200 mg/L (Becker, 2013), which supports the idea that uranium in the COA is related to carbonate concentrations. Becker (2013) indicates that uranium sorbs onto iron oxides in the COA, supporting the idea of lessened uranium sorption with more abundant uranium-carbonate complexes in the COA.

Selenium

Like arsenic, both selenium (IV) and selenium (VI) have been shown to sorb onto iron oxides less effectively at higher pH in a lab setting (Rovira et al., 2008). Selenium concentrations in the COA also exhibit a similar depth dependence as arsenic and chromium (Table 1) which supports the idea that higher pH groundwater facilitates selenium mobilization.

Aquifer response to pumping

In addition to geochemical water-rock reactions, the interconnectedness of individual sandstone beds in the COA plays a large role in understanding how it will respond to increased pumping. A 15-day USGS pumping test on a Norman well in the COA showed that individual sandstone beds have little to no interconnectivity. Out of three monitoring wells less than 50 ft from the production well (with shallow, intermediate, and deep screens), only the deep-screened monitoring well showed a head change in response to the production well, which was screened at the same depth as the deep monitoring well (Mashburn et al., 2018). This suggests that over short time scales, there is little interaction between individual sand lenses which Norman wells are screened through. The results from the week-long pump test and the eight-day recovery period (Fig. 8) show that the head in screened sand lenses changes very quickly in response to pumping and recovers relatively slowly, as it takes the entire recovery period for the head to approach pre-pumping levels.

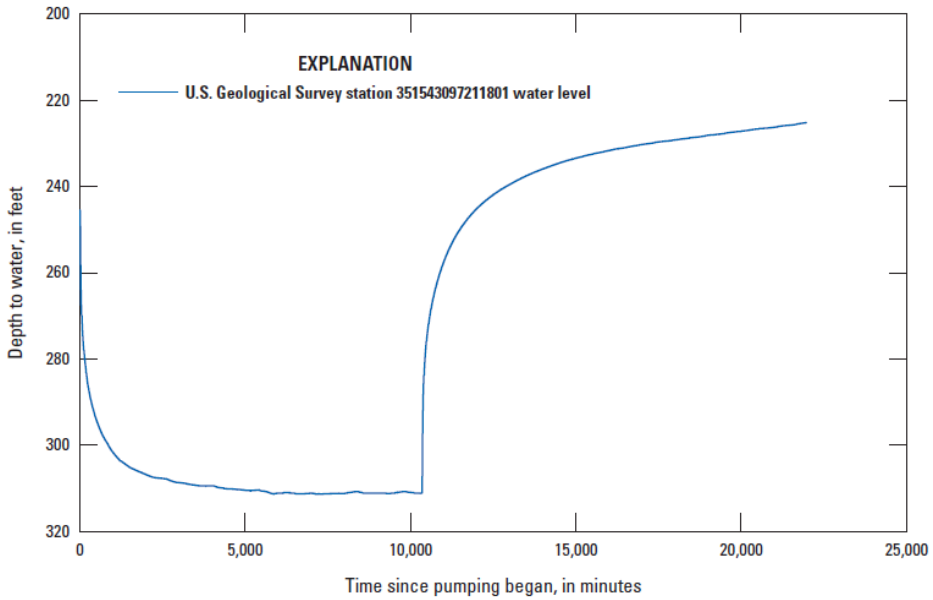


Figure 8: Norman pump test results, from Mashburn et al. (2018).

Hypotheses

General flow patterns and chemistry changes with pumping

Since the COA has sequences of individual, minimally connected sand bodies, groundwater would likely flow down inactive wells and short circuit the normally long flow paths into the deep COA, as observed in in the High Plains Aquifer near York, NE (Ayotte et al., 2011; Clark et al., 2008) and in the Upper Floridian Aquifer (Pichler et al., 2017). Large downward hydraulic gradients around pumping wells during periods of inactivity facilitate this downward flow (Clark et al., 2008). As discussed above, hydraulic head recovers in Norman wells in around 8 days, so induced downward flow could happen in any wells that were recently pumped.

The City of Norman well construction and use patterns also support potential down-well flow. Most City of Norman wells have between 5 and 21 screened zones (with most falling around 10 screens) which usually correspond to individual sand units, although some screens go through multiple units (Smith et al., 2009). This would allow shallow groundwater to enter one screen, flow down the well, and exit a deeper screen in the direction of groundwater flow (Fig. 9). The pumping pattern for well 20 (Fig. 9) is consistent with pumping patterns for most other City of Norman wells. All rapidly switch from off to on at short time intervals, which would allow high hydraulic gradients during non-pumping periods and therefore down-well flow. There are also longer-term variations in pumping (Fig. 9). “Increases in pumping” in this thesis therefore refers to both an increase in the number of days on for a given well over sample-specified time interval and an increase in the mean volume of water pumped over the same sample-specified time interval. These variables likely correlate together, as increasing the number of days with non-zero pumping would also drive up the mean pumping rate farther from zero.

During inactive periods under high hydraulic gradients, young, low pH, oxygenated groundwater would flow down the well and, to some degree, out of the well screens in the direction of groundwater flow, reacting until reaching equilibrium with the aquifer material (Fig 9). Downward groundwater flow in the well would cease during pumping, and the resulting cone of depression would induce some flow against the general groundwater flow direction in the zone of influence around the pumping well (Fig. 9). This would make higher pH, older water flow back through aquifer material previously equilibrated with the lower pH water coming out of the well screen. As pumping rates increase, the depression in the potentiometric surface around the well would increase and high pH groundwater would react along the previous low pH groundwater flow path for a greater distance. Several sources above discussed the importance of changing redox conditions between oxic and anoxic, and the effect that can have on trace metals. This is not a likely control here because the COA is almost entirely oxic or post-oxic (Parkhurst et al. 1996). Based on the conceptual models for groundwater flow through the system and previous literature discussed above, several hypotheses were developed to predict how increased pumping rates in the COA would affect concentrations of U, Se, As, and Cr in the groundwater.

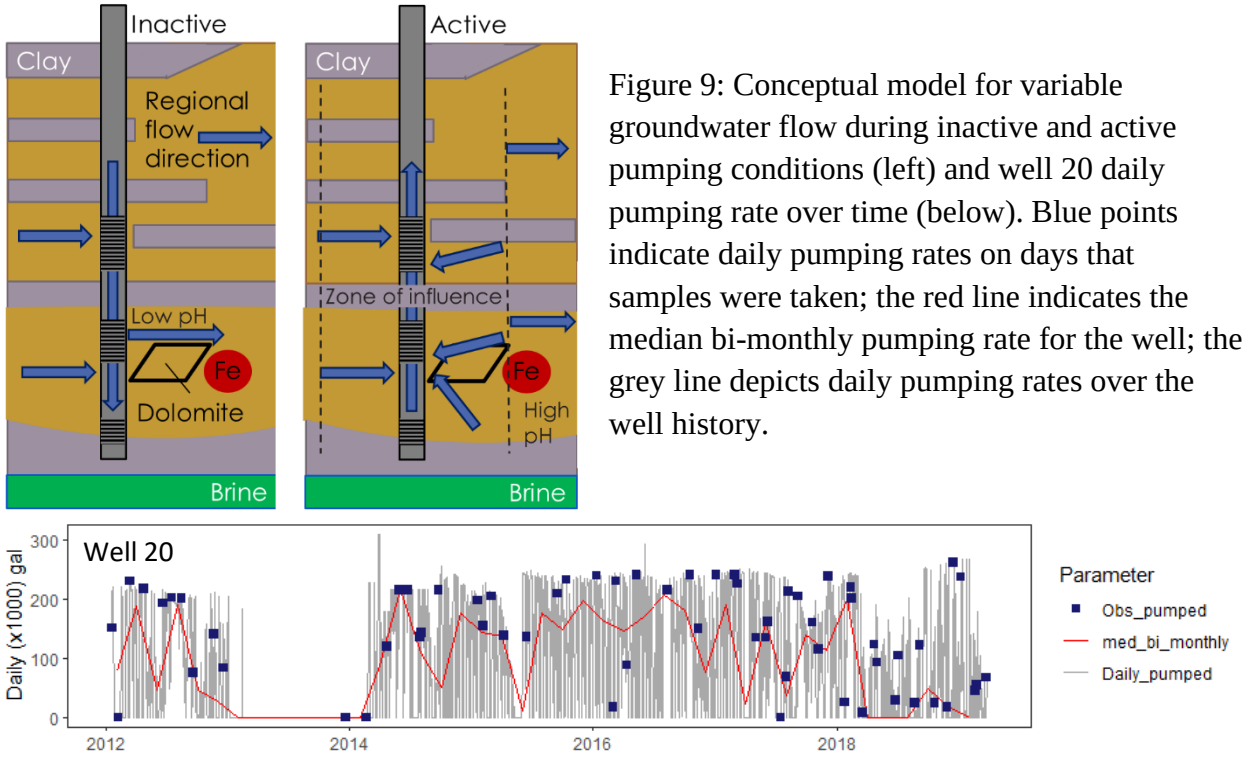


Figure 9: Conceptual model for variable groundwater flow during inactive and active pumping conditions (left) and well 20 daily pumping rate over time (below). Blue points indicate daily pumping rates on days that samples were taken; the red line indicates the median bi-monthly pumping rate for the well; the grey line depicts daily pumping rates over the well history.

Trace metal-pumping relationships

A positive correlation between uranium and pumping is likely, assuming that carbonate concentrations have a greater role in uranium mobility than calcium does. The groundwater flowing down wells would have low pH, low carbonate, and high calcium. The native groundwater deep in the aquifer would typically have very low calcium (from ion exchange) but much higher carbonate. When the well is off (with downward hydraulic gradients), low

carbonate, low pH water would flow down the well, reduce the pH and carbonate concentrations at depth, and sorb any free uranium onto iron oxides. Pumping would gradually replace the high Ca, low carbonate water with native low Ca, high pH water along the flow path. The resulting pH increase would desorb uranium from the aquifer sediments and generate a positive correlation between pumping and uranium. Arsenic and selenium are also likely sorbed onto iron oxides at low pH and desorb at high pH. Thus, they would also likely show positive correlations with pumping under the same mechanism.

Positive or negative correlations between pumping and chromium could occur based on the well depth and whether dolomite is saturated or undersaturated in the downward-flowing groundwater. PHREEQC models in Parkhurst et al. (1996) between samples with at least a 2,500-year age gap predicted dolomite dissolution in the model with the youngest age. Smith et al. (2009) showed that calcium-magnesium-bicarbonate groundwater (indicative of dolomite dissolution) is only prevalent near the potentiometric high for the COA, the area with the youngest water. Therefore, it is possible for down-well groundwater to be undersaturated with respect to dolomite, but it would only occur very close to aquifer recharge areas in the unconfined aquifer.

If the low pH water flowing down the wells is undersaturated with respect to dolomite, it could dissolve dolomite, oxidize chromium, and release Cr(VI) as it flows into the deep aquifer from the well screens. This would generate a negative correlation between pumping and total chromium because increasing pumping frequency would decrease the amount of downward flow in the well and reduce the Cr mobilization process. If the downward-flowing water is already saturated with respect to dolomite, a positive correlation between pumping and total chromium is expected because no Cr(VI) is mobilized when the well is inactive. Without Cr(VI) mobilization during inactive conditions, old, deep groundwater would be more likely to dissolve dolomite (and oxidize/mobilize chromium) than the downward flowing water because it has had more time to undergo ion exchange and has had more exposure to clays (due to longer flow paths) than the younger water. That older, higher Cr groundwater would get pulled into the well as the young, low pH groundwater that flowed down the well gradually gets replaced during pumping. This would also depend on the well depth, though, as the deepest groundwater in shallow wells would not be able to undergo the same degrees of ion exchange. Based on this understanding of chemical influences on Cr, As, Se, and U, we expect increases in pumping to increase arsenic, selenium, and uranium concentrations and either increase or decrease chromium concentrations depending on the dolomite saturation conditions.

Methods

Field Sampling

15 City of Norman wells with historical data were sampled from the well head during March and April 2021 in order to provide additional time-points for pumping-chemistry comparisons. Tubing was connected to the sample port nearest to each well, which fed the sample water into

the bottom of a bucket which was tilted slightly to allow flow along one section of the lip; this served as a basic flow-through cell. A multi-parameter probe continually collected pH, oxidation-reduction potential (ORP), dissolved oxygen (DO), conductivity, total dissolved solids (TDS), temperature, and pressure data from the bottom of the bucket near the inlet, which limited interference from the atmosphere. The flow-through cell was also left out of the sun whenever possible to preserve temperature accuracy. Once pH, conductivity, and TDS stabilized, four readings (one every two minutes) were collected for each field parameter. Then, we filled sample jars for metal (500 mL), anion (100 mL), and organic matter (1 L) analysis, stored them with ice, and turned the water off. The probe stayed in the flow-through cell for at least twenty additional minutes at each site without water running to get a more accurate DO measurement, because DO consistently dropped while the other measurements stabilized. Most of the wells had very low turbidity and were not filtered before being put in the sample jars; one well had visible colloids and was filtered through a 0.45 μm filter before the metal and anion collection. An Erlenmeyer flask was also filled with 100 mL of water (using a graduated cylinder for accuracy). The Hach Alkalinity Test Kit, Model AL-DT was used (along with its outlined procedure) with the Erlenmeyer flask to determine alkalinity from titration. Back at the lab, metal samples were acidified with 1 mL of 15 M nitric acid. The Atomic Spectroscopy Lab at Colorado School of Mines analyzed major and trace metal concentrations with ICP-AES and major anion concentrations with IC.

For quality control, the multi-parameter probe was calibrated three times over the course of the field work. No major corrections were needed, suggesting the instrument remained consistently calibrated during the field work. Additionally, anything that was in contact with the sampled water was rinsed with native water at least three times prior to storage (including the bucket, all sample containers, the Erlenmeyer flask, and the graduated cylinder). The bucket, Erlenmeyer flask, graduated cylinder, and multi-parameter probe were also rinsed with DI water and air dried after each field day. Results were added to the other data sources as described below. All field parameters (other than DO, which was reported as its 5^{th} measurement after the water stopped flowing) were reported as averages of the four readings collected once the readings stabilized.

Modeling

Well-head models

PHREEQC inverse modeling was used to determine plausible geochemical reactions that could produce a final well head sample chemistry from an initial well head sample chemistry at an earlier point in time (well_head_models folder in PHREEQC_inputs.zip). Well 5, 31, 33, and 36 samples were used in the models because they had the only well-head samples with complete major ion chemistries which were also sampled in 2021. Well 23 well-head samples were also modeled over time because the USGS collected a series of samples with variable pumping rates over a three-day period from well 23, which were helpful for looking at short-term pumping/geochemical interactions. Inverse modeling has generally been applied to consider chemistry differences between two wells along a flow path at a single (or approximately single)

point in time (Carucci et al., 2012; Kuells et al., 2000). PHREEQC inverse models take the initial solution and calculate the element mole transfers into/out of user-specified phases that are required to produce the final solution, within a margin of error for both individual elements and overall charge balance (Parkhurst and Appelo, 2013). Because these models only account for changes in element concentrations, there are no thermodynamic or kinetic considerations for inverse models. Consequently, there is no fundamental difference between running calculations over a flow path at a similar time and running calculations over a single well through time; solutions to the inverse models will still calculate the same mole transfers between solutions. Redox and valence states are calculated in the mass balance, however. For example, model solutions will account for S^{-2} and S^{+6} even if only one is included in a sample (Parkhurst and Appelo, 2013). Pyrite was included in model phases specifically to deal with these redox considerations. The PHREEQC default database was used for all models.

Major ions (Mg^{+2}, Ca^{+2}, Na^{+}, K^{+}, HCO_3^{-} , Cl^{-}, and SO_4^{-}) along with field parameters like pH and temperature (along with redox conditions and dissolved gasses in some cases) are generally the main components considered in PHREEQC inverse modeling since they play the largest single role in determining compositions, activities, etc. (Carucci et al., 2012; Kuells et al., 2000). We used the suite of major ions (with HCO_3^{-} expressed as alkalinity) along with iron, pH, and temperature to describe our solutions. Pe was left at the default value of 4 and water density was left at the default value of 1 kg/L. Iron was included in the solutions because of the important role of iron and iron oxides in the COA (Smith et al., 2009; Breit, 1998). CO_2 (g) and O_2 (g) were left out of the initial solutions because we had no capacity to measure CO_2 (g) in the field and DO consistently failed to stabilize to an accurate reading. CO_2 was added as a phase to the inverse model calculations, however, because it was required for some of the inverse models to generate solutions. By including CO_2, we assumed that enough CO_2 was present to facilitate the mole transfers calculated in the model solutions, which is a plausible assumption since most NWIS samples had measurable CO_2.

In addition to CO_2, barite, dolomite, pyrite, hematite, and goethite were included as phases for the mass-balance calculations, based on the observed mineralogy of the COA (Breit, 1998). Sodium, calcium, and magnesium exchange sites (NaX, MgX, and CaX) were also included as phases due to the observed importance of Na exchange with Mg and Ca in COA clays (Smith et al., 2009). All uncertainties were set to 0.1. Well 23 also required calcite as a phase to generate model solutions.

Smith et al. (2009) also discussed the potential importance of brine interactions in generating observed water chemistry. Therefore, rather than including just the initial and final well head solutions in each inverse model, we included both the initial solution and a brine as input solutions leading to the final well head solution. The brine solution was first described in Parkhurst et al. (1996). Interestingly, this brine is slightly acidic (pH = 5.713), which they attribute to it having evolved in place from sea water. This is inconsistent with the pH trends shown in Smith et al. (2009), which generally increase with depth. We examined the validity of a

low pH brine by looking at waters produced from rock during oil production (Fig. 10), which were assembled by the Oklahoma Geological Survey (OGS) into the Dowell Brine Database (2015).

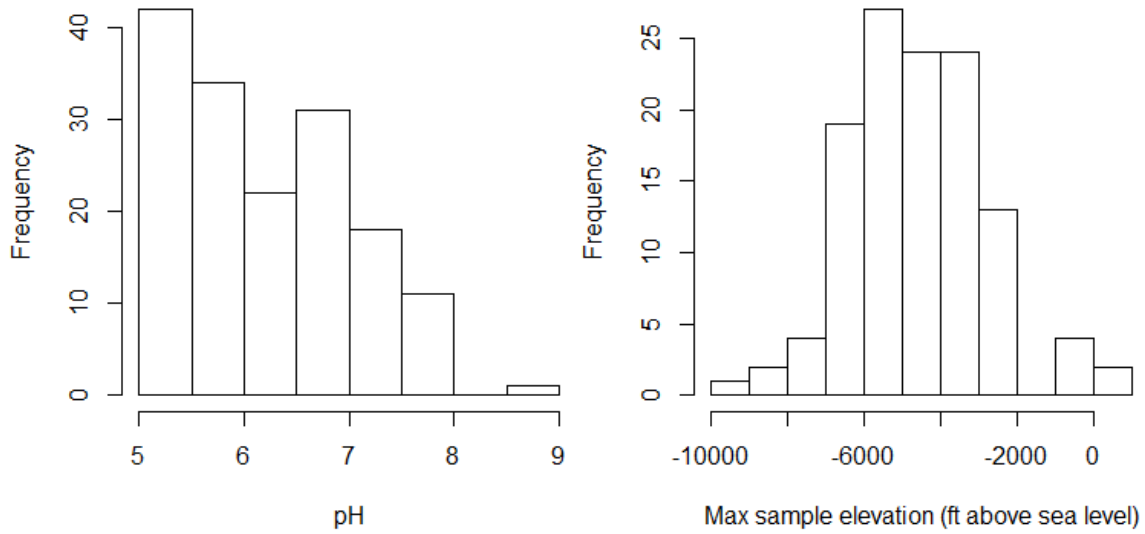


Figure 10: pH and sample depth of produced waters below the COA.

Samples from this database were filtered to only include samples from the spatial extent of the COA. The elevation of the base of fresh water in the COA ranges from more than 200 ft to less than 900 ft (Mashburn et al., 2018). With elevations ranging from approximately 859 to 1300 ft and a maximum reported total well depth of 766 ft for Norman wells, all wells in Norman pull water from above sea level, while most of the produced waters came from substantially below sea level and the reported brine threshold (Fig. 10). The two produced waters in the bin above sea level had pH values of 7.39. The lower pH value of Parkhurst et al. (1996), used also in this study, fits within the observed values of all regional brines but not the observed values from only regional brines in the same depth-range (Fig. 10). Increasing the model brine pH to 8 produced no change in the maximum brine contributions for the well-head inverse models; therefore, brine pH was left at 5.713 and likely had minimal effects on the model solutions.

Charge balance percent errors for the initial and final well head samples for wells 23, 5, 31, 33, and 36 ranged from 11.69% to -13.23%. The majority of well 23's samples had between 5 and 7% charge balance error, suggesting a potential sampling error. Due to the data limitations, charge balance errors of less than 7% were kept, but the two solutions with charge balance errors greater than ±7% were removed. Well 33 only had 2 samples, one of which with a 11.69% charge balance error, so it was not included in the model results. The major assumptions in this model set up are: (1) the phases encompass all the major geochemical reactions, (2) all other changes with time are negligible (infiltration, leaking between sand bodies, etc.), (3) there is sufficient CO_2 in the system to produce the calculated mole transfers, and (4) the modeled system has sufficient time to reach equilibrium.

Mixing models

The simplest alternative hypothesis to the proposed geochemical reactions occurring in response to pumping changes is a change in mixing proportions within the well. If pumping increases lead to a greater volume of water being pumped from a screened interval that contains high arsenic (and high pH), for example, arsenic and pH will increase in the well head sample regardless of any changes in geochemical reactions. Therefore, in addition to the well head inverse models described above, mixing models were also created to determine if depth-dependent samples could mix in varying proportions to produce the final well head solutions at variable pumping rates and conditions (mixing_only folder in PHREEQC_inputs.zip). Depth dependent USGS data were collected by attaching a screened PVC access tube to the pump column in each well; the depth sampler hose would then be extended into the access tube at the designated depth and the sample would be collected (Smith et al., 2009). Many of the sample depths did not correspond perfectly with a screened interval; it is unclear whether this was due to discrepancy between the well log screen locations and the screen locations determined by camera or if screen position was simply not considered during sampling.

In setting up these models, we assumed that the sample depth best represented the chemistry of the closest screened interval, and that the closer the sample was to the nearest screen, the more representative it was of the chemistry (Appendix II). If, for example, samples were collected within the depth range of two screened intervals and a third sample was collected between the two screened intervals, the third sample was thrown out because we assumed that the two samples collected at screened interval depths would be more representative. If two samples were collected on either side of a screened interval, they were mixed in PHREEQC before being used in the inverse model using inverse distance weighting as the mixing coefficients (Appendix II). This weighting reflects the assumption that samples closer to the screened interval would be more representative than samples farther away.

Wells 2, 5, 23, and 36 were chosen for these models because they all included multiple sets of depth-dependent samples, each collected within a few days of a well-head sample. Once individual solutions were allocated and/or mixed to best reflect the chemistry at individual screens, they were treated as initial solutions for the inverse modeling. The brine discussed above was also treated as an initial solution. The final solution for each inverse model run was the well-head sample that was collected at around the same time as the respective set of depth-dependent samples. No phases were included, as the models were intended to test the possibility of only mixing changes generating the well-head chemistry changes. All the major elements were included as balances, and all uncertainties were set to 0.1. Iron was not included in these models because concentrations were small compared to the major ion concentrations and, since phases were not included in the models, it was not needed to test the model objective about mixing feasibility.

The final solution for well 36 in its 2004 solution set had a charge balance error of -13.24%, so that entire solution set was not used. The 615 ft depth in the second set of solutions for well 2

had a charge balance error of -18.98%, so it was removed from the solution set. As discussed above, many of the well-head solutions for well 23 all had 5-7% charge balance error; however, they were still used. All other solutions used in the models had a charge balance error of ±5% or less. The major assumptions used in this set of inverse models were: (1) no significant reactions occur as the groundwater travels up the pump column, (2) the temperature of each sample is the same as the final temperature leaving the well head, (3) gas effects on sample chemistry are negligible, (4) no significant changes in groundwater chemistry occur at the depth-dependent sample depths between the time that they were collected and the time that the well-head sample was collected (up to two days), and (5) the depth-dependent samples encompass the entire range of groundwater chemistries entering the well. Assumption 5 is perhaps the least valid (but still necessary) assumption because, following the sample to screen allocation procedure described above, each model run for each well had at least two screens without allocated samples.

Post-modeling processing

Once all the models described above were prepared and ran in PHREEQC, they generated the number of moles of each specified phase that would have to precipitate/dissolve/exsolve to produce the final solution chemistry from the initial solution chemistry (mole transfers). These results were nonunique, so anywhere from 1 to over 100 sets of mole transfers existed for any given model. R was used to calculate the range of mole transfers for each phase over all the model solution sets. Each phase that had its entire range of mole transfers below 0 was considered to precipitate, while each phase with an entire range of mole transfers above 0 was considered to dissolve. Ion exchange was treated slightly differently. Based on ion exchange experiments done on cores drilled around the COA, the clay-sized fraction of COA material has median exchangeable ion concentrations of 1.95,17.5, and 14 meq/100 g for NaX, CaX, and MgX, respectively (Parkhurst et al., 1996). Mole transfers of NaX, CaX, and MgX in the PHREEQC model solutions were converted to final exchanger composition with the following formula:

$$\text{Median exchangeable ions} - \frac{\text{mol MX}}{\text{1 L}} \times \frac{\text{1000 mmol}}{\text{1 mol}} \times \frac{\text{1(Na) or 2 (Mg, Ca) meq}}{\text{1 mmol}} \times \frac{\text{0.22 L solution}}{\text{1 L total}} \times \frac{\text{1000 L total}}{\text{1 m}^3} \times \frac{\text{1 m}^3}{\text{1930 kg}} \times \frac{\text{0.1 kg}}{\text{100 g}} = \text{final exchangeable ions}$$

Moles per liter of exchanger are subtracted from the median exchangeable ions because positive mole transfer always means transfer to solution in PHREEQC; a positive mole transfer would be taking moles of ions out of the clays and therefore subtract from the exchangeable ions in the clays. The ‘0.22 L solution / 1 L total volume’ factor accounts for the 0.22 median porosity of the COA (Parkhurst et al. 1996). The ‘1930 kg / 1 m^3’ density of the aquifer rock comes from the mean Garber Sandstone density reported by Belt and Paxton (2005). The shallow COA (< 300 ft) has median exchangeable ion concentrations of 1.95, 18.5, and 14.5 meq/100 g for NaX, CaX,

and MgX, respectively, while the deep COA has median concentrations of 1.55, 16.5, 9.2 meq/100 g, respectively (Parkhurst et al., 1997). The 300 ft cutoff for the shallow vs. deep COA is discussed in Smith et al. (2009). According to these reported exchanger compositions, declines in NaX and/or the ratio of MgX to CaX indicates a greater contribution from deeper waters; an increase in NaX or MgX to CaX indicates a greater contribution from shallower waters.

Data Preparation

The R statistics software was used to compile and organize five different datasets for calculating trends between trace metal concentrations and pumping rates. These include: (1) general water quality parameters regularly sampled by the City of Norman, (2) field data, (3) compliance data recording concentrations of chemicals with EPA-mandated safe minimum concentrations (collected by the City of Norman), (4) National Water Information System (NWIS) data stored by the USGS for select Norman wells, and (5) pumping data showing how much water was pumped from each well each day. Each dataset is described in detail in Appendix I; the datasets required different types of pre-processing and standardization before they could be fully combined and utilized as a single dataset.

Different agencies report different data, so we used approximate formulas to estimate magnesium and calcium concentrations (within 5% error) when possible (Appendix I). Conductance was also converted to specific conductance to unify the parameters. Not all chemical components were collected (or could be estimated) for each sample, so non-parametric statistics on individual ions were used for correlation analyses. Some of the data (75 instances out of 15,839 observations of individual parameters) had corrected measurements in addition to original measurements assigned to the same date, time, well, and parameter, so processes were constructed to pick the likely corrected measurements (Appendix I). Specifically, rules were constructed to pick the lower number from a set with likely misplaced decimal points, report the measurement as below detection for duplicate samples with conflicting detection statuses, and average together the two measurements if they were sufficiently close (less than 10 units) with no way to differentiate them. Although averaging sets of duplicate measurements is not ideal when more information can be obtained, it does halve the error that would occur from selecting the wrong measurement. Finally, some wells had multiple samples taken from a given well on the same day. These would be tied to the same pumping information since only one volume per well per day was reported. Therefore, any measurements of the same parameter in the same well on the same day were averaged together, but only for calculations involving pumping.

New metrics for understanding trace metal-pumping relationships

Change in interval mean daily pump rate

Two new metrics are introduced in this study: the change in interval mean daily pump rate and the quadrant correlation metric. The change in interval mean daily pump rate is a way to track general changes in pumping conditions between pairs of consecutive samples. Consider, for example, two arsenic samples taken out of well 5: 10.9 μg/L on 11/9/05 and 8.1 μg/L on 11/17/05 (Fig. 11). The change in interval mean pumping rate is -60,000 gal/day, calculated by

taking the mean of daily pumping rates over the sample interval (November 9 to November 17, 2005, inclusive) minus the mean of daily pumping rates over the same amount of time prior to the first sample, which is November 1 to November 9, 2005 (Fig. 11).

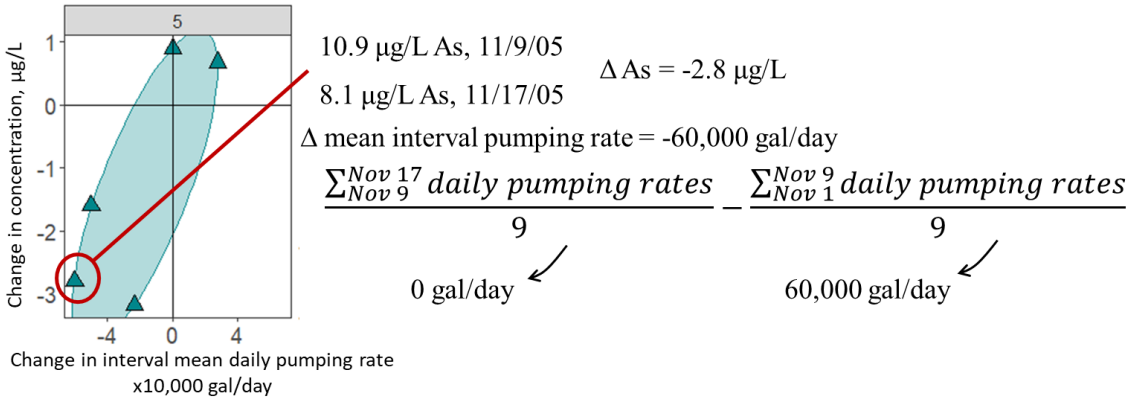


Figure 11: Example calculation for the change in interval mean daily pumping rate using two samples from well 5. The cross plot on the left shows all pairs of consecutive samples for well 5 using the same calculation technique. The change in interval mean daily pumping rate provides the x-coordinate, while the y-coordinate is the change in groundwater arsenic concentration during the considered interval.

Calculating a metric for changing pumping conditions in this way has several benefits. Firstly, it is time-scale independent; it works for samples that are days apart and years apart. This is advantageous in a system like the COA where there is a lot of time-scale variation in the available data. Secondly, using the mean instead of the median ensures that the metric is sensitive to any pumping days in the time interval. Lastly, the change in interval mean daily pumping conditions has the advantage of taking into account both the pumping conditions before the first sample was taken and the conditions between samples. For this example, knowing that the well was on recently before the two arsenic samples were collected aides in the interpretation that arsenic concentrations decreased with lessened pumping.

Quadrant correlation metric

The quadrant correlation metric is a robust way to numerically evaluate the strength of trend between pumping and trace metal concentrations in a given well. It essentially counts the number of samples in the upper right and lower left quadrant of a change in pumping vs. change in trace metal concentration cross plot (Fig. 11) and then subtracts the number of samples in the upper left and lower right quadrants. Finally, it divides this result by the total number of samples. For well 5 in Fig. 11, the quadrant correlation metric is: $\frac{(1+3)-(0+0)}{5} = 0.8$. 1 is the strongest possible positive correlation, and -1 is the strongest possible negative correlation, so 0.8 indicates that in general, there is a direct relationship between the change in pumping and the change in arsenic concentration; in other words, increases in pumping produce increases in arsenic concentrations and decreases in pumping produce decreases in arsenic concentrations.

The quadrant correlation metric is helpful to use here because most established correlation methods rely on individual samples (individual concentrations) rather than pairs of samples as

used here. Additionally, the metric is completely independent of the magnitude of change; it makes no difference if a point has a 100 μg/L change in trace metal concentration or a 1 μg/L change in trace metal concentration with a given change in pumping conditions. This is valuable because of the mixed timescales in the data; longer time scales could logically produce larger changes in trace metals, so this procedure helps equalize those points.

Kendall's Tau Correlation Analysis

In addition to the new quadrant correlation metric, the more established Kendall's Tau correlation analysis can provide useful information about the system because it is also non-parametric and robust; it orders the quantities in one variable and calculates how well the quantities in the other variable are ordered in relation to that. For example, points (1,2); (2,3); (3,4) would have a Kendall's Tau correlation coefficient close to 1 because each increase in x has an increase in y. (1,4); (2,3); (3,2) would have a correlation close to -1 because each increase in x has a decrease in y. The change in mean daily pumping volume over the interval and the change in trace metal concentrations between samples could not be used as inputs for these calculations because concentrations themselves (rather than changes in concentration) were necessary for the calculation.

If arsenic concentrations went from 10 to 8 to 6 μg/L over a sequence of samples, for example, using change in concentration for the correlation coefficient would pick up no trend (due to a tied concentration change of -2 μg/L across each interval) while the concentration would pick up the clearly substantial trend. The same would occur if the change in mean interval pumping rate was used. Since raw concentrations rather than change in concentration are required, any pumping metric assigned to a given concentration must be independent of surrounding samples. Therefore, mean concentration over the week leading up to each sample (Appendix I) was chosen to compare with trace metal concentrations because: (1) the aquifer returned to equilibrium in one week during the Norman pumping test and (2) the metric is independent of surrounding samples. Although necessary, it is also beneficial to use different time scales to look at pumping correlations. The new metrics discussed above assume that the pumping conditions over the entire time interval reflect concentration changes. If the trace metal concentrations better reflect short term pumping conditions over only a few days prior to the sample being taken, the Kendall's Tau correlation between trace metal concentrations and week-long pumping rates would be more likely to detect it.

Change in interval mean pumping rate vs. change in concentration would still be a valuable metric for the application of Kendall's correlation coefficient analysis, because it would help determine if the amount of change in concentration relates to the amount of change in pumping rate. However, this is outside the scope of the current thesis, which seeks only to address if changes in pumping produces a meaningful change in trace metal concentration and not how the magnitudes might relate. Once the method was determined, a matrix of values was computed with three values for each parameter for each well: the Kendall's Tau correlation coefficient between the trace metal concentration and the mean daily volume of water pumped from the well

(averaged over the week leading up to each sample), the p-value testing the null hypothesis that the true correlation coefficient was equal to 0 (no relationship), and the sample size. The function "cenken()" in the "NADA" library calculated both the correlation coefficient and the p-value (respectively). The function takes arguments of both the measurements and a list of TRUE and FALSE values, TRUE if the measurement is below detection and FALSE if the measurement is not below detection. It uses the Akritas-Theil-Sen method to compute the Kendall's Tau correlation coefficients, which is recommended for censored data by the USGS in their book *Statistical Methods in Water Resources* (Helsel et al., 2020).

Trace metal hazard mapping

Data from NWIS and the array of Norman data discussed above was utilized to compute hazard maps of elevated trace metal risk for each trace metal along with an overall water quality hazard map. The field data collected for this study and City of Norman data was divided into wells that had USGS data and wells that did not have USGS data prior to processing for the maps; this prevented sites from being counted twice. Two maps were generated from each parameter: one well-head map for the entire COA and one depth-dependent map around Norman only.

Many methods exist to map trace metal hazards, such as concentration, water quality index (WQI), heavy metal pollution index (HPI), metal index (MI), etc. (Mahapatra et al., 2020; Yadav et al., 2018). Concentration maps are difficult to apply in the COA because many of the wells only have concentrations below detection. There is also a lot of variation in the frequency of sampling; some wells have 15 measurements for any given trace metal and others have one or no measurements per trace metal, so it would also be necessary to compute some statistic to display overall concentrations from a given well. Multi-metal metrics like HPI, MI, and WQI are difficult to apply to the COA because not every well has measurements, even below detection measurements, for each parameter. Metrics that sum multiple metals would be lower than the true value for wells that are missing metals. Because of these constraints, a high-concentration limit was constructed for each trace metal, set at 10% lower than the maximum contaminant levels (MCLs) set by the EPA. Then, the number observations above the high-concentration limit were counted for each well/parameter and divided by the total number of observations for that well/parameter (called the fraction of samples above the high-concentration limit, or FSAL). FSALs were applied to both individual trace metals and as an aggregate metric over all four trace metals of interest (chromium, selenium, arsenic, and uranium). The aggregate FSAL was constructed by adding all of the trace-metal samples that exceed their respective high-concentration limits in the well and then dividing by the total number of samples. For example, a well with two arsenic samples (10 μg/L and 2.1 μg/L), one selenium sample (30 μg/L), and no other samples would yield a fraction of 1/3 because only one of the three samples exceeds its high concentration limit.

The FSAL method has multiple benefits in a data-limited system. It is independent of number of samples, so wells that have a lot of samples of each trace metal do not require any additional processing steps. Also, it incorporates below detection values without additional assumptions. If

the detection limit for a particular sample is below the high-concentration limit for a given trace metal, it can be counted in the FSAL calculation. There were a few cases where the detection limit was above the high-concentration limit (like arsenic concentrations of < 110 μg/L with a high-concentration limit of 9 μg/L, for example). These samples were omitted from the mapping.

Once the FSAL was computed for each trace metal for each well (along with the aggregate calculation described above), the wells were plotted using ArcGIS Pro. Then, ordinary kriging was used to extrapolate the fraction of samples above the high-concentration limit across the entire COA, again for each trace metal and the aggregate map. Ordinary kriging assumes that there is an underlying spatial pattern to point value variations and calculates an empirical relationship for that variation with distance (called a variogram) from a user-specified model (the spherical model was used here, which assumes that points stop correlating after a finite distance threshold). Then, ordinary kriging uses the local mean from the 12 nearest points around the point of interest, along with the variogram and the distances from the nearest points to the point of interest, to determine the expected value at the point of interest (Li and Heap, 2008). This process then repeats for a grid of points which become the centers of 1000 m x 1000 m raster cells, the same cell size used by Mashburn et al. (2018) to map hydraulic conductivity across the COA.

Kriging was chosen for the interpolation method because, when tested to generate points removed from a surface, it was less sensitive to sample size and more visually faithful to the original surface than other widely used methods such as the multiquadric radial basis function (RBF) and inverse distance to a power (IDP), even though it generated slightly more error than the multiquadric RBF function (Yang and Holder, 2000). “Visual accuracy” dictates how errors are distributed on the map; poorer visual accuracy means that the errors are highly concentrated in certain areas of the map, even if the amount of net error is lessened. Visual accuracy is more desirable in this case than the minimum net error because a zone with highly concentrated errors could be misinterpreted in terms of water quality. Additionally, the lack of sensitivity to sample size made kriging more applicable than other methods because different areas of the COA ranged wildly in sample point density. Ordinary kriging was also recommended by Li and Heap (2014) for estimating continuous variables without a global mean or information from secondary variables. This best fit the needs of constructing a hazard map without instilling any assumptions about the normality (in the case of the global mean) or correlation with any secondary variables. Kriging can generate a surface through an entire area regardless of the number of nearby points, but it is less accurate (as all interpolation methods are) when there are less nearby points. Therefore, regions in the COA greater than 5 km (~3.1 miles) away from any sample point were colored white to indicate lack of data in the region.

Results and Discussion

Historical major ion chemistry changes

One of the main reasons for studying chemistry change with pumping around Norman, OK is the drastic change in the COA potentiometric surface between the 1980's and 2009 (Fig. 5). All non-depth-dependent USGS and field data for sites within a 6.5 mile radius of the center of the 2009 cone of depression in Norman were gathered and plotted to determine how the chemistry in that region of the aquifer has changed overall through time (Fig. 12). 6.5 miles (~10.5 km) was chosen because it covered the entire region where groundwater flow paths would likely be affected by the cone of depression, based on the potentiometric surface (Fig. 5). Data from 1986-1988 and 2003 both have 3 samples falling in the lower carbonate/bicarbonate region, followed by 2004 and 2005-2008 with 2 samples apiece in the lower carbonate/bicarbonate region, followed by 2021 with one sample in that region (Fig. 12). This reflects a general increase in pH and carbonate over time. There was also a known increase in pumping from wells 20 and 13 from the 1980's to 2009.

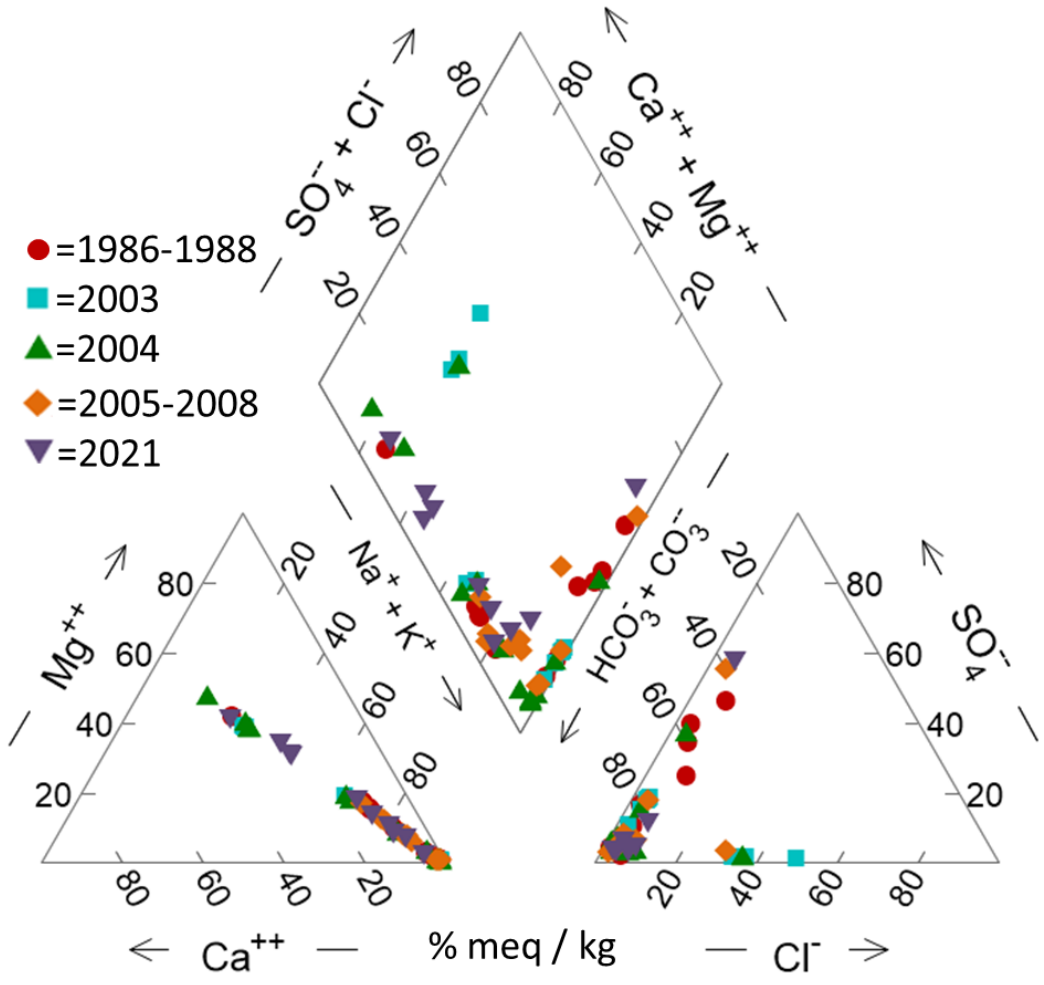


Figure 12: Piper plot of historic groundwater chemistry over area of greatest potentiometric surface change. Date-interval divisions were chosen based on data abundance.

Modeling

Week to year-scale inverse models

Several sets of models were generated using PHREEQC to determine if any dominant geochemical reactions would be associated with changing trace metal concentrations on time scales ranging from one week to 16 years. Trace metals were not included in the models themselves but are shown with the model results. Changes in concentration from below a detection limit (dl) to above were still reported as the change in concentration to half of the dl or from half of the dl when the dl was below the observed value. For example, a change in As concentrations in well 5 from < 110 µg/L to 10.9 µg/L was not reported but a change in Se from 9 µg/L to < 8 µg/L was reported as -5 µg/L. Reactions that occurred between an earlier sample and a later sample for wells 5, 31, and 36 were investigated with PHREEQC inverse modeling using changes in well-head major ion chemistry. For these wells, increases in interval mean daily pumping rate consistently produced increases in As, Se, and Cr concentrations; decreases in the interval mean daily pumping rate consistently produced decreases in As, Se, and Cr (Fig. 13). Since Fig. 13 plots changes in concentration, any points in the upper right and lower left quadrants follow this pattern. The change in interval mean daily pump rate also matches the change in pH for all but one set of samples (Table 2).

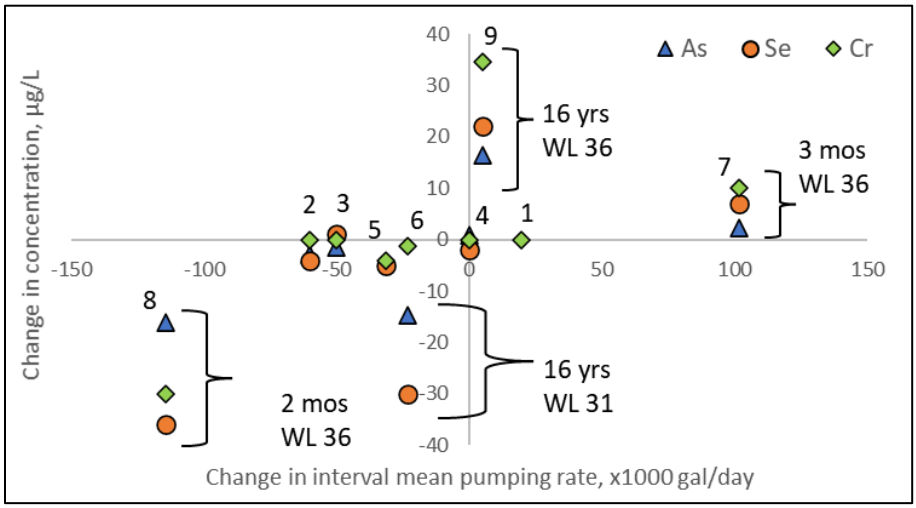


Figure 13: Changes in trace metal concentrations vs. changes in mean pumping rate for wells (WL) 5, 31, and 36. Integers above each set of trace metal points refer to the model run that they correspond to.

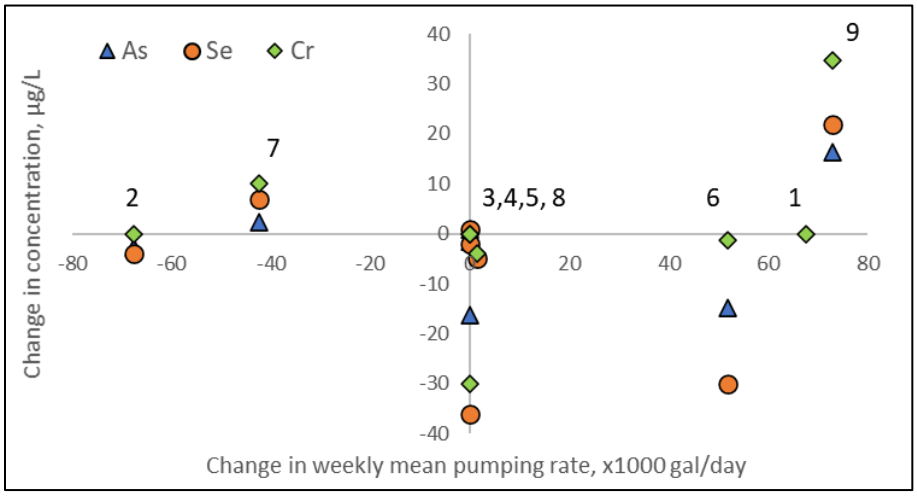


Figure 14: Change in trace metal concentrations vs. changes in weekly mean pumping rate for wells 5, 31, and 36. The integers above each set of points indicates the corresponding model run.

Interestingly, the magnitudes of trace metal changes were well-specific, as well 5 (integers 1-5 in Fig. 13) produced less trace metal variations overall. In wells 31 and 36, longer time periods between samples produced greater changes in trace metal concentrations than shorter time periods between samples at comparable or even lower changes in mean daily pumping rate across the sample interval. Uranium was below detection for each sample and was therefore excluded. Note that the change in interval mean daily pumping rate does a better job at separating out the changes in trace metal concentrations into distinct quadrants than the mean daily pumping rate aggregated only over the week leading up to the sample date (Fig 14).

Geochemical reactions for these wells also followed interesting trends. PHREEQC produces numerous solutions to any given inverse problem. To interpret these results, the range of moles transferred to/from all phases was calculated for each inverse model solution set. If the entire range of results for a given model and phase transfer is positive, dissolution is likely and is shown with a "D" (Table 2). If the entire range of results for a given model and phase transfer is negative, precipitation is likely and is shown with a "P" (Table 2). Relative changes in pH, As, Cr, and Se are shown alongside model results for context.

Model, Well	pH	As	Cr	Se	Hem-atite	Goet-hite	Dolo-mite	NaX	MgX	CaX	CO_2	Brine fraction	Δ $\overline{\text{Int}}$ pumping
1, 5	↑	NA	↔	NA	D	P	P	1.9, 2.0	14	17.5	E	2.25e-05, 6.44e-06	↑
2, 5	↓	↓	↔	↓	P	D	D	1.9	14	17.5	D	1.42e-06	↓
3, 5	↑	↓	↔	↑	NO SOLUTIONS								↓
4, 5	↑	↑	↔	↑				1.9, 2.0	14	17.5		0, 5.86e-06	↔
5, 5	↓	↓	↓	↓	NO SOLUTIONS								↓
6, 31	↓	↓	↓	↓	P	D	D	1.9, 2.0	14	17.5	E	1.94e-05, 2.82e-05	↓
7, 36	↑	↑	↑	↑				1.9, 2.0	14	17.5		0, 9.44e-06	↑
8, 36	↓	↓	↓	↓	NO SOLUTIONS								↓
9, 36	↑	↑	↑	↑	D	P	P	1.9	14	17.5	E	1.20e-05, 9.33e-06	↑

Table 2: Summary of geochemical reactions predicted by PHREEQC in wells 5, 31, and 36. The ↔ symbol indicates no change in concentration. "E" refers to gas exsolution, "D" refers to dissolution, and "P" refers to precipitation. Letters are only listed in cells in which the entire range of model solutions yielded the indicated phase transfer. CaX, MgX, and NaX show the range of final clay exchange site compositions (in meq/100 g) based on mole transfers in the respective PHREEQC solutions and an initial median exchange site composition of 1.95, 14, and 17.5 (respectively), from Parkhurst et al. (1996). Δ $\overline{\text{Int}}$ pumping refers to the change in interval mean pumping rate between the sampling interval and the same length of time before the sample interval.

Pyrite and barite were also included as phases in the inverse models; no mole transfers of barite occurred in any model solutions and only negligible amounts of pyrite dissolved/precipitated, on the order of $1x10^{-8}$ or less. "NA" is reported for changes in As and Se in model 1 (Well 5, Table

2) because the first samples for both contained concentrations below detection limits. NaX, MgX, and CaX (exchange sites) all have final site compositions very close to the initial site compositions, so ion exchange likely did not contribute much to the composition changes in the model solutions. Likewise, the brine contributed very little, if at all, to the final solution composition. The model with greatest increase in interval mean pumping rate over the time period, model 7, had at least one solution with no additional brine contribution whatsoever (Table 2). All the models that did not produce solutions had imbalances in sulfate. This is a major limitation of the current knowledge of chemical evolution in the COA. Smith et al. (2009) discuss potential sulfate contribution from the confining layer above the COA, but it is not well constrained.

There are no notable trends in CO_2 mole transfers as they relate to changes in pH and trace metal concentrations; CO_2 exsolution is linked to both trace metal increases and decreases. When the models predict significant hematite dissolution, trace metals/pH increase and goethite is predicted to dissolve (Table 2). Similarly, trace metals/pH decrease when goethite is predicted to precipitate and hematite is predicted to dissolve. Although this seems significant, it is simply a function of PHREEQC's mass balance algorithm and does not relate to iron concentrations. Iron concentrations do increase from the initial to final solutions in the two models that predict hematite dissolution, but one of solution sets with predicted hematite precipitation has increased iron concentrations as well. The fact that goethite dissolves with pH decrease likely factors into PHREEQC's H^+ balance rather than any physical phenomena, therefore, because goethite dissolution adds H^+ to solution.

Dolomite dissolution with pH/trace metal decrease and dolomite precipitation with pH/trace metal increase (as predicted in the model results) is reflected in the solutions, unlike the iron oxides. Both starting solutions have more Mg and Ca than the ending solutions in the models that predict dolomite precipitation. Both starting solutions have less Mg and Ca than the ending solutions in the models that predict dolomite dissolution. Although thermodynamically possible, dolomite precipitation is not actually observed in the COA; nor is it likely at the observed pressures, temperatures, and general groundwater chemistries in the COA (Parkhurst et al., 1996). Model-predicted dolomite precipitation simply reflects decreases in Ca and Mg from initial to final solutions, perhaps masking ion exchange behavior not captured by these models.

Decreased pH leading to dolomite dissolution is opposite to the process described in Smith et al. (2009), wherein ion exchange in the deep COA facilitates dolomite dissolution and drives pH up instead of down. However, the two ideas are not mutually exclusive because of the different time scales and mechanisms for dolomite dissolution; the ion exchange to dolomite dissolution process occurs over hundreds to tens of thousands of years over the entire flow path in the deep COA. Low pH, dolomite-undersaturated groundwater flowing down the well during inactive periods between samples on the scale of weeks to years could yield dolomite dissolution with a pH decrease without disrupting the thousand-year trends observed by Smith et al. (2009). Both

models that predicted dolomite dissolution occurred in periods of lessened pumping; both periods that predicted dolomite precipitation occurred in periods of increased pumping.

Hour-scale inverse models

While wells 5, 31, and 36 are excellent indicators of trace metal changes over time intervals spanning weeks or months (models 2, 3, 4, 7, and 8) to many years (models 1, 5, 6, and 9), well 23 provides a good indicator of trace metal changes over hours to days. The USGS took seven samples from well 23 at varying instantaneous flow rates (measured in gallons per minute, or gpm) from 09/16/2003 to 09/19/2003. Note that this is a different metric than the pumping rate metrics used above, measured in x1000 gal/day, because of the shorter time interval. Other than relatively static chromium concentrations, trace metal concentrations tended to increase over the course of the day and then decline overnight (Fig. 15). This fits with the trace metal trends in the longer time-scale wells because well 23 was not otherwise pumped during the 4-day sample period. Therefore, longer gaps with lessened pumping (overnight periods in this case) decreased general trace metal concentrations. Intervals with increased pumping (subsequent samples on the same day) produced increases in trace metal concentrations. It is worth noting that the only increase in chromium concentrations occurred during a decrease in arsenic, selenium, and uranium, and the only decrease in chromium occurred while the other three metals increased.

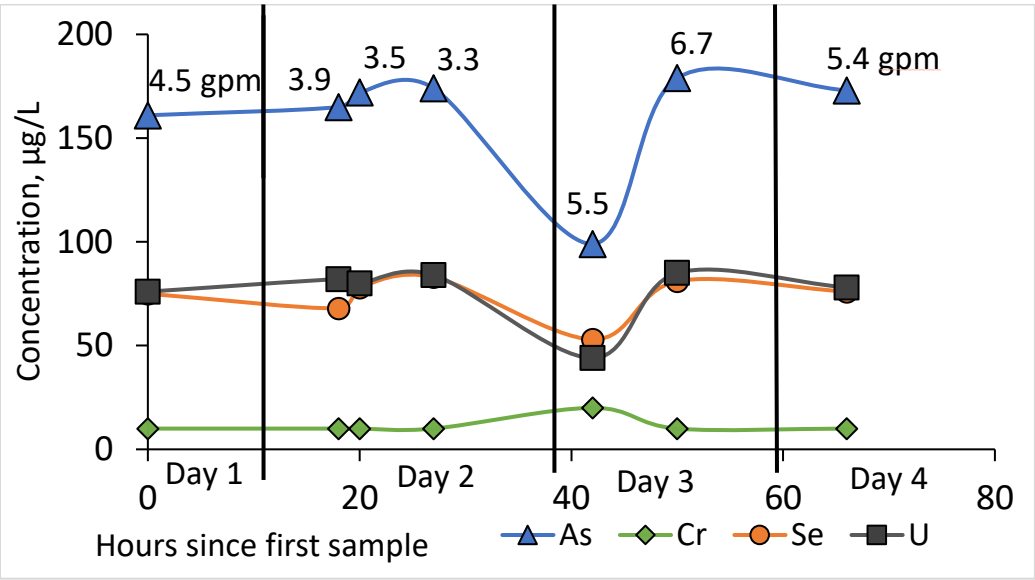


Figure 15: Changing trace metal concentrations with time in well 23. Numbers above the arsenic points indicate the instantaneous flow rate (in gallons per minute, or gpm) that each sample was collected at, which apply to all four trace metals. Each section of the graph corresponds to a sample day.

Magnitudes of trace metal increase/decrease varied wildly. One possible explanation for the differences in magnitude between increases in trace metal concentrations through the day could be the flow rate. Day 3 had a much more drastic change in flow rate over the course of the day than day 2, and it had a similarly drastic increase in trace metal concentrations. In addition to the

variable flow rate well head samples for well 23, an additional set of depth-dependent samples was collected less than a year later. Two sets of inverse models were created: one set to generate the final well-head solutions from the previous well-head solutions (Table 3) and another set to generate each final well-head solution from the set of depth-dependent samples (well_23_mix_and_phases folder in PHREEQC_inputs.zip, Table 4). The correlation between trace metal increases (arsenic and selenium especially) and predicted dolomite precipitation was consistent between the well-head sample models in well 23 (Table 3) and the longer-term models for the other wells (Table 2), as both model intervals with dolomite precipitation had increasing arsenic and selenium concentrations. Ion exchange is much more prevalent in the well 23 well-head model set (Table 3) than the models for wells 5, 31, and 36 (Table 2). Some of the model solutions for the final exchange composition in well 23 show a reduction in MgX relative to CaX, suggesting a greater influence of deep water as discussed in the methods. This could relate to the temporal differences between the models; at short time scales, the well-head groundwater chemistry might be less influenced by reactions and more by differences in water mixing.

Model, Day	pH	As	Cr	Se	U	Hem-atite	Goe-thite	Cal-cite	Dolo-mite	NaX	MgX	CaX	CO_2	Brine fraction
1,2	↓	↑	↔	↓	↑					1.9, 2.0	14	17.5	E	0, 9.93e-07
2	↔	↑	↔	↑	↓				P	1.9, 2.0	14	17.5	E	0, 6.15e-07
2	↓	↑	↔	↑	↑		P			1.9, 2.0	2.3, 14.0	17.5, 29.2		0, 4.63e-06
2,3	↑	↓	↑	↓	↓	P	D	D		1.9	8.3, 14.0	17.5, 23.3	D	1.91e-07, 8.26e-06
3	↓	↑	↓	↑	↑	D	P		P	2	10.4, 14.0	17.5, 21.1	E	0
3, 4	↔	↓	↔	↓	↓		D	P		1.9, 2.0	14	17.5	D	0, 1.07e-07

Table 3: Well-head model solutions for well 23. Table follows the same conventions as Table 2.

Models that predicted reactions between individual screened intervals to produce the final well-head solutions in well 23 (Table 4) did not show many significant trends in the range of mixing fractions that matched the observed trace metal trends. The maximum mixing fraction from the 510 ft sample (the second number in column 4, Table 4) does decrease over each sampling day and increase over each night, and the 445 ft sample has decreasing maximum contributions over the first day of sampling, but no other patterns fit with the pattern of trace metal concentrations increasing during the day and decreasing during the longer periods of inactivity. Additionally, the minimum mixing fraction from the model solutions (the first number in columns 2-6, Table 4) are nearly always 0 or less than 0 in some cases.

Flow rate	420 ft	445 ft	510 ft	550 ft	560 ft	Hem-atite	Goe-thite	Cal-cite	Dolo-mite	NaX	MgX	CaX	CO_2
4.5	0, 1.7	0, 1.78	0, 2.68	0, 0.02	0, 0.02	D	P	D	P	2	13.9, 14	17.5	E
3.9	0, 1.51	0, 1.58	0, 2.38	0, 0.02	0, 0.02	D	P	0	P	2, 2.1	13.9, 14.1	17.4, 17.5	
3.5	0, 0.77	0, 0.78	0.38, 1.6	0, 0.02	0, 0.02	D	P	0	P	2	14	17.5	E
3.3	0, 0.082	0, 0.08	1.07, 1.2	0, 0.01	0, 0.01	D	P		P	1.9, 2	14	17.5	E
5.5	0, 0.96	0, 0.99	0.31, 1.82	0, 0.02	0, 0.02	D	P		P	2	14	17.5	D
6.7	-19.4, 0.83	0, 20.9	0, 0.96	0, 0.05	-0.17, 0.04			P	D	1.9, 2	14	17.5	E
5.4	0, 1.38	0, 1.44	0, 2.57	0, 0.02	-1.65, 0.02				P	2	14	17.5	E

Table 4: Mixing coefficients and phase transfers for depth-dependent PHREEQC model solutions of well 23. Models are shown in chronological order. Columns 2-6 indicate the range of mixing proportions for the solutions taken at those depths, rounded to 2 decimal places in the format ‘minimum mixing proportion, maximum mixing proportion’. Flow rate is shown in gpm. The other columns follow the same notation as Tables 2 and 3.

Dolomite, hematite, and goethite show consistent phase transfers across the depth-dependent models (Table 4), which likely stems from a missing initial solution. Of the 7 screened intervals in well 23, only 5 had samples that could be attributed to them (Table 5). All the final, well-head solutions except the ones taken at 5.5 and 3.3 gpm had Mg concentrations below all of the input solution concentrations, so it is very likely that a screened interval across a sand body with lower Mg concentrations was not sampled. A sixth initial solution at 485 ft was excluded from the well 23 models because of the screen allocation method described above; the two nearest screens were 466 to 480 ft and 488 to 513 ft. Both 510 ft and 485 ft were closest to the screen starting at 488 ft, but 510 was within the screen and therefore deemed more representative.

Mixing-only models

The simplest alternative hypothesis to the proposed mechanisms for trace metal change in the COA is that changes in mixing proportions between groundwaters from individual sand layers at different pumping rates generate the observed trace metal variations. In order to test this hypothesis, more PHREEQC inverse models were generated to determine if the final well-head solutions could be produced from different mixing proportions in the depth-dependent initial solutions for wells 23, 5, 2, and 36, all of which had sets of depth-dependent samples taken within a few days of well-head samples. Two out of the five sets of final solutions could be generated from the corresponding set of initial solutions for well 5, one out of three for well 36 (with one model excluded for charge balance error), zero out of two for well 2, and zero out of seven for well 23 (Table 5). In other words, none of the seven final solutions at different flowrates for well 23 could be produced from different combinations of the five depth-dependent initial samples without allowing some reactions to occur.

It is impossible to entirely rule out the possibility of this hypothesis because the models may fail to converge due to either input missing solutions from unsampled screened intervals or non-modeled reactions occurring in the system. These mixing models had anywhere from 2 to 6 solutions less than the actual number of screens per the well log (Table 5). Relaxing the assumptions used to assign depth-dependent samples to the nearest and most representative well screens could help increase solution numbers, but there would still not be any modeled well with a sample for each screen. There seems to be no relationship between model convergence and number of screens accounted for, however. Successful model runs in wells 5 and 36 do not have more samples allotted to screened intervals than the unsuccessful ones, and although well 2 does have less allotted samples than the others, well 23 does not and still has no solutions (Table 5). So, although missing input solutions do not allow us to fully reject the hypothesis that the variation in trace metals with pumping is simply due to changes in mixing fractions, it seems less likely due to the lack of trends between number of solutions and model success for the mixing models.

Well	Number of allotted screened intervals in:			Number of models	
	successful model(s)	failed model(s)	well log total	Success	Fail
23	NA	5	7	0	7
5	4-5	4-5	7	2	3
2	NA	2-5	8	0	2
36	3	3-4	7	1	2

Table 5: Number of allotted screened intervals in successful vs. unsuccessful inverse mixing models. NA values indicate that there were no successful models for that well.

Summary of modeling trends

The modeling results above show that dolomite dissolution most closely mirrors changes in trace metals for all of the inverse models, including the shorter-term models for well 23 and the longer-term models for wells 5, 31, and 36. Every instance where the entire range of inverse model results yielded dolomite dissolution was linked to lower concentrations of arsenic and selenium, and every instance of dolomite precipitation occurred with higher concentrations of arsenic and selenium. Uranium was only measurable in well 23 but showed similar increases and decreases as the other trace metals in all but one set of solutions. Chromium patterns were not as strong (as it varied less than the other trace metals) but followed the same trends as the other metals in wells 5, 31, and 36. The only change in chromium concentration in well 23 was opposite the changes in the other trace metals, but it was only a slight variation.

We ruled out the possibility that the trace metal/dolomite relationship was just due to PHREEQC mass balancing procedures like the hematite and goethite phases, because, unlike iron concentrations, the solutions' calcium and magnesium concentrations reflected the same increases and decreases that the model results predicted. Established literature on trace metal mobility in the COA has not shown any direct relationship between dolomite and arsenic,

selenium, or uranium, however, so the relationship between them is likely not one of causation. More likely, lower pH, dolomite-undersaturated groundwater flowing down the well dissolved dolomite and facilitated trace metal sorption onto the iron oxides present in the COA. High pH conditions (which generates higher dolomite saturation levels) also desorb trace metals. Dolomite dissolution would be directly related to chromium mobilization through manganese oxidation as outlined above (Fig. 7).

Well 23 provides a potential counterexample for the pH-driven arsenic and selenium mobilization mechanisms proposed above. Much of the literature on arsenic and selenium mobilization highlights the importance of pH conditions, and pH generally changes similarly to these trace metals in the other wells. Yet, large variations in arsenic and selenium occur in well 23 without corresponding variations in pH. Across all well 23 samples, pH only varies between 8.8 and 9. If pH increases during increased pumping conditions truly facilitate all or most of the variation in arsenic/selenium, there would be much less variation in those trace metals here. Still, dolomite precipitation is predicted on both multi-sample days with increasing trace metal concentrations, which indicates abundant carbonate even though dolomite precipitation has not been observed (Breit, 1998). More work needs to be done on short-term geochemistry response to pumping to determine how these reactions compare to the longer-term trends.

Contributions from brine were essentially negligible through all of the well-head models, with a maximum mixing fraction of 2.82e-5. 7 out of the 12 successful well-head inverse models did not even require the brine to generate solutions. Additionally, the models that did require brine to run successfully did not have higher pumping rates overall. Well-head model 7 (well 36), with the largest positive change in interval mean daily pumping rate of all the models, did not require the brine for its model solutions. Model 2 had the largest negative change in pumping conditions and had a non-zero brine contribution. As a notable exception, the only non-zero brine contribution for well 23 came from the inverse model between the last sample on day two and the first sample of day three and had the largest positive change in flow rate for that well at 2.2 gpm.

Ion exchange in wells 5, 31, and 36 was also fairly negligible, as the final exchange composition only varied by 0.05 meq/100 g from the initial solution, and only for NaX. CaX and MgX values remained constant. Well 23 model solutions included a wide range of ion exchange contributions. NaX varied by the same small amount as in the other wells, but final MgX composition varied from the initial composition (14 meq/100 g) to 2.3 meq/100 g and CaX varied from the initial composition (17.5 meq/100 g) to 29.2 meq/100 g. So, some model solutions had very little ion exchange activity while some had a lot. Those with higher ion exchange show more CaX preferentially incorporated over MgX. That suggests a greater proportion of deep groundwater in the system, because the deep COA sample cores discussed in Parkhurst et al. tended to have a greater ratio of CaX to MgX in clay composition (1996). These findings have implications for potential time-scale variation in how the aquifer responds to

pumping. However, none of the models required this change in ion exchange to generate at least one solution, so evidence for it is insufficient.

There are strong positive correlations between the trace metal concentration changes and the change in mean daily pumping rate over the sample interval for the modeled wells (Fig. 13, Table 2), which support the hypothesis above. Under non-pumping conditions, there could be sufficient downward gradients around the well to make low pH groundwater flow into the well and sorb some of the free trace metals onto iron oxides (reducing arsenic, selenium, and uranium concentrations). PHREEQC predicts that this groundwater may dissolve dolomite as well. Then, increasing pumping rates pulls increasingly higher pH water back out of the well, desorbing trace metals along the flow paths (increasing arsenic, selenium, and uranium concentrations). This not only works for the longer-term models but for well 23 as well. The overnight periods between samples could have been long enough to facilitate some downward flow and reduce trace metal concentrations; then, repeated pumping to get samples over the course of the day would increase the concentrations again. One potential issue with this understanding of the system is that dolomite dissolution should increase chromium concentrations, so chromium should show opposite trends as arsenic, uranium, and selenium. This does occur in well 23 but not in the long-term well-head models. When similar change in concentration and mean pumping rate calculations were applied over the entire well field, positive correlations and consistent trace metal behaviors in general were much less prevalent. In fact, some wells had negative correlations between pumping and trace metal concentrations (Fig. 16, Table 6).

Trace metal trends in the overall well field

Although all of the wells used in modeling are in relative agreement that increasing the mean daily pumping rate relative to the previous mean daily pumping rate at any time scale will increase trace metal concentrations, they represent a relatively small subset of Norman wells. Applying the same calculation methods of change in interval mean daily pumping rate vs. change in trace metal concentration to the entire well field makes it clear that many wells do not follow the same trends (Fig. 16).

Individual trace metals exhibit different relationships between mean pumping rate and measured well-head concentration. When it has a correlation, chromium has a positive correlation with change in mean pumping rate in every instance except in well 33 (Fig. 16), so the modeled wells are fairly representative for chromium. Selenium is not often reported, so it will be omitted from this section. Arsenic and uranium display an array of trends across the well field, ranging from strongly positive to strongly negative, and often have trends in opposing directions (Fig. 16). Not all of the samples from Fig. 13 and 14 are represented in Fig. 16 because all changes in trace metal concentrations with below detection limit values (either initial or final) were removed due to code limitations.

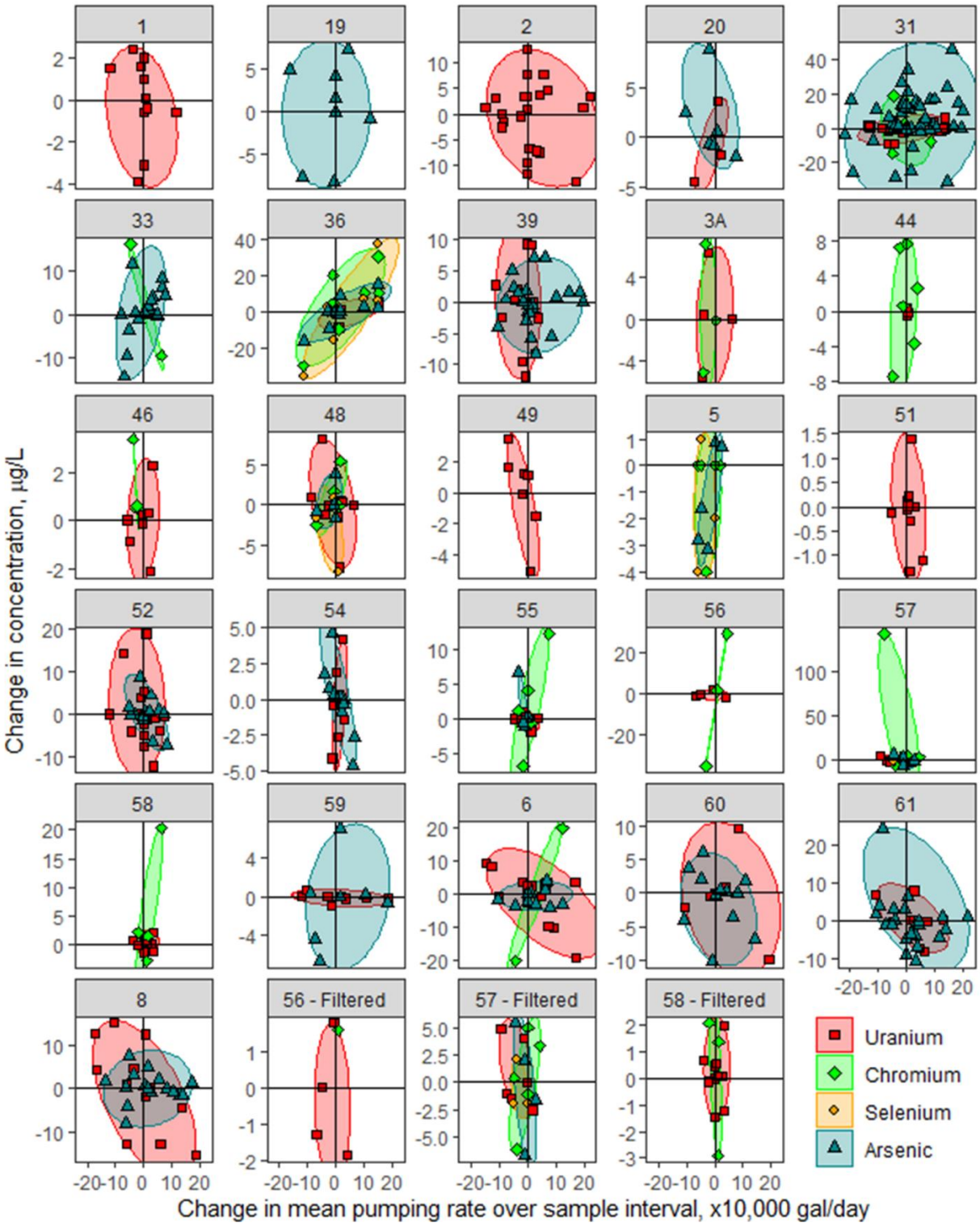


Figure 16: Trace metal trends with pumping across Norman wellfield. Each subplot represents a single well, and each point represents a set of consecutive samples for a particular trace metal. The x value for each point is the change in interval mean daily pumping rate (measured in x10,000 gal/day). The y value for each point is the change in trace metal concentration from the

earlier sample to the later sample. By calculating the values in this way, any increases in value over time are positive and any decreases in value over time are negative. Points are color coded by trace metal. Ellipses mark the extent that all points for that trace metal fall within and therefore do not correspond to a particular statistic. Wells 56, 57, and 58 have additional panels where the largest changes in trace metal concentrations are filtered out.

Even though the change in mean interval daily pumping rate correlated better than the weekly mean pumping rate with changes in trace metal concentrations for the wells used in modeling (Figs. 13 and 14), correlations may still occur over other time intervals. Kendall's Tau correlation coefficients between trace metal concentrations and the weekly mean pump rate over the week leading up to each sample was calculated to determine if the different metric would yield substantially different correlations than the interval mean pumping rate (Table 6). It also helps determine the validity of the assumption that trace metal concentrations reflect the pumping conditions over the entire sample interval rather than just a narrow window of pumping conditions close to when samples are taken.

Uranium	Arsenic	Chromium	Selenium	pH	Wells
-0.190	*-0.300*	*0.095*	*-0.067*	✔0.280	57
-0.100	*NA*	*0.667*	*-0.167*	✔0.377	56
0.067	*0.100*	*0.400*	*-0.333*	0.070	48
-1.000	*-0.048*	*NA*	*0.300*	-0.022	20
-0.100	*0.000*	*0.333*	*0.067*	0.126	3A
-0.105	✔-0.235	*0.250*	*-0.214*	0.024	31
✔0.603	✔0.365	*NA*	*NA*	✔0.502	39
-0.043	*0.133*	*NA*	*0.300*	0.275	2
-0.327	-0.182	*0.100*	*-0.400*	0.152	6
0.300	✔0.329	*0.167*	*-0.333*	✔0.305	33
✔-0.545	*0.167*	*NA*	*NA*	✔0.429	51
-0.298	✔-0.448	*NA*	*NA*	0.195	52
-0.345	✔-0.577	*NA*	*NA*	-0.013	54
0.156	*0.524*	*0.333*	*0.200*	0.148	55
-0.182	*NA*	*0.500*	*0.500*	✔0.275	58
-0.356	-0.105	*NA*	*NA*	0.141	59
✔0.491	-0.029	*NA*	*NA*	✔0.224	60
-0.436	✔-0.486	*NA*	*NA*	0.148	61
-0.107	*0.500*	*NA*	*NA*	0.052	49
NA	*NA*	*0.667*	*NA*	0.158	43
NA	*NA*	*0.500*	*0.167*	0.146	47
NA	*0.167*	*NA*	*NA*	0.169	45
-0.667	*0.167*	*-0.143*	*0.333*	-0.079	44
0.028	*0.200*	*0.333*	*NA*	0.168	46
✔-0.449	*0.036*	*NA*	*0.167*	-0.295	1
NA	-0.212	*NA*	*NA*	✔-0.473	19
-0.238	✔-0.408	*NA*	*NA*	✔0.215	8
NA	*0.333*	*NA*	*NA*	✔0.261	38
NA	*0.167*	*NA*	*NA*	✔0.266	40
✔0.655	0.276	✔0.409	0.197	✔0.402	36
-0.048	*-0.214*	*-0.200*	*0.067*	✔0.269	5

Table 6: Kendall's Tau correlation coefficients between trace metal concentrations and mean daily pumping rate over a week. Green checkmarks indicate results that are statistically significant, and the red italic text indicates that the wells have less than 10 samples of that specific trace metal.

As with the change in interval mean daily pump rate vs change in concentration, the Kendall's Tau correlation coefficients show that no trace metal is strongly correlated with pumping across all wells. All trace metals have a mix of positive and negative Kendall's Tau correlations as well. pH, however, is almost always positively correlated with pumping rates, which fits with the Piper plot showing higher carbonate and bicarbonate concentrations in a high-pumping area (Fig. 12). Only one of the 13 statistically significant Kendall's correlations between weekly mean pump rate and pH is negative (Table 6). Applying the quadrant correlation method to the trends in Fig. 16 yield generally similar results to the Kendall's Tau correlation coefficients shown here. Of the 13 statistically significant Kendall's Tau correlations, 9 had the same sign as the quadrant correlation metric. The magnitudes of the quadrant correlations did not match well with the magnitudes of the Kendall's Tau correlations; only 2 of those 9 quadrant correlations would be considered strong (±0.5 or higher).

Well properties influential in trace metal-pumping correlation variation

In order to determine how positively or negatively trace metal concentrations correlate with changes in pumping rate, the trends in Fig. 16 were numerically quantified into the quadrant correlation metric. Then, these values were plotted against different quantifiable characteristics of the wells: (1) the fraction of days that the well was turned on between the first to last sample, (2) the mean daily pumping rate (in x1000 gal/day) between the entire period from the first to last sample, (3) the average aquifer rock percent sand at the well location, and (4) the standard deviation of daily pumping volume data from the first to last sample, in x1000 gal/day (Fig. 17, Fig. 19, Fig. 21). Each point represents the pumping/percent sand/etc. information for a single well.

The raw data for the percent sand calculation comes from the COA modeling work by Mashburn et al., where the COA was divided into 1000 m x 1000 m x 100 ft grid cells, each with a percent sand value related to the hydraulic conductivity (2018). Percent sand values in their model were constructed by a combination of outcrop and well lithologic/gamma logs. This data was combined with the land elevation and the well data (both construction and location) to determine which raster cells matched each well laterally and with depth. Well construction information was limited; for wells with known screens, the raster cells with elevations matching the first screen to the total depth in the vertical stack at the well location were averaged to get the percent sand. For wells with known total depths, raster cells from the land surface to the total depth were averaged for the percent sand. Raster cells for wells without screen or total depth information were averaged from the land surface to the last non-zero raster value.

Arsenic

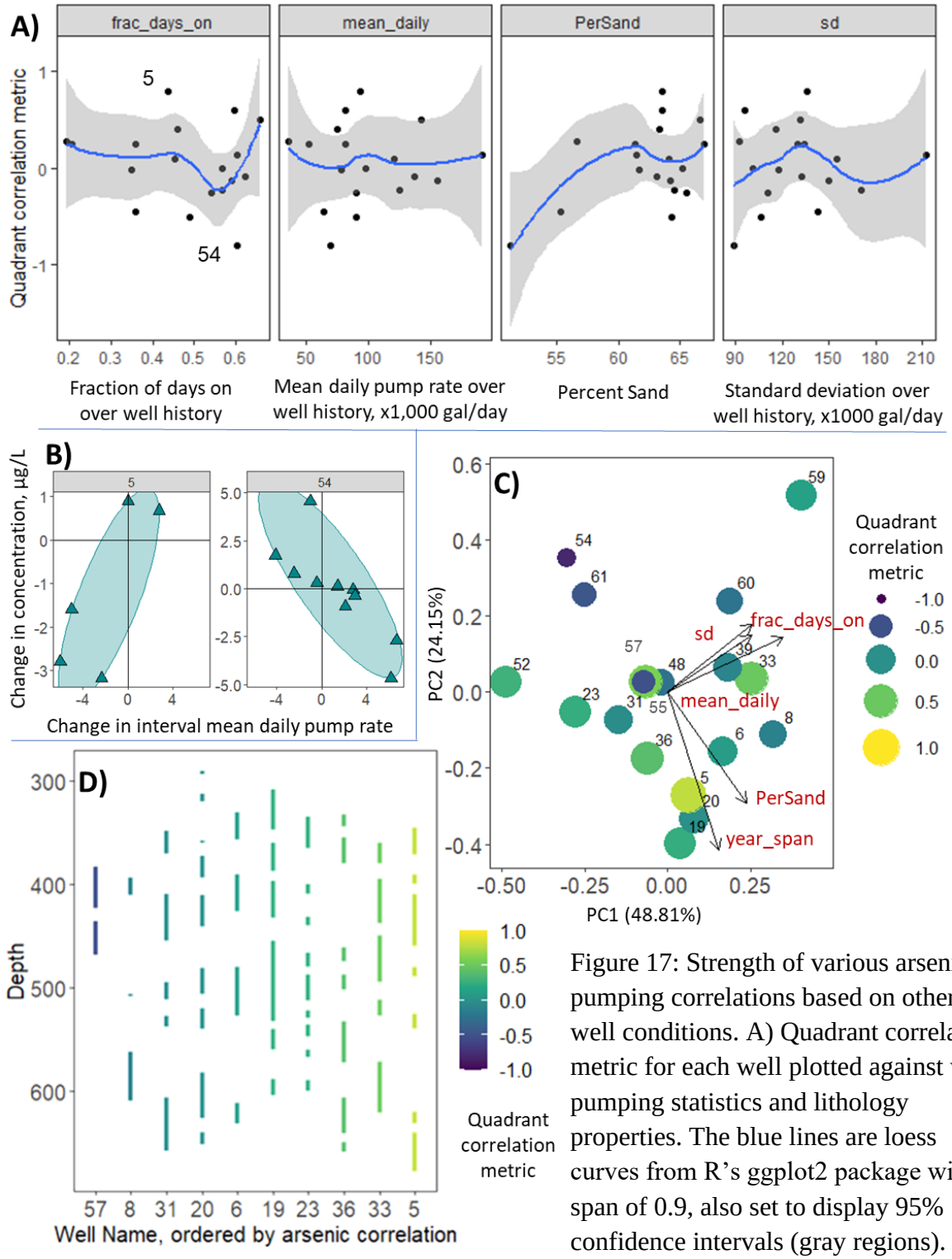


Figure 17: Strength of various arsenic-pumping correlations based on other well conditions. A) Quadrant correlation metric for each well plotted against well pumping statistics and lithology properties. The blue lines are loess curves from R's ggplot2 package with a span of 0.9, also set to display 95% confidence intervals (gray regions). B) Cross-plots showing the wells with the highest and lowest quadrant correlations, from Fig. 16. C) A PCA plot of all the well properties from panel A along with year_span, the number of years between the first and last sample dates for each well. Points vary with color and size based on the quadrant correlation metric between arsenic and pumping. D) A cross section map showing

screens for all available Norman wells that produced arsenic variations; not all wells had known well screens so there are more points in panels A and C than there are in panel D.

Wells with no correlations between pumping and arsenic concentrations have higher standard deviation than wells with either a lower or higher correlation; wells with lower percent sand are more likely to have negative correlations between pumping and arsenic (Fig. 17). All of the strong arsenic correlations (both positive and negative) also fall between 70,000 and 100,000 gal/day for mean pumping rate over the entire sample period for any given well. PCA analysis indicates that percent sand and the number of years spanning well sampling are the greatest driving forces for arsenic correlation variations; the lowest of both corresponding to more negative arsenic correlations (Fig. 17c). Standard deviation, mean daily pumping rate over the interval, and fraction of days on have a smaller influence on the overall distribution of points in the PCA plot, but there is an increase in correlation with these variables along a line from well 57 to well 59, which matches the left leg of the trend in these variables in Fig. 17a.

The percent sand and mean daily pump rate controls on arsenic – pumping correlations are consistent with the conceptual model for down-well flow in the aquifer. When a well first turns on, it pulls some deeper, older groundwater in the zone of influence from different flow paths and raises the pH close to the well, which leads to arsenic desorption from iron oxides (Fig. 18). While it is possible that flow path modification could bring in younger, lower pH waters as well, the Kendall's Tau correlation coefficients and the wellfield-wide changes in carbonate concentrations with time generally support increased pH with pumping. Once the well is turned off, down-well flow lowers the pH in a fixed zone of influence around the deep well screens and some of the arsenic sorbs back onto the iron oxides.

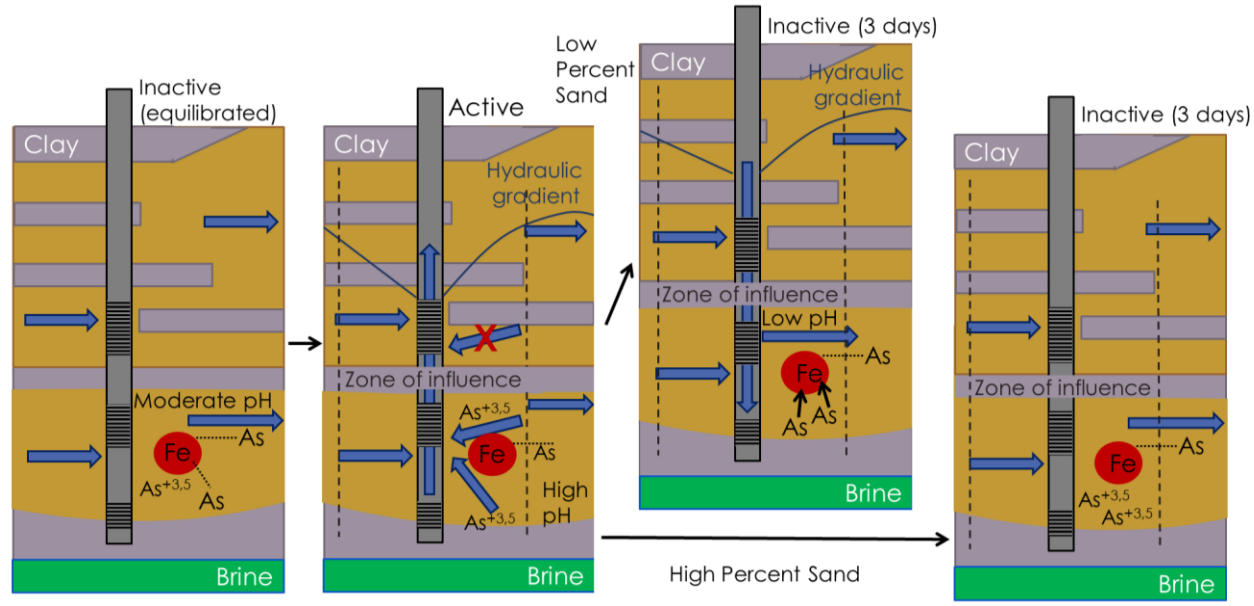


Figure 18: Conceptual model for arsenic variation in wells with variable percent sand

This process only occurs before the aquifer re-equilibrates to non-pumping conditions and the cone of depression terminates. Downward hydraulic gradients would last longer under lower percent sand conditions because of the decreased hydraulic conductivity, so the downward flow would last longer and sorb more arsenic onto iron oxides in lower percent sand wells. In these cases, every pumping event could initialize more downward flow and sorb enough arsenic back onto the iron oxides to outweigh any arsenic increases during the actual pumping event, thus generating negative correlations with pumping. Under higher percent sand conditions, the downward flow after pumping would last a shorter time and the arsenic increases during pumping would more likely outweigh any arsenic decreases after pumping, generating positive correlations between arsenic concentrations and pumping rates.

As an example, consider a well which is sampled and pumps at a constant rate for two days, then is off for four days, then pumps for a day and is sampled again. At a low percent sand, the hydraulic gradient would likely allow flow down the well for the entire four-day period. The four days off would outweigh the three days on and there would be a decrease in arsenic. At a higher percent sand, the hydraulic gradient may only allow flow down the well for a single day. Then, the three days of pumping would outweigh the one day with flow down the well, so pH would be higher and arsenic concentrations would increase. The fraction of days on would be 3/7. The change in mean daily pumping rate would depend on the conditions that produced the sample in the week prior to the first sample, because the example sample interval covers a week. If, for example, the well was off for the week prior to the first sample, the change in mean daily pumping rate would be positive.

Most of these arsenic variations would occur within the well's zone of influence because that is where the most drastic pH oscillations would occur between pumping and non-pumping conditions (Fig. 19). The low pH groundwater flowing down the well would have a fixed zone of influence, but the zone of influence during pumping conditions would increase with higher pumping volumes. Higher pumping would therefore likely weaken the arsenic-pumping correlations because more water would come into the well from outside the area with most pH change and therefore reduce any change in concentrations.

Chromium

For chromium, positive correlations with pumping occur in wells that run between 40 and 60% of the time overall (Fig. 19). Wells with higher pump rates and standard deviations occur with more negative correlations, and wells with negative correlations also have higher percent sand than wells with positive correlations. Lastly, wells with no correlation tend to have shallower screens (Fig. 19). The chromium PCA does not show very much variation with the plotted variables, but there are two nearly perpendicular trends: one from low correlations with high percent sand/mean daily pump rate to higher correlations with lower percent sand/mean daily pump rate and one from high standard deviation/low correlation to low standard deviation/high correlation. Thus, there is an exceedingly complex array of conditions which determine how a well might respond to increased pumping, depending on how the hydraulic gradient might

respond to the change in pumping and whether that would outweigh the pH increase from pumping itself, which in turn depends on how frequently the well is being used, the chemistry of the downward flowing water, possibly the screen depths depending on the trace metal, etc.

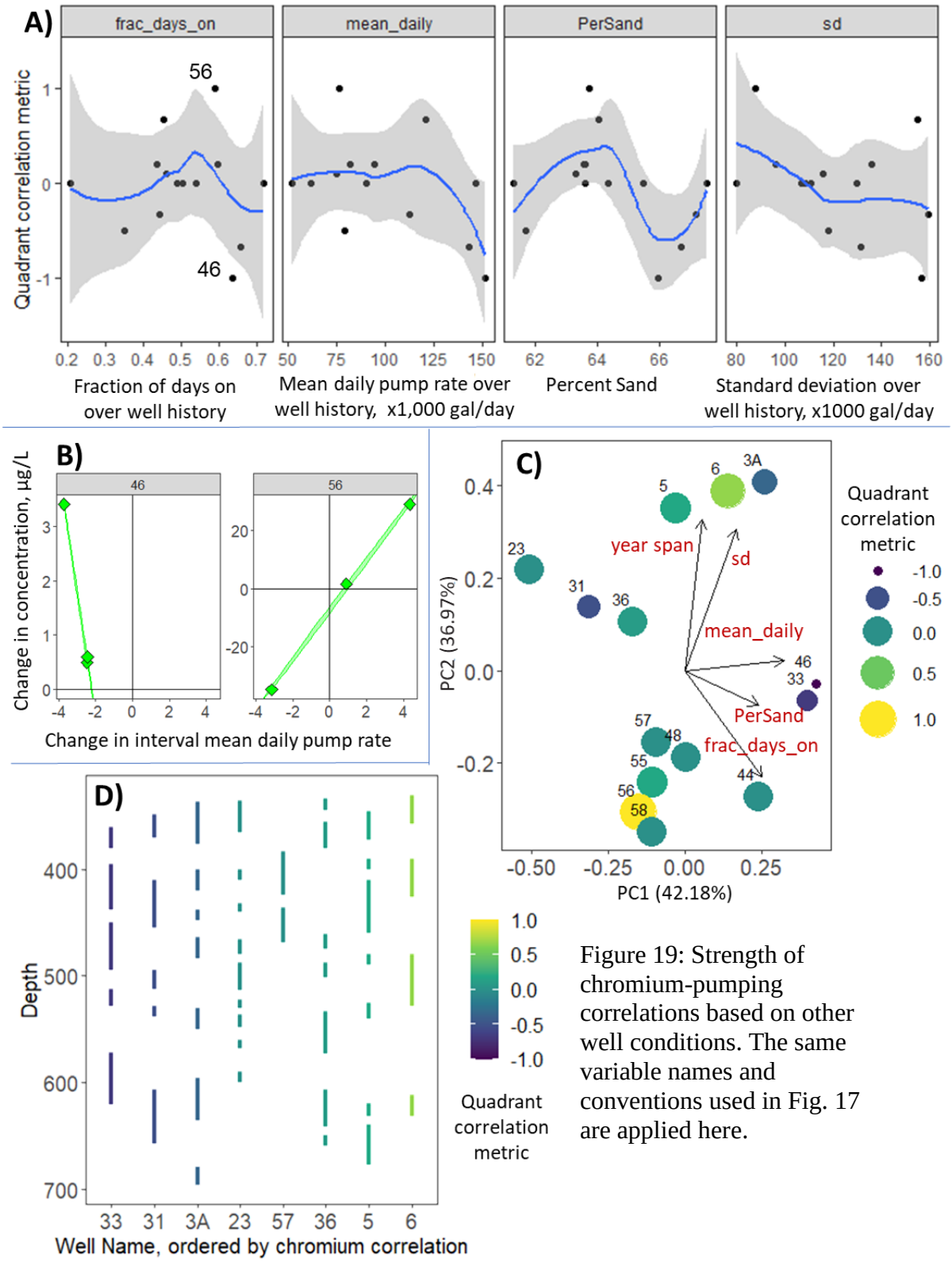


Figure 19: Strength of chromium-pumping correlations based on other well conditions. The same variable names and conventions used in Fig. 17 are applied here.

Positive correlations between chromium and pumping in wells with lower percent sand is consistent with the conceptual model for groundwater flow down wells, with the assumptions that (1) the down-well flowing groundwater during non-pumping conditions is undersaturated with respect to dolomite and (2) that native groundwater present in the sand bodies is saturated with respect to dolomite after extensive ion exchange and dolomite dissolution drives up the pH (Fig. 20, top). In this context, the high pH groundwater pulled through the zone of influence during pumping would not dissolve dolomite along the new flow paths into well screens (Fig. 20). Then, once the well is turned off, lower pH, dolomite-undersaturated groundwater would flow through the same area and dissolve dolomite, releasing manganese and oxidizing chromium in the process (Fig. 20). Low percent sand wells would facilitate more chromium release by prolonging the period of downward flow after each pumping occurrence, producing a positive correlation between pumping and chromium. The relatively narrow pumping frequency with positive pumping – chromium correlations supports this interpretation, as the well would have to be on often enough to generate downward hydraulic gradients but not so often that down-well flow and dolomite dissolution does not have time to occur.

As discussed above, dolomite-undersaturated downward-flowing groundwater is only expected extremely close to the recharge areas of the aquifer, as it becomes saturated soon after infiltration. Therefore, it is most likely in the unconfined portions of the aquifer in wells that are screened near the surface. An alternative mechanism for wells in the confined portion of the aquifer better fits the well frequency control on chromium-pumping correlations (Fig. 20). If we assume that the groundwater pulled into the well along different flow paths during active pumping conditions undergoes sufficient ion exchange to become undersaturated with respect to dolomite again, dolomite could dissolve during active pumping conditions and release chromium. In this conceptual model, down-well flow could not dissolve dolomite, as even the shallower, ‘lower pH’ waters in this region of the COA are near or at dolomite saturation. However, chromium concentrations in the COA are strongly depth dependent (Christenson et al., 1998). Therefore, this downward-flowing shallow groundwater would likely contain less chromium than the native groundwater and thus lower chromium concentrations during inactive well conditions.

This reduction in chromium during inactive periods and increase in chromium during active periods fits well with the well frequency (fract_days_on) control on chromium-pumping correlations. Since both increases and decreases in chromium require significant periods of pumping, it is logical that the lowest fraction of days on produces no correlation. From there, increasing the fraction of days on allows down-well flow to occur more frequently and reduces chromium concentrations every time the well is turned on, producing negative correlations between pumping and chromium. Further increasing the fraction of days on allows the active pumping-chromium mechanism to outweigh the inactive chromium mechanism, therefore increasing chromium concentrations and producing positive correlations between pumping and chromium. Interestingly, the wells with the highest observed pumping frequencies dip back to 0

or negative correlations between pumping and chromium. Well 46 does lie in the unconfined portion of the aquifer and has high percent sand, so its negative correlation with chromium, even at high fractions of days on, is still consistent with the first conceptual model. Well 33 is in the confined portion of the aquifer and, according to conceptual model 2, should have a positive or 0 correlation between pumping and chromium. It is possible that more low-Cr groundwater gets pulled into the well at high fractions of days on (and high pumping rates) and produces negative correlations between pumping and chromium.

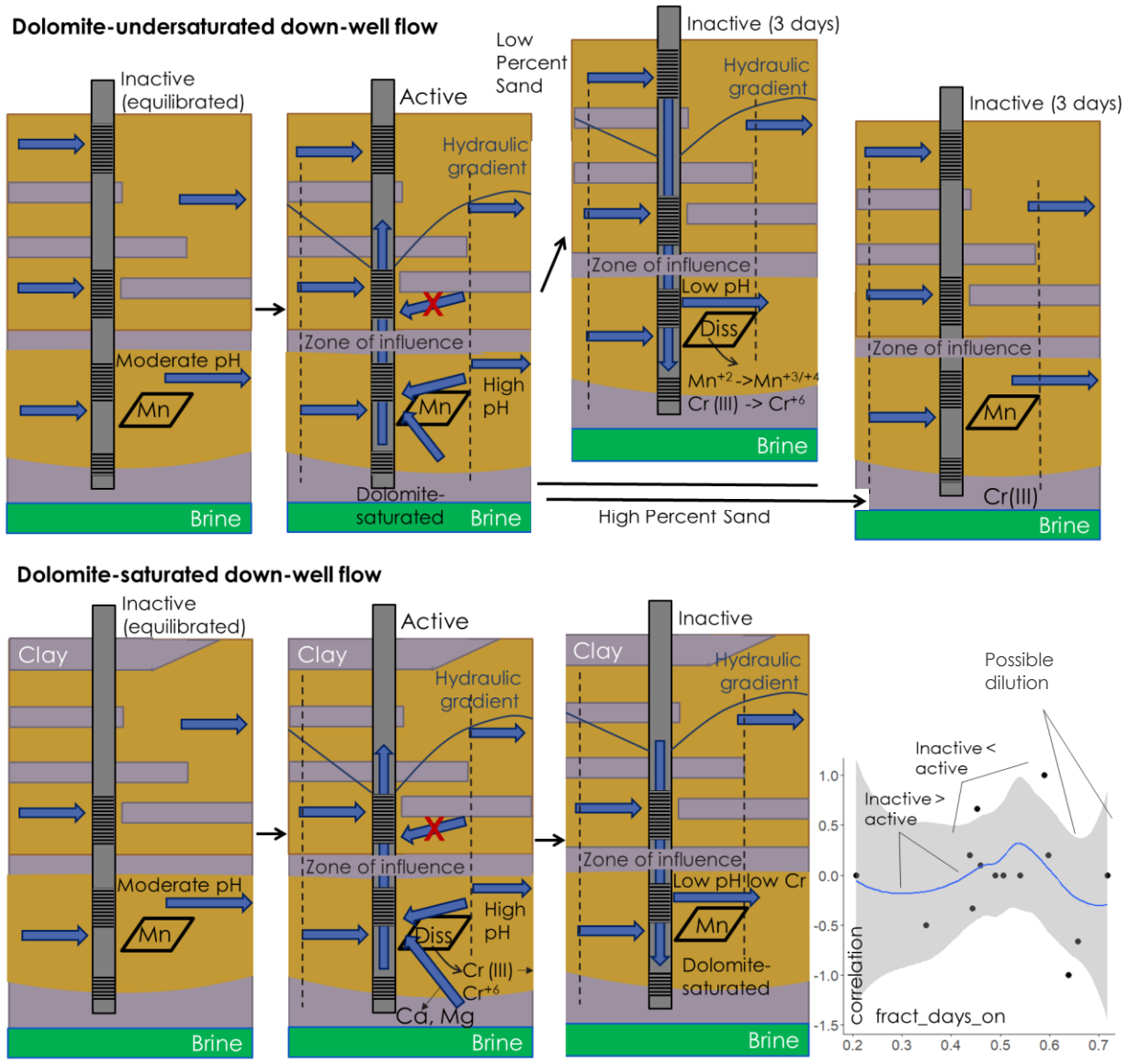


Figure 20: Conceptual models for chromium variation with two possible sets of dolomite saturation conditions. Top: Conceptual model for chromium-pumping variation with percent sand with down-well groundwater undersaturated with respect to dolomite, which would be most likely in the unconfined portion of the aquifer in wells with very shallow screens. Bottom: Conceptual model for chromium-pumping variation with well frequency, with down-well groundwater saturated with respect to dolomite.

Uranium

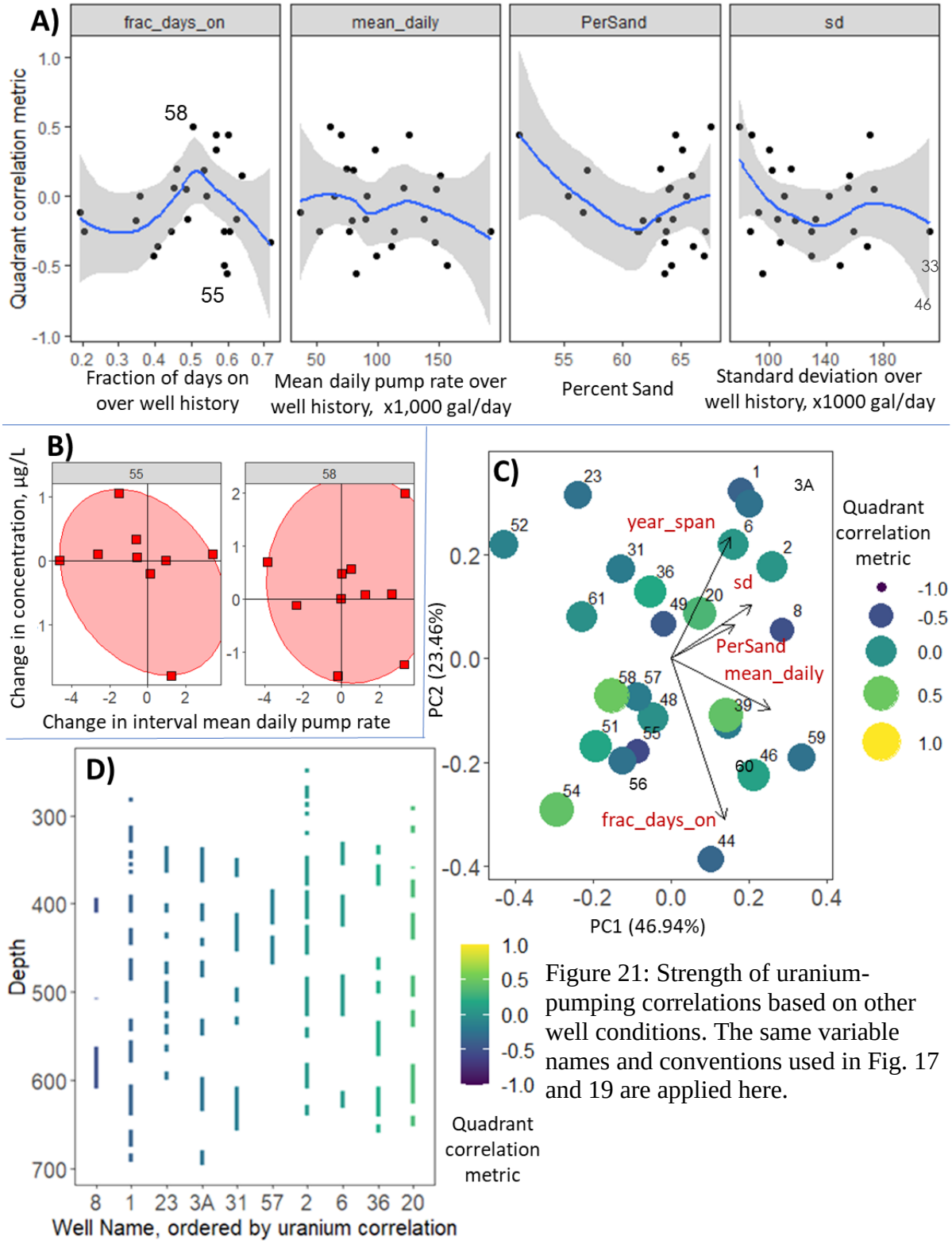


Figure 21: Strength of uranium-pumping correlations based on other well conditions. The same variable names and conventions used in Fig. 17 and 19 are applied here.

Uranium correlations are generally weaker than the arsenic and chromium correlations, which can be seen in the narrower range of y-values in Fig. 21a. and the cross plots in Fig. 21b. Like with chromium, positive uranium correlations only occur in wells that are on between ~45% to 65% of the time overall. Negative correlations between uranium and pumping only occur in high percent sand wells, while positive correlations between uranium and pumping occur with both high and low percent sands. There are no clear trends in the PCA plot between the quadrant correlations and any of the plotted variables.

Based on the conceptual model for groundwater flow during pumping/non-pumping conditions, pH is expected to generally rise during pumping, especially in the down-flow direction in the zone of influence (third panel in Fig. 22) as older groundwater gets pulled against the direction of groundwater flow and from other flow paths that only exist during pumping conditions. This will reduce uranium sorption onto iron oxides due to the more abundant uranium-carbonate complexes under the higher pH environment. When the well is subsequently turned off, assuming that the percent sand is sufficiently low to allow down-well flow for a sustained period, low pH groundwater will flow back down the well and more uranium will sorb back onto the iron oxides. The net increase or decrease in uranium will depend on the pumping frequency.

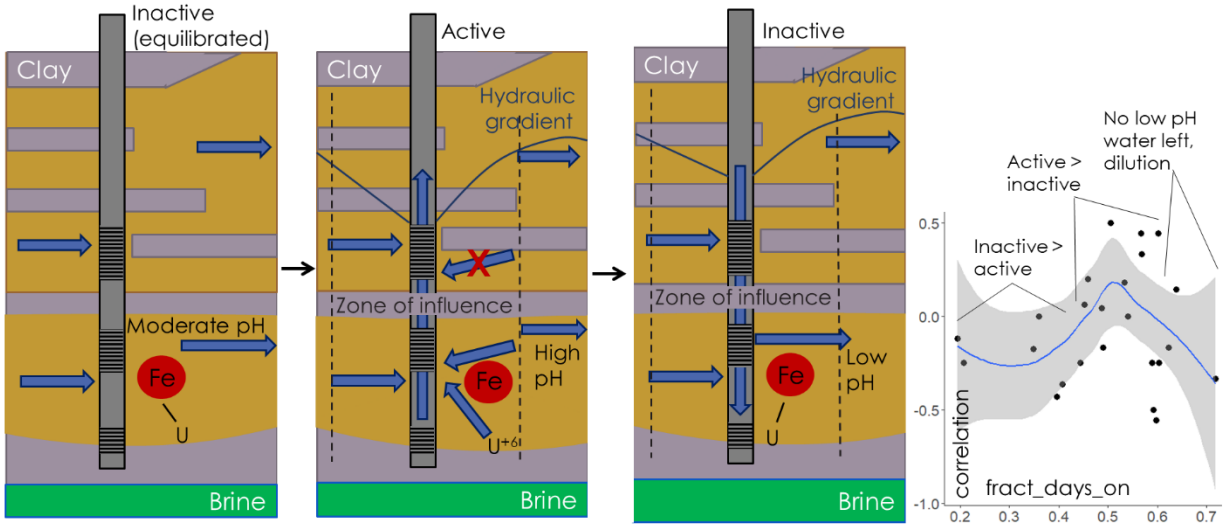


Figure 22: Conceptual model for uranium concentration changes with pumping

When the well is not on very often, any pumping instances would allow low pH water into the well (outweighing the pH increases during the actual pumping itself) and reduce the amount of uranium around the well. Thus, inactive flow would outweigh the active flow and produce negative correlations between pumping and uranium concentrations. Then, when the fraction of days on increases slightly, pumping conditions start to outweigh the down-well flow after pumping, leading to net uranium increase and thus positive correlations between pumping and uranium concentrations. Interestingly, further increases in the fraction of days on leads to negative correlations between uranium and pumping again. Positive and negative correlations both depend on pH oscillation in the well, so it's possible that too high of a pumping frequency

ends down-well flow and thus removes all the low pH water, removing the chance for positive correlations.

Shallow groundwater tends to be higher in Ca and Mg than the deep groundwater (Smith et al., 2009), so it is possible that calcium enrichment from down-well flow could reduce the predicted uranium sorption under lower pH conditions. Analyzing available NWIS data with above-detection uranium concentrations (Fig. 23) indicates that high calcium/sulfate concentrations in the COA facilitate some uranium mobility at lower pH/carbonate concentrations but have less effect at higher carbonate concentrations. High Ca/SO_4 also does not produce as high of uranium concentrations as high carbonate concentrations can. Therefore, the reduction in pH should play a larger role than the increase in Ca with the downward-flowing groundwater during inactive conditions and reduce uranium concentrations.

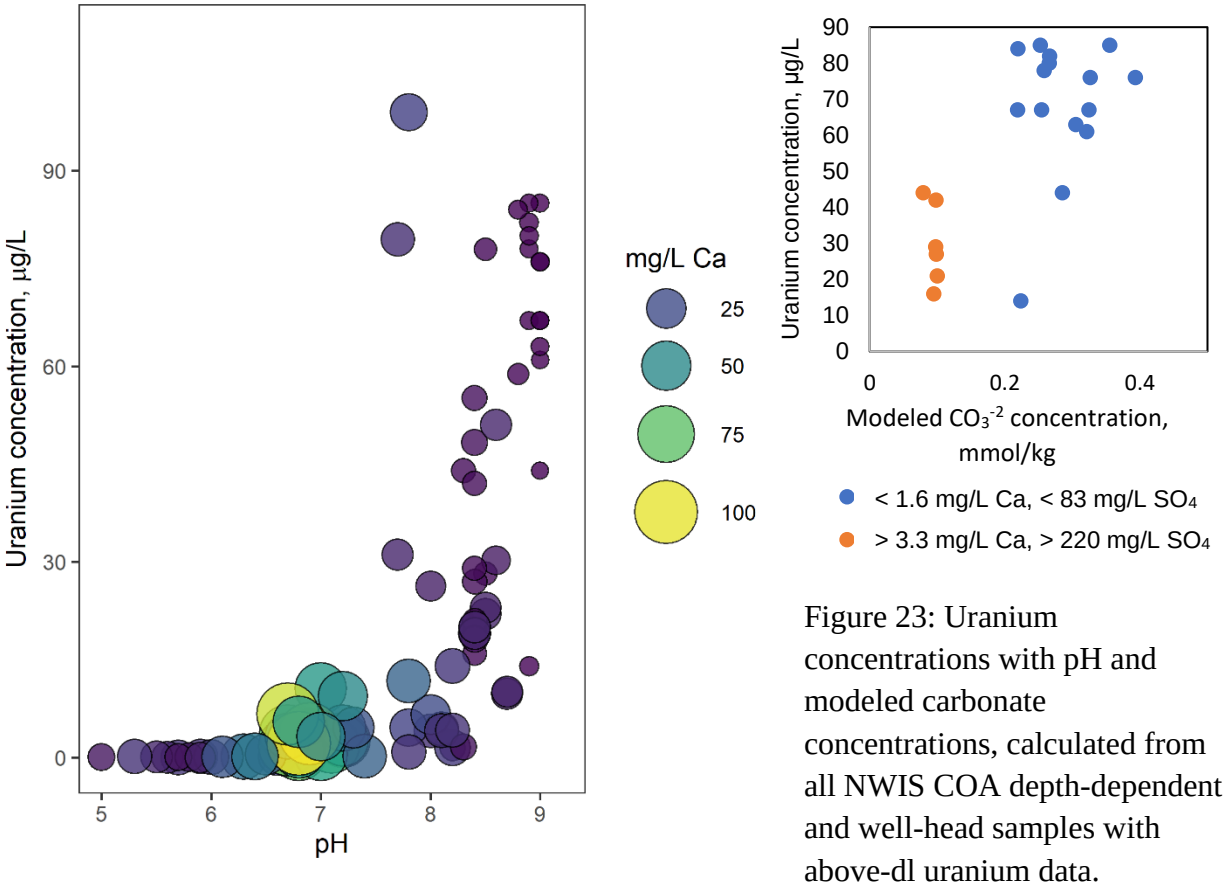


Figure 23: Uranium concentrations with pH and modeled carbonate concentrations, calculated from all NWIS COA depth-dependent and well-head samples with above-dl uranium data.

Hazard Maps

A positive correlation between pumping and trace metals only represents degrading water quality if the well is already located in an area with high trace metal concentrations. Therefore, the quadrant correlations used in Fig. 17, 19, and 21 are plotted with the FSAL, or the fraction of samples with concentrations greater than or equal to 10% below the EPA maximum contaminant limit (MCL) for each trace metal. The EPA MCLs are 10 µg/L for total arsenic, 30 for total uranium, 50 for total selenium, and 100 for total chromium (US EPA, 2015). There is some

debate about the chromium MCL because it accounts for total Cr instead of Cr(VI), which is the more toxic form of the metal, and because most of the human toxicity data used to set the limit comes from occupational inhalation rather than ingestion with drinking water (Moffat et al., 2018). Therefore, the World Health Organization (WHO) total chromium limit of 50 μg/L is used as well (World Health Organization, 2006).

Around the City of Norman, water quality generally improves from west to east for most trace metals except chromium, which has most samples close to or exceeding 50 μg/L near the center of Norman's well field (Figs. 24-26). Chromium rarely comes close to the EPA limit of 100 μg/L in well head samples, so the chromium map shows almost no issues with water quality (Fig. 25). These maps only include well-head samples, however, so samples from individual depths can (and do) reach higher concentrations (Tomlinson, 2019). Just because an area has a high FSAL number does not inherently mean that people with wells in those areas have unsafe water. The City of Norman combines water from wells with known water quality issues with other sources (surface water, other wells, etc.) which get concentrations back within safe levels before water gets to the public (Rachel Croft, personal communication, 23 October 2020).

The maps for water quality over the entire COA generally show most issues with water quality on the western, confined portion of the aquifer as well (Appendix III). Uranium notably has more samples close to its high-concentration limit on the eastern and southeastern portions of the aquifer, at least for well-head samples. This could relate back to the importance of flow paths for uranium, as longer, shallow flow paths from west to east lead to higher Ca concentrations and can mobilize some uranium in conditions with otherwise low carbonate. Alternatively, the uranium distribution could again relate to spatial variation in uranium in the sediment, which could be variation in uranium form, abundance, redox conditions, etc. Although uranium has the highest concentrations in this area, arsenic also presents some water quality issues in the eastern portion of the aquifer. Combining the hazard maps with individual well trends, we make recommendations about wells that are at highest and lowest risk for degraded water quality under higher pumping conditions in the implications section below.

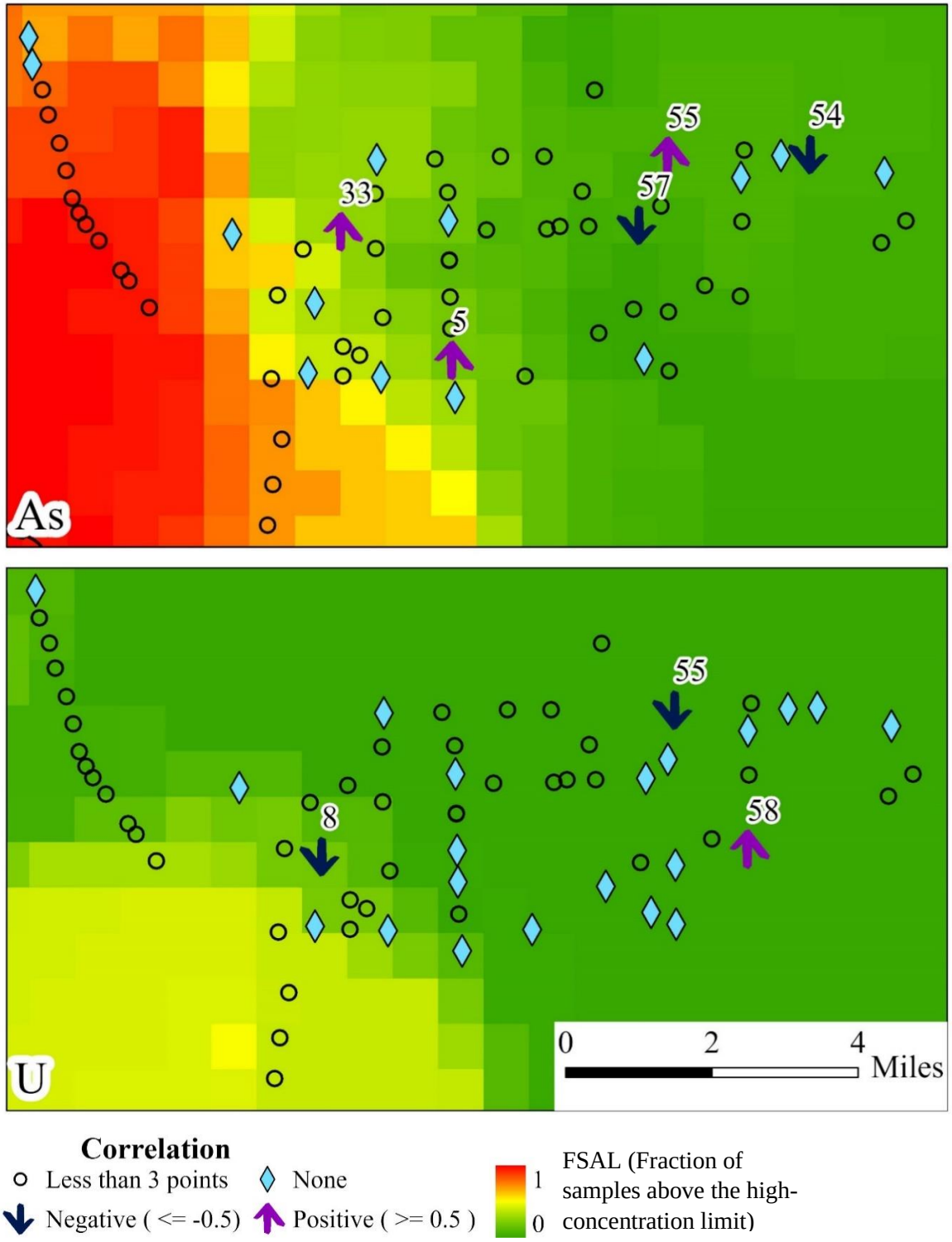


Figure 24: Arsenic and uranium hazard for the Norman well field with pumping correlations. The quadrant correlation metric is the same used in Fig. 17. Wells without needed data for robust correlation metric analysis are shown with circles. Wells with sufficient data that have correlations between -0.5 and 0.5 ("weaker" correlations) are shown as blue diamonds. The high-concentration limit for arsenic is 9 μg/L and the high-concentration limit for uranium is 27 μg/L.

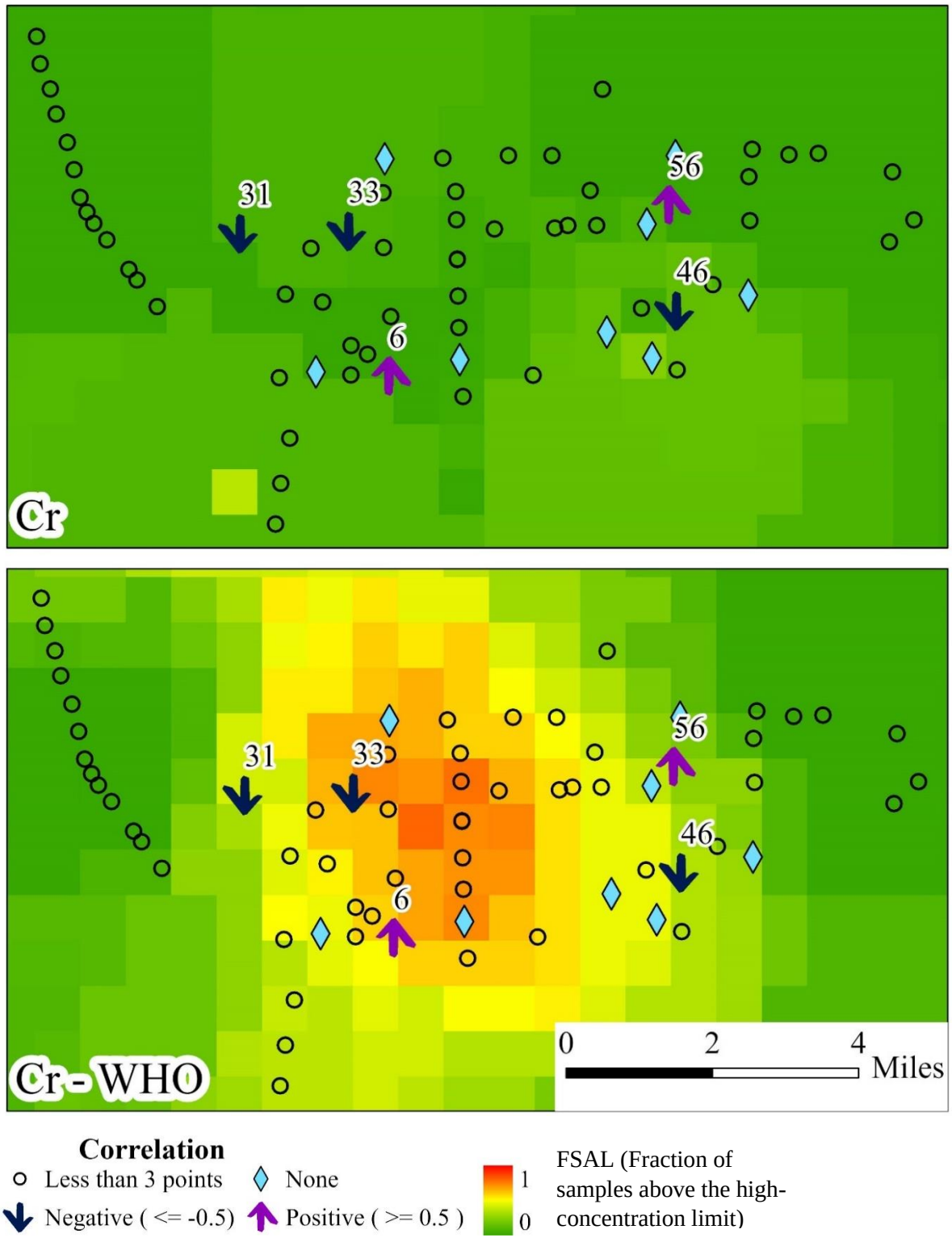


Figure 25: Chromium hazard for the Norman well field with pumping correlations. The top panel uses the EPA MCL of 100 µg/L (a high-concentration limit of 90 µg/L) for total Cr; the lower panel uses the WHO recommended MCL of 50 µg/L for total Cr (a high-concentration limit of 45 µg/L).

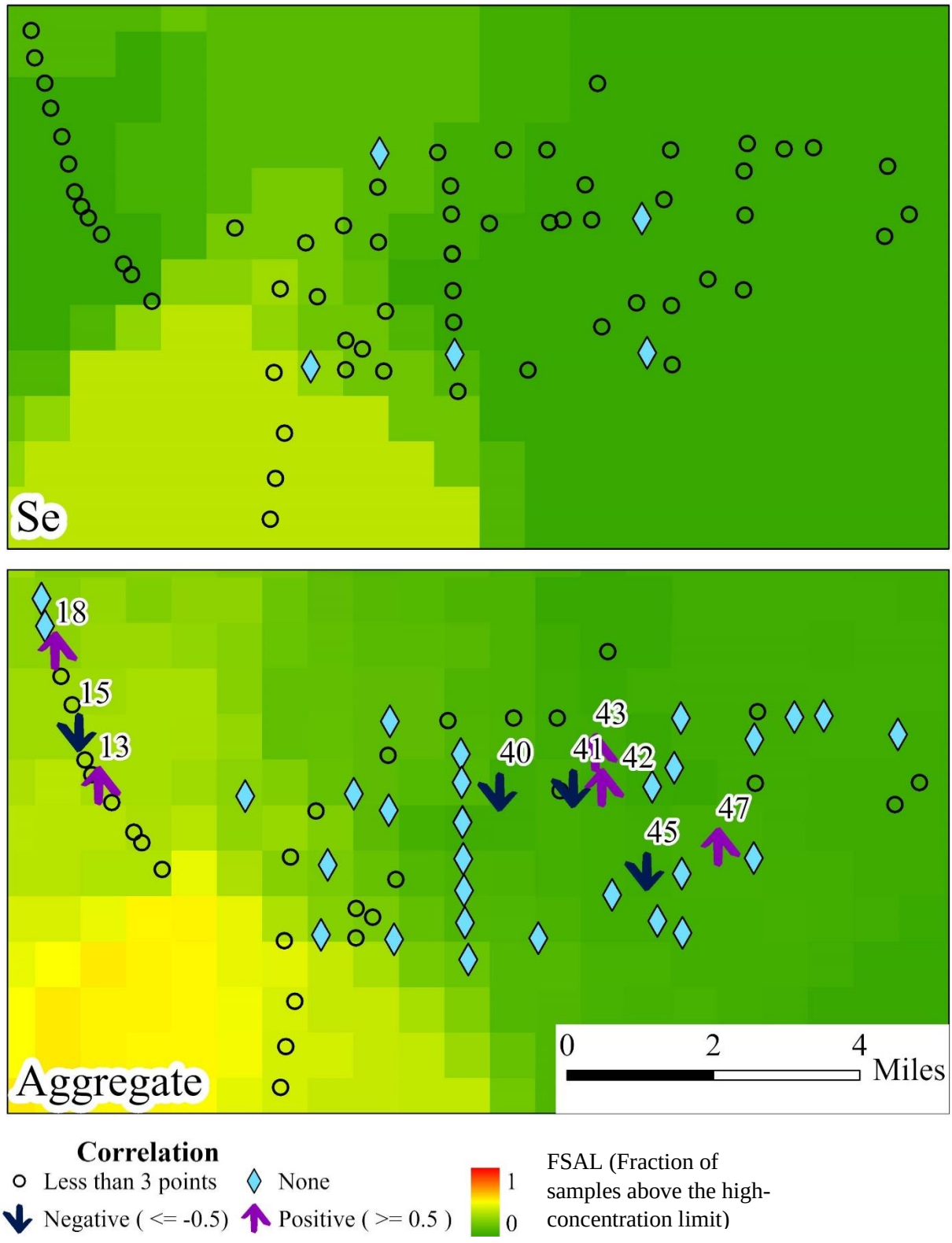


Figure 26: Selenium and aggregate trace metal hazard for the Norman well field with pumping correlations. The selenium high-concentration limit is 45 µg/L. The aggregate trace metal hazard

map displays the fraction of arsenic, chromium, uranium, and selenium samples that are above their high-concentration limits. The EPA 100 μg/L limit for chromium was used; with the WHO guidelines instead of the EPA guidelines, the cells in the center of the aggregate map especially would have higher values.

Depth-specific maps of aggregate trace metal hazard were also constructed to determine most and least hazardous well screen locations for water quality. As in Mashburn et al. (2018), the maps were split into 100 ft deep layers, with 1000 m by 1000 m raster cells. There were few samples for each depth class, but general trends could still be determined (Appendix III). No major issues for water quality began until the 400 to 500 ft screen range, where the FSAL exceeded 0.5 in a narrow region around well 23. Other than that region, most of the rest of the Norman well field stayed below a FSAL of 0.25 from 400 to 500 ft. The 500 to 600 ft screen range showed a larger band of moderately impacted water quality in western Norman (FSAL > 0.5), still centered around well 23. By 600-700 ft, there were fewer samples and the majority of the well field has an FSAL close to 0. Most of the depth dependent data used came from depth – dependent sampling as part of the USGS/EPA arsenic investigation discussed in Smith et al. (2009), mostly in west Norman. Ultimately, the lack of water quality issues with depth in eastern Norman is likely due to a combination of the lack of data and the confining layer over western Norman making generally longer flow paths.

Conclusions

Analysis and modeling of Central Oklahoma Aquifer data aggregated from a variety of sources demonstrates a complex array of well parameters and water flow conditions which may dictate how any given well responds to increased pumping. The trends discussed cannot robustly be used to make any accurate prediction of how a well will respond in the future, as they reflect past patterns in an incompletely understood system. Any change in that system which was not considered here could easily change a well's behavior with respect to trace metals and pumping. Outside of the incomplete assumptions, the aggregation of data from different sources, each with its own procedures, instruments, reporting methods, and parameters, can produce artificial variation in the data. Consider two instruments, one that consistently reads slightly higher than the other instrument (within the calibration threshold) for the same parameter. Switching instruments could yield a false change in measurement which was, in reality, due to a change in instrument.

Many statistically significant relationships were still found in the data, which at the very least can show where increases in pumping is more likely to increase trace metal concentrations based on past behavior. Simple linear models between individual parameters could not describe the trends and variability across the COA or even the Norman well field, however, so even these statistically significant relationships are only valid when additional conditions are met. Despite very limited predictive capabilities, the results shown here do provide several significant findings which can shed some light on the complex interplay between anthropogenic and natural sources of variation in aquifer chemistry. A series of hypotheses were evaluated based on the existing

work, which may then be tested in future carefully-designed experiments and with other/larger data sets.

Hypothesis 1: Low pH groundwater flows down the well under non-pumping conditions. Then, when the well pumps, higher pH water gets pulled into the area of well influence.

There is substantial evidence to support this hypothesis. The Kendall's Tau correlation coefficients showed that a strong, statistically significant relationship exists between pH and pumping; more pumping generates higher pH values (Table 6). Additionally, the pH of the western portion of Norman's well field generally increased with pumping based on the Piper plot results (Fig. 12). Lastly, the pH change in the long-term well head models matched the sign of the change in interval mean pumping rate in all but one set of samples (Table 2).

Hypothesis 2: Arsenic and selenium concentrations increase with pumping because the pH increase that comes with pumping desorbs them from iron oxides in the COA.

The wells selected for the well-head modeling results (23, 5, 31, and 36) generally support this hypothesis (Fig. 13). There is some data to support this hypothesis for other wells too, but only for arsenic. Selenium was simply not reported frequently enough to make any accurate assessment of how it responds with pumping. The modeling data and conceptual models for trace metal mobilization suggest that they should behave similarly, but the trace metal/pumping rate cross plots show that this is not the case for at least well 48 (Fig. 16).

For arsenic to show a positive correlation with pumping, the well must have high percent sand and moderate pumping volumes. If the percent sand is too low, high hydraulic gradients produced whenever the well is turned on would facilitate more downward flow, and the As sorption during this downward flow would outweigh the As desorption during actual pumping. Too high of pump volumes would simply obscure the correlations due to high volumes of water from many flow paths.

Hypothesis 3: Uranium will increase with pumping in deep wells because increased carbonate concentrations prevent it from sorbing onto iron oxides.

Well 23 data weakly support this hypothesis because overnight periods between well-head samples reduced uranium concentrations and concentrations increased again throughout each day of sampling (and well pumping). Other wells in the wellfield show that positive correlations between uranium and pumping do occur, but only in wells with a relatively narrow pumping frequency between 45 and 60%. If the pumping frequency is too low, the decreased carbonate concentrations from down-well flow outweigh the increases in carbonate from pumping itself and there is a negative correlation between uranium and pumping. If the pumping frequency is high, there is not enough reduction in uranium (or pH) during down-well flow to generate uranium increases with pumping.

Hypothesis 4: Chromium will increase with pumping in wells if the downward flowing solution is already saturated with respect to dolomite but may decrease with pumping if the downward flowing solution is not saturated with respect to dolomite.

It is impossible to determine the validity of this hypothesis with the given data because it is difficult to determine whether or not the hypothetical downward-flowing solution is saturated with respect to dolomite. The correlations themselves cannot be used as evidence for either part of the hypothesis because most observed trends (in almost any well) could fit into either hypothesis. There are conditions that could produce negative correlations between chromium and pumping even if the downward-flowing solution is saturated with respect to dolomite (low pumping frequency) and conditions that could produce positive correlations between chromium and pumping even if the downward flowing solution is not saturated with respect to dolomite (high percent sand).

Other assumptions

All four hypotheses rely on the assumption that observed trace metal trends with pumping in the COA cannot be attributed to simple variations in mixing coefficients between screened intervals. The lack of model solutions for the mixing models, lack of correlation between mixing model success and number of incorporated well screens, and the definite evidence of dolomite dissolution in some of the well-head models are all consistent with the assumption that mixing alone cannot account for changes in well chemistry with time.

Future research

The modeling results were somewhat limited by lack of knowledge about sulfate fate and transport through the aquifer system, with all three failed well-head models having sulfate imbalances. Smith et al. (2009) discuss the importance of contributions from sulfate-bearing rocks in the Hennessy group, but the degree of communication between the Hennessy and the aquifer is unknown, along with the potential impacts on trace metal concentrations.

Another way to improve the meaningfulness of the modeling results would be to improve the data for the mixing models. The only way to truly test changing mixing fractions in response to aquifer pumping would be to set up controlled experiments where samples are taken from each of the well screens both before and after a fixed period of pumping, along with a wellhead sample. In order for variations in trace metals to truly be from mixing fraction changes only, the depth dependent solutions would be nearly constant from before to after the period of pumping and both could generate the wellhead sample. The depth-dependent USGS data used in this study simply did not have enough sample points to truly test this idea. Taking samples from the first well screen for each well and modeling how saturated they are with respect to dolomite is also the only way to determine the validity of hypothesis 4.

The last major question posed by the PHREEQC models was how such large changes in trace metal concentrations could occur in well 23 without the accompanying pH changes. Even though the changes in trace metal concentrations line up with some other factors associated with the

conceptual model (namely dolomite phase change and change in pumping activity), pH is thought to be the main control on arsenic and selenium mobilization, so without the pH change, it is still unknown how dolomite phase change or change in pumping activity could have any effect on those trace elements.

The study could also benefit from increased camera work to determine missing well construction information. Not all wells have total depths listed, for example, and even less have screen information. Additional camera work on missing wells could facilitate better understanding of how well depth (and screen depth more specifically) could generate more positive or negative chromium/uranium correlations stated in the hypotheses. There is already some qualitative evidence pointing towards these relationships, but more well construction data would make any conclusions much more robust.

Implications

Using the maps in Figs. 24-26, it is possible to predict which wells are most and least eligible for increased pumping to meet rising water demand in Norman, OK. For arsenic, wells 33 and 5 are the least suitable for increased future pumping because they both have positive correlations between arsenic concentration and pumping rate, plus fall into a region where some of the samples have high arsenic. Wells 57 and 54 are the most suitable because they have negative arsenic-pumping correlations and are in a lower-arsenic portion of the well-field (Fig. 24). Based on the available uranium data, increased pumping would not put well water at risk for high uranium concentrations. No wells in the region with heightened uranium had strong positive correlations with pumping (Fig. 24). The same is true for selenium (Fig. 26). Increased pumping in wells 6 and 56 is most likely to cause water quality problems with chromium, while 31 and 46 would be the safest for increased pumping (Fig. 25). Based on aggregate overall trace metals (Fig. 26), wells 40, 41, and 45 are most eligible for increased pumping. Even through wells 42, 43, and 47 show positive trace metal-pumping correlations, the risk of high trace metals from a regulatory standpoint would still be low because of generally low trace metal concentrations in the region. Wells 13 and 18 (which are no longer active) would show greatest risk of increased trace metal concentrations with pumping.

Outside of the strict regulatory sense for the City of Norman, this study can also be used to make some general predictions about trace metal-pumping correlations elsewhere in the COA. Wells in the entire western, confined portion of the COA would likely show similar arsenic patterns as wells under similar conditions in west Norman. Lower conductivity wells might actually experience negative correlations with pumping, but sufficiently high conductivities and moderate pumping rates would likely produce positive correlations with pumping (Appendix III). Uranium is the other main trace metal that frequently reaches concentrations near or above its 30 μg/L EPA limit, mostly in the eastern COA. Even low carbonate concentrations could produce 30-40 μg/L uranium under high Ca concentrations (Fig. 23), which is likely the case here. Assuming the system is similar to that around Norman, this region is most likely to produce negative

correlations with pumping because it reflects mobilization under static, non-pumping conditions. Why this is the case is still somewhat unknown for deeper wells. Selenium and chromium do not show elevated concentrations other than around Norman, which may actually be a testament to the proposed chromium mobilization along deep wells. Lastly, wells that pump between 45 and 60% of the time are most likely to have positive uranium and chromium correlations with pumping; avoiding that pumping frequency can help reduce risk of elevating uranium and chromium concentrations.

Outside of the COA, the methods and calculation metrics used here can be applied to other aquifer systems to determine other relationships between pumping and trace metal concentrations. The importance of calcium in uranium mobilization at lower pH can be used and studied in other aquifer systems. Also, the processing methods especially were built to handle a wide range of data, from municipal EPA compliance reports to complete metal and anion samples from the USGS. To apply the methods developed in this thesis to other aquifer systems, the only requirements are pumping records and sample records with dates/times. If there is not enough pumping data to cover the entire period between samples and the same period of time before the first sample like the pumping metric used here, the weekly pumping conditions leading up to each sample can still yield meaningful results with the Kendall's Tau correlation analysis.

References

Ayotte, J.D., Belaval, M., Olson, S.A., Burow, K.R., Flanagan, S.M., Hinkle, S.R., Lindsey, B.D., 2015. Factors affecting temporal variability of arsenic in groundwater used for drinking water supply in the United States. Sci. Total Environ. 505, 1370–1379. https://doi.org/10.1016/j.scitotenv.2014.02.057

Ayotte, J.D., Szabo, Z., Focazio, M.J., Eberts, S.M., 2011. Effects of human-induced alteration of groundwater flow on concentrations of naturally-occurring trace elements at water-supply wells. Appl. Geochem. 26, 747–762. https://doi.org/10.1016/j.apgeochem.2011.01.033

Baird, R.B., Eaton, A.D., Rice, E.W., 2018. 2340 HARDNESS, in: Standard Methods For the Examination of Water and Wastewater, Standard Methods for the Examination of Water and Wastewater. American Public Health Association. https://doi.org/10.2105/SMWW.2882.025

Becker, C., 2013. Groundwater Quality and the Relation Between pH Values and Occurrence of Trace Elements and Radionuclides in Water Samples Collected from Private Wells in Part of the Kickapoo Tribe of Oklahoma Jurisdictional Area, Central Oklahoma, 2011 (Scientific Investigations Report No. 2012–5253), Scientific Investigations Report. USGS.

Belt, K., Paxton, S.T., 2005. GIS as an aid to visualizing and mapping geology and rock properties in regions of subtle topography. GSA Bull. 117, 149–160. https://doi.org/10.1130/B25463.1

Breit, G. N., 1998. The Diagenetic History of Permian Rocks in the Central Oklahoma Aquifer. In Ground-Water-Quality Assessment of the Central Oklahoma Aquifer, Oklahoma: Results of Investigations. U.S. Geological Survey, Water-Supply Paper 2357-A.

Brown, C., Colabufo, S., Coates, J., 2002. Aquifer geochemistry and effects of pumping on ground-water quality at the Green Belt Parkway Well Field, Holbrook, Long Island, New York (Water-Resources Investigations Report 2001-4025), Scientific Investigations Report. https://doi.org/10.3133/wri014025

Carucci, V., Petitta, M., Aravena, R., 2012. Interaction between shallow and deep aquifers in the Tivoli Plain (Central Italy) enhanced by groundwater extraction: A multi-isotope approach and geochemical modeling. Appl. Geochem. 27, 266–280. https://doi.org/10.1016/j.apgeochem.2011.11.007

Christenson, S., Havens, J. S., 1998. Ground-Water-Quality Assessment of the Central Oklahoma Aquifer, Oklahoma: Results of Investigations. U.S. Geological Survey, Water-Supply Paper 2357-A.

City of Norman, 2018. City of Norman Public Utility Report 2018. https://www.normanok.gov/sites/default/files/documents/2020-06/Current-NUA-Report.pdf (accessed 29 December, 2020).

Clark, B.R., Landon, M.K., Kauffman, L.J., Hornberger, G.Z., 2008. Simulations of Ground-Water Flow, Transport, Age, and Particle Tracking near York, Nebraska, for a Study of Transport of Anthropogenic and Natural Contaminants (TANC) to Public-Supply Wells (Scientific Investigations Report No. 2007–5068), Scientific Investigations Report. USGS.

de Graaf, I.E.M., Gleeson, T., van Beek, L.P.H. Rens, Sutanudjaja, E.H., Bierkens, M.F.P., 2019. Environmental flow limits to global groundwater pumping. Nature 574, 90-94,94A-94N. http://dx.doi.org.ezproxy.lib.ou.edu/10.1038/s41586-019-1594-4

Fendorf SE, Zasoski RJ (1992) Chromium(III) oxidation by δ-MnO2. 1. Characterization. Environ Sci Technol 26:79–85

Haack, S. K. & Luukkonen, C. L. Relation Between Organic-Wastewater Compounds, Groundwater Geochemistry, and Well Characteristics for Selected Wells in Lansing, Michigan. U.S. Geological survey, Scientific Investigations Report 2013–5139.

Helsel, D., Hirsch, R., Ryberg, K., Archfield, S., Gilroy, E., 2020. Statistical Methods in Water Resources: Chapter 3 of Section A, Statistical Analysis, in: Book 4: Hydrologic Analysis and Interpretation. U.S. Geological Survey.

Hudak, P.F., 2018. Associations between Dissolved Uranium, Nitrate, Calcium, Alkalinity, Iron, and Manganese Concentrations in the Edwards-Trinity Plateau Aquifer, Texas, USA. Environ. Process. 5, 441–450. https://doi.org/10.1007/s40710-018-0296-5

Ji-Zheng, H., Meng, Y.-T., Zheng, Y.-M., Zhang, L.-M., 2010. Cr(III) oxidation coupled with Mn(II) bacterial oxidation in the environment. J. Soils Sediments 10, 767–773. https://doi.org/10.1007/s11368-009-0139-0

Kuells, C., Adar, E.M., Udluft, P., 2000. Resolving patterns of groundwater flow by inverse hydrochemical modelling in a semiarid Kalahari basin. Tracers and Modelling in Hydrogeology, in *Proceedings of the TraM'2000 Conference,* Liège, Belgium, May 2000. IAHS Publ. no. 262.

Li, J., Heap, A.D., 2014. Spatial interpolation methods applied in the environmental sciences: A review. Environmental Modelling & Software 53, 173–189. https://doi.org/10.1016/j.envsoft.2013.12.008

Li, J., Heap, A.D., 2008. *A Review of Spatial Interpolation Methods for Environmental Scientists.* Geoscience Australia, Record 2008/23, 137 pp.

Mahapatra, S.R., Venugopal, T., Shanmugasundaram, A., Giridharan, L., Jayaprakash, M., 2020. Heavy metal index and geographical information system (GIS) approach to study heavy metal contamination: a case study of north Chennai groundwater. Appl. Water Sci. 10, 238. https://doi.org/10.1007/s13201-020-01321-0

Mashburn, S.L., Ryter, D.W., Neel, C.R., Smith, S.J., Correll, J.S., 2018. Hydrogeology and Simulation of Groundwater Flow in the Central Oklahoma (Garber-Wellington) Aquifer, Oklahoma, 1987 to 2009, and Simulation of Available Water in Storage, 2010–2059 (Scientific Investigations Report No. 2013–5219), Scientific Investigations Report. USGS.

Mekonnen, M.M., Hoekstra, A.Y., 2016. Four billion people facing severe water scarcity. Sci. Adv. 2, e1500323. https://doi.org/10.1126/sciadv.1500323

Mitchell, E., Frisbie, S., Sarkar, B., 2011. Exposure to multiple metals from groundwater—a global crisis: Geology, climate change, health effects, testing, and mitigation. Metallomics 3, 874–908. https://doi.org/10.1039/C1MT00052G

Moffat, I., Martinova, N., Seidel, C., Thompson, C.M., 2018. Hexavalent Chromium in Drinking Water. J. AWWA 110, E22–E35. https://doi.org/10.1002/awwa.1044

Morales, K.H., Ryan, L., Kuo, T.-L., Wu, M.-M., Chen, C.-J., 2000. Risk of Internal Cancers from Arsenic in Drinking Water. Environ Health Perspect 108, 655–661. https://doi.org/10.1289/ehp.00108655

Nico, P.S., Zasoski, R.J., 2000. Importance of Mn(III) availability on the rate of Cr(III) oxidation on δ-MnO2. Environ Sci Technol l34:3363–3367

Oklahoma Geological Survey, 2015. Dowell Brine Database [WWW Document]. *Brine data.* URL http://www.ou.edu/ogs/data/brinedata.html (accessed 3.20.21).

Parkhurst, D.L., Appelo, C.A.J., 2013. Description of Input and Examples for PHREEQC Version 3—A Computer Program for Speciation, Batch-Reaction, One-Dimensional Transport, and Inverse Geochemical Calculations, in: Book 6, Modeling Techniques, Techniques and Methods. USGS.

Parkhurst, D. L. Christenson, S., Breit, G. N., 1996. Ground-Water-Quality Assessment of the Central Oklahoma Aquifer, Oklahoma-Geochemical and Geohydrologic Investigations. U.S. Geological Survey, Water-Supply Paper 2357.

Pichler, T., Renshaw, C.E., Sültenfuß, J., 2017. Geogenic As and Mo groundwater contamination caused by an abundance of domestic supply wells. Appl. Geochem., Environmental and Health Roles of Geogenic Arsenic 77, 68–79. https://doi.org/10.1016/j.apgeochem.2016.03.002

Riedel, T., Kübeck, C., 2018. Uranium in groundwater – A synopsis based on a large hydrogeochemical data set. Water Res. 129, 29–38. https://doi.org/10.1016/j.watres.2017.11.001

Rodell, M., Famiglietti, J.S., Wiese, D.N., Reager, J.T., Beaudoing, H.K., Landerer, F.W., Lo, M.-H., 2018. Emerging trends in global freshwater availability. Nature 557, 651-659,659A-659M. http://dx.doi.org.ezproxy.lib.ou.edu/10.1038/s41586-018-0123-1

Rovira, M., Giménez, J., Martínez, M., Martínez-Lladó, X., de Pablo, J., Martí, V., Duro, L., 2008. Sorption of selenium(IV) and selenium(VI) onto natural iron oxides: Goethite and hematite. J. Hazard. Mater. 150, 279–284. https://doi.org/10.1016/j.jhazmat.2007.04.098

Silvester E, Charlet L, Manceau A (1995) Mechanism of chromium (III) oxidation by Na-buserite. J Phys Chem 99:16662–16669

Smith, S.J., Paxton, S.T., Christenson, S., Puls, R.W., Greer, J.R., 2009. Flow Contribution and Water Quality with Depth in a Test Hole and Public-supply Wells: Implications for Arsenic Remediation through Well Modification, Norman, Oklahoma, 2003-2006. EPA.

Tortorelli, R.L., 2009. Water Use in Oklahoma 1950–2005 (Scientific Investigations Report No. 2009–5212), Scientific Investigations Report. USGS.

Tomlinson, Z. T., Maples, B. L., Dee, K. T., Elwood Madden, A. S., 2019. Applying clustering methods to hydrologic data from the Central Oklahoma Aquifer (COA) to better spatially constrain changes in water chemistry. In *Oklahoma Governor's Water Conference and Research Symposium,* Midwest City, OK, Dec. 4, 2019.

US EPA, 2015. National Primary Drinking Water Regulations. https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations (accessed 22 October, 21).

U.S. Geological Survey, 2019. Specific conductance: U.S. Geological Survey Techniques and Methods, book 9, chap. A6.3, 15 p., https://doi.org/10.3133/tm9A6.3. [Supersedes USGS Techniques of Water-Resources Investigations, book 9, chap. A6.3, version 1.2.]

Wilkins, K., 2019. Hydrologic Survey and Simulation of Water Available in Storage. Oklahoma Water Resources Board. https://www.owrb.ok.gov/studies/groundwater/pdf/Garber-Wellington-Staff-Report-2019.pdf (accessed 21 January 2021).

Westrop, Jeffrey & Weeks, Brett & Hu, Qin-Hong & Gallagher, Brian & Swindle, Andrew & Elwood Madden, Andrew. (2016). Manganese-Bearing Dolomite Dissolution Drives Hexavalent Chromium Occurrence in the Central Oklahoma Aquifer. In *GSA Annual Meeting,* Denver, CO*, Sept. 28, 2016.* https://doi.org/10.1130/abs/2016AM-281658

World Health Organization, 2006. Guidelines for drinking-water quality: first addendum to the third edition, volume 1: recommendations. Geneva: WHO. https://www.who.int/water_sanitation_health/dwq/gdwq0506.pdf

Xing, L., Guo, H., Zhan, Y., 2013. Groundwater hydrochemical characteristics and processes along flow paths in the North China Plain. J. Asian Earth Sci. 70–71, 250–264. https://doi.org/10.1016/j.jseaes.2013.03.017

Yadav, K.K., Gupta, N., Kumar, V., Choudhary, P., Khan, S.A., 2018. GIS-based evaluation of groundwater geochemistry and statistical determination of the fate of contaminants in shallow aquifers from different functional areas of Agra city, India: levels and spatial distributions. RSC Adv. 8, 15876–15889. https://doi.org/10.1039/C8RA00577J

Yang, X., Hodler, T., 2000. Visual and Statistical Comparisons of Surface Modeling Techniques for Point-based Environmental Data. Cartography and Geographic Information Science 27, 165–176. https://doi.org/10.1559/152304000783547911

Appendix I: R processing procedure

General City of Norman data

The City of Norman keeps a dataset of general parameters that are regularly sampled; the parameters used here are total and Ca hardness, M (total) alkalinity, pH, conductivity, TDS, temperature, fluoride, and chloride concentrations. Total and Ca hardness were used to estimate Ca and Mg concentrations, and conductivity was converted to specific conductance. The other parameters were directly used in the correlation analysis.

According to *Standard Methods for the Examination of Water and Wastewater* by the American Public Health Association, the American Water Works Association, and the Water Environment Federation, total hardness can be estimated as the sum of the Mg and Ca concentrations in water after converting the cation concentrations to mg/L as $CaCO_3$ with the formula below (Baird et al., 2018).

Hardness, mg/L as $CaCO_3$ = 2.497 * Ca concentration, mg/L + 4.118 * Mg concentration, mg/L

To test the validity of this approximate formula, NWIS groundwater quality data was collected from across the COA with hardness, Ca, and Mg concentrations (USGS parameter codes 00900, 00915, and 00925 respectively). One measurement with 0 mg/L Mg was removed; 1048 samples were used. Then, hardness was calculated using the formula above and compared to the actual hardness to compute percent error (Fig. I-1). This assumes that the USGS hardness parameter 00900 is directly measured rather than calculated. The median percent error was -0.20%, with a 95% median confidence interval of -0.22% to -0.17%. The confidence interval was calculated using the binomial probability distribution as shown in the USGS publication *Statistical Methods for Water Resources* (Helsel et al., 2020). Additionally, Mg concentration was calculated from the formula and compared to the measured Mg concentration (Fig I-1). This produced a median percent error of 0.42%, with a 95% median confidence interval of 0.38% to 0.49%. Because of the very low percent error, the approximate formula was used to first calculate Ca concentration (using the formula Ca, mg/L = Ca hardness as $CaCO_3$ / 2.497) and Mg concentration (using the formula Mg, mg/L = (Total hardness as $CaCO_3$– Ca hardness as $CaCO_3$) /4.118).

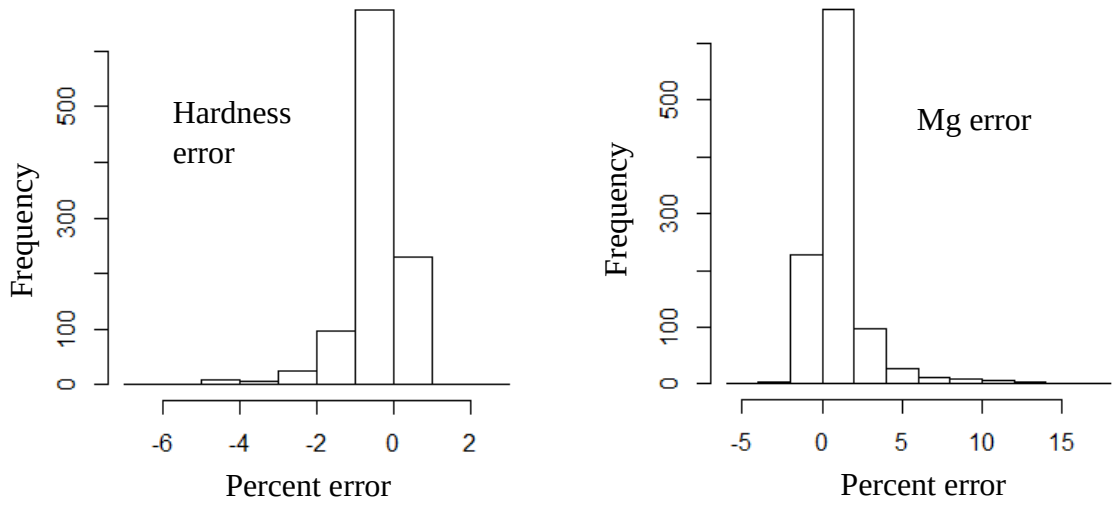


Figure I-1: error in hardness and Mg approximations; calculated from all NWIS COA groundwater sample data with known hardness, calcium, and magnesium concentrations. The left plot is the percent error between the known hardnesses and the values calculated from calcium and magnesium concentrations for each sample. The right plot is the percent error between the known magnesium values and the values calculated from calcium and hardness for each sample.

Conductivity was also converted to specific conductance. The formula below was used for the conversion, which comes from the USGS National Field Manual for the collection of Water-Quality Data (USGS, 2019). $K_{25°C}$ is specific conductance, K is conductivity, t is temperature in °C, and α is an adjustment factor between 0.019 and 0.02 which is accurate for most natural waters (pH > 4 and 5°C < t < 35°C). All COA groundwater samples used in the study fall within this range, so 0.02 was used for α.

$$K_{25°C} = \frac{K}{1 + \alpha(t - 25°\text{C})}$$

No other parameters were calculated from the general City of Norman data. To merge this dataset with other datasets, it was converted from a list of samples with multiple parameters (specific conductance, Ca, Mg, etc.) into a list of individual observations using the select and gather functions in the R package “tidyverse”, with code sample shown below:

```
library("tidyverse")

new_dataset = old_dataset %>%

  select(Date, Time, Source, Fluoride, Chloride, Calcium, Magnesium, Alk_as_CaCO3, TDS, pH,
  Sp_Conductance) %>%

   gather(key = "Parameter", value = "Result", -Date, -Time, -Source)
```

Essentially, the code selects the parameters illustrated in the select function, then splits them into individual rows with the gather function such that each row has one of the selected parameters

and the observed value for the selected parameter. Each parameter keeps the date, time, and source (well name) that was in the same row of the old dataset as the selected parameter. This result is then saved as new_dataset. An example with a smaller dataset is shown below:

Date	Time	Source	Fluoride	Chloride	→
1	2	3	4	5	→
6	7	8	9	10	→

→

Date	Time	Source	Parameter	Result
1	2	3	Fluoride	4
1	2	3	Fluoride	9
6	7	8	Chloride	5
6	7	8	Chloride	10
↓	↓	↓	↓	↓

Like the general City of Norman data, our field data came in rows where each row was a complete sample. Therefore, we used the same select and gather functions to select all parameters of interest and organize them into a list of observations instead of a list of samples. No parameter calculations were needed. Norman EPA compliance data and NWIS data (discussed below) are both formatted as rows of individual observations rather than rows of samples, so this transposition was necessary in order to combine the general City of Norman data and field data with the other data sources.

Norman EPA compliance data

The EPA compliance data was originally formatted in a list of individual parameters and results, so it did not need the transformation used for the general dataset. However, several other processing methods were required. First, confidence intervals were removed from the data and values below detection limits were replaced with the detection limit as the result value by removing the < symbol. An additional column was then added to the data with the value "TRUE" if the concentration fell below detection limits. This additional column was later used in the Kendall's Tau correlation coefficient calculations, which is discussed above. Initially, 29.7% of the EPA compliance data contained non-detects. After subsetting to only include parameters in this study (major ions, fluoride, nitrate, arsenic, chromium, uranium, and selenium), the number of non-detects decreased to 14.0%.

Subsetting the compliance data to parameters of interest for the study also took several steps. Since each parameter was input as a table entry rather than a column label, there was a lot of variation in the entries themselves that needed to be standardized. Arsenic measurements, for example, were labeled as "Arsenic, total" and ARSENIC in the dataset. Several of the entries also had extra spaces, which R reads as unique parameters. In order to standardize the parameter names, the "grepl" function in R was used to search for any parameters that included the uppercase and lowercase (except for the first letter) versions of the parameters. The key reason for using this instead of other subsetting functions was the include feature, which selects parameters that are not exact matches to the search parameters if the name of the search parameter is somewhere in the character string. For example. the grepl function treats "Arsenic,

total" and "Arsenic" the same if searching for the string "Arsenic". A snippet of the code used to do this is shown below:

data_subset = data[grepl('Arsenic|ARSENIC|. . .|last_parameter' , data$parameter) ,]

Essentially, the code above searches for any character strings in data frame "data" that contain the sequence "Arsenic" or "ARSENIC" in the parameter column and returns those rows in the data_subset (the "|" symbol is the logical operator OR in R). This way, the code can capture any combination of parameter names as long as they are spelled correctly and the first letter is capitalized. Adding the additional code line "unique(data$parameter)" can determine if the two conditions above are met. This data met those conditions; therefore, no additional character terms were needed.

Once the data subset was produced from all of the parameters of interest to this study, the variable parameter names were replaced with consistent ones by using the grepl function to generate indices of equivalent parameter names ("Arsenic" and "ARSENIC" for example) and then replacing the characters with a unified name. In addition to the parameter names, the sources (well names) and units were unified as well. R function str_remove_all() in the tidyverse package removed "well" and "WELL" characters in front of the numbers that the City of Norman used to label their wells. Lastly, the grepl function was again used (along with the which function) to find rows where arsenic, selenium, chromium and uranium concentrations were reported in mg/l (instead of the more widespread µg/l). Once the row numbers were determined, it was only necessary to multiply the results by 1000 and change the unit name.

NWIS Data

The R package 'dataRetrieval' (specifically the function 'readNWISqw()') was used to read data directly from the U.S. Geological Survey National Water Information System (NWIS) database into R. Only parameters that were already measured with other datasets were called; if parameters carried overlapping information, the option with the largest number of measurements was used. In addition to parameters, the function also needed a list of wells to call. The USGS published a list of City of Norman wells with their corresponding USGS well names in a 2009 report; this well list was used to determine well site numbers for the function call (Smith et al., 2009). Like the EPA compliance data above, the dataset produced from this R function was formatted as a list of individual observations and similar processing methods were required. NWIS data contains a remark column with a "<" value in every row that contains a value below detection limit; this was easily converted into values of "TRUE" to be used in the later calculations. Unlike the previous data, each parameter was listed as its respective code, so there was no need to make the names uniform. Code names were replaced with the parameters they represented, USGS well names were replaced with City of Norman well names, units were added based on parameter code descriptions, and the data was added to the other datasets with the R function "rbind()".

All of the other data sources consist of well-head measurements, while NWIS data contains both well-head measurements and depth-specific measurements. Depth-specific data were used for hazard mapping but needed to be filtered out of the combined Norman dataset because shallow depth-specific measurements do not show the variations in trace metals seen in deeper samples. Likewise, deeper samples could overemphasize certain trends. Keeping the depth-specific measurements in this dataset could therefore obscure some of the trends. There are four steps to filter out these types of data from an NWIS data call:

1. Add parameter 00003 to the readNWISqw() function to pull depth information from NWIS.
2. Make a subset of data containing only parameter 00003 and the columns date, time, source (well name), and result with subset() and/or select() and filter().
3. Change the name of the result column to something else, like “Depth”.
4. Join this subset to the original dataset by date, time, and source.

Assuming that all samples have a unique combination of date, time, and source, the four steps above will generate an extra column that easily distinguishes the depth-specific data from the other data. There were a number of samples with repeat date, time, source, and parameter values which needed to be aggregated (75 total observations across all parameters). Some duplicated samples were clearly corrections to typos; for example, arsenic entries with values of 1500 and 1.5 μg/L. Others were less obvious, like alkalinity measurements of 335 and 340 mg/L $CaCO_3$ reported for the same date and time, for example. For duplicate entries with differences greater than 10 units, the minimum value was taken to ensure that decimal point corrections were used in place of the original values. Duplicate entries with differences less than 10 were averaged together. Although duplicate samples (not intended for analysis) are collected regularly with environmental data for quality assurance (QA) purposes, they are not a likely major contributor in the duplicate data here. Firstly, QA data would likely record slightly different sample times than the original samples unless both sets of samples were collected very quickly. Secondly, the QA data would have to be misreported as regular field data to get mixed in with the collected data. Even if some QA data did make it through these two steps, there was no way to differentiate QA data from regular field data with the published information, so averaging repeat measurements halved the error that would come from using a duplicate sample in place of the regular field sample.

Duplicate entries with arsenic concentrations of < 1 and 1.2 μg/L were be averaged together, for example, so there also had to be a rule for how to report samples with at least one value. If the sample(s) reported as not below detection had a greater concentration than the sample(s) reported as below detection, not below detection was listed. If the samples reported as below detection had an equal or greater value to the samples reported as not below detection, below detection was listed. Note that this only occurred for differences of less than 10 μg/L or mg/L.

Pumping data and aggregate pumping metrics

The City of Norman keeps a record of the volume of water taken from each well on each day. A for loop in R iterated through each sample record and computed the mean daily volume of water extracted over each week leading up to the sample day (including the sample day, so 8 total days). Some days were missing from the record, so instead of using the mean() function, the sum() function was used divided by 8 total days to get the mean, which assumes that the well was not pumped on the day with missing data (Fig. I-2). The same for loop computed the number of days on (pumping volume > 0) for each weekly period. Then, using the "left_join()" function in the dplyr R package, the pumping information was joined to the existing sample data discussed above by the date and well ID parameters. Thus, each day on which a sample was collected from any given well was linked to a specific volume of water extracted and the pumping conditions on the week prior to the sample being extracted.

Daily pumping rates in well N

Date	Daily pumping volume (x1000 gal/day)
10/10/15	0
10/11/15	0
10/12/15	50
10/13/15	100
10/14/15	75
10/15/15	60
10/16/15	20
10/17/15	0
10/18/15	80
10/19/15	90

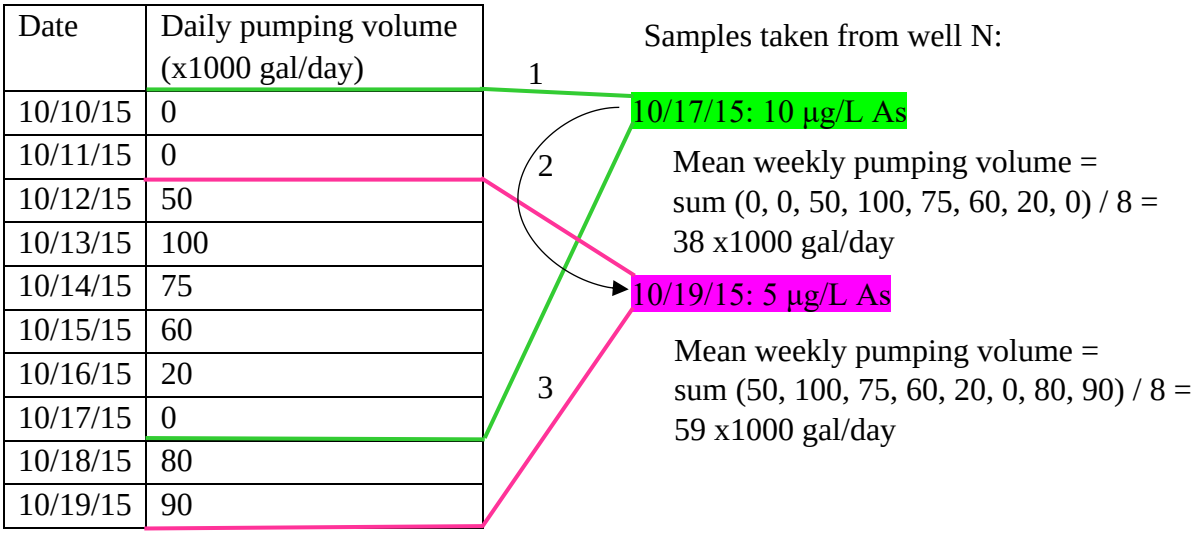
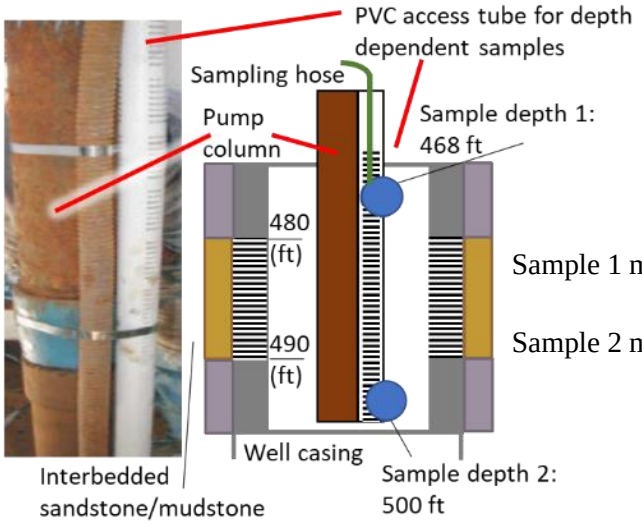


Figure I-2: Process for computing the weekly mean pumping rate for each sample taken from the well. In step 1, the code looks at the well's daily pumping volumes starting 8 days before the well was sampled and ending on the day that the well was sampled, then computes the mean pumping rate. In step 2, the code iterates to the next sample. Step 3 is a repeat of step 1 except the sample date is different, so the 8-day daily pumping window produces a different mean weekly pumping volume.

Appendix II: PHREEQC modeling supporting information

Sample calculation for mixing two samples close to a well screen



Sample 1 mixing fraction: $\frac{500-490}{(480-468)+(500-490)} = 0.45$

Sample 2 mixing fraction: $\frac{480-468}{(480-468)+(500-490)} = 0.55$

Figure II-1: USGS depth-dependent sampling procedure from Smith et al. (2009). A sample calculation for determining the mixing fractions for two samples near a single screened interval from 480-490 ft using inverse distance weighting is also included.

Well construction and lithology for wells used in modeling

Wells 2 and 5

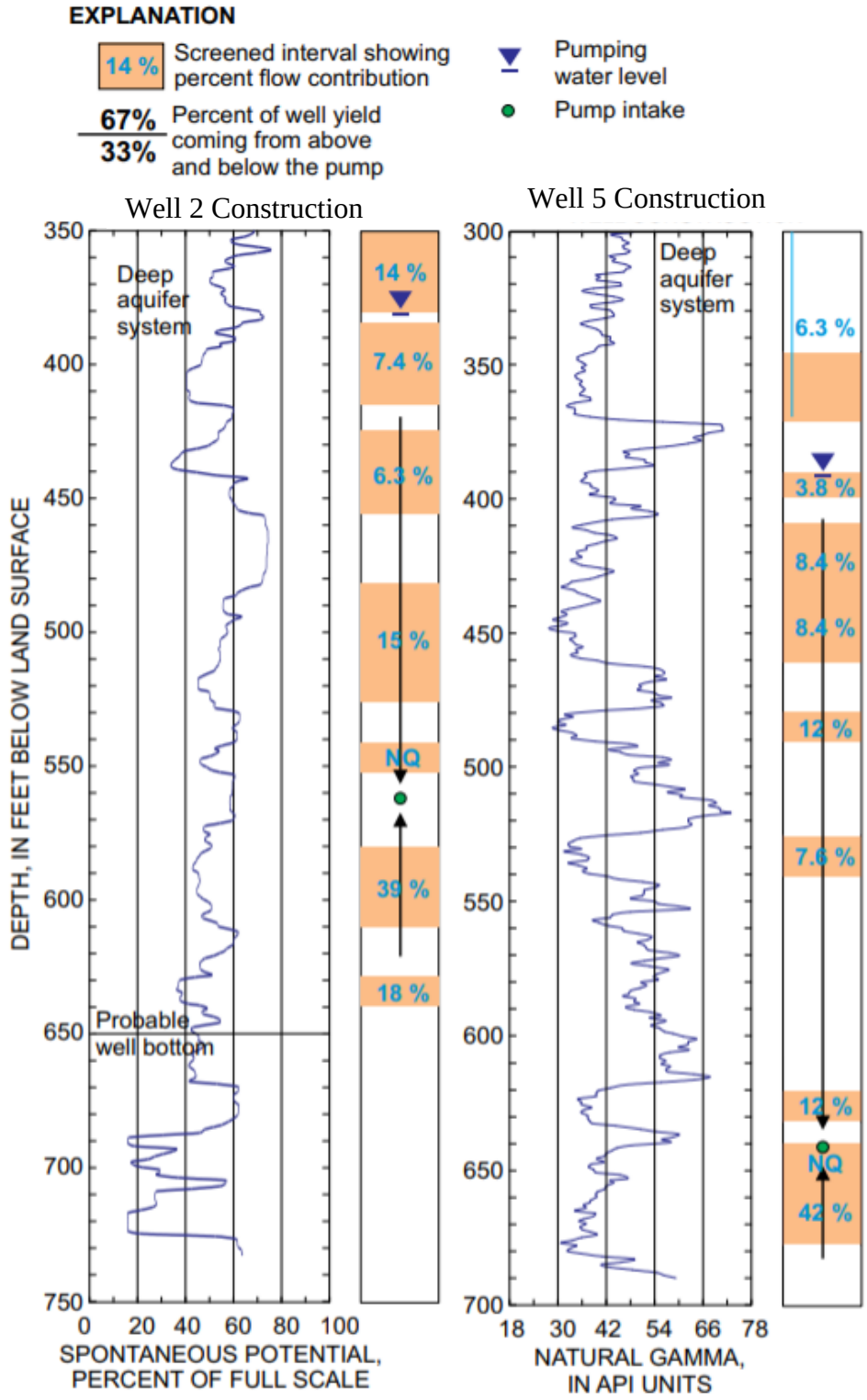


Figure II-2: Well 2 and well 5 screen, percent flow contribution, and well-log information from Smith et al. (2009). Spontaneous potential roughly corresponds to lithology, where deflection to the left indicates generally sandier units. Small natural gamma also indicates sandier units.

Wells 23 and 31

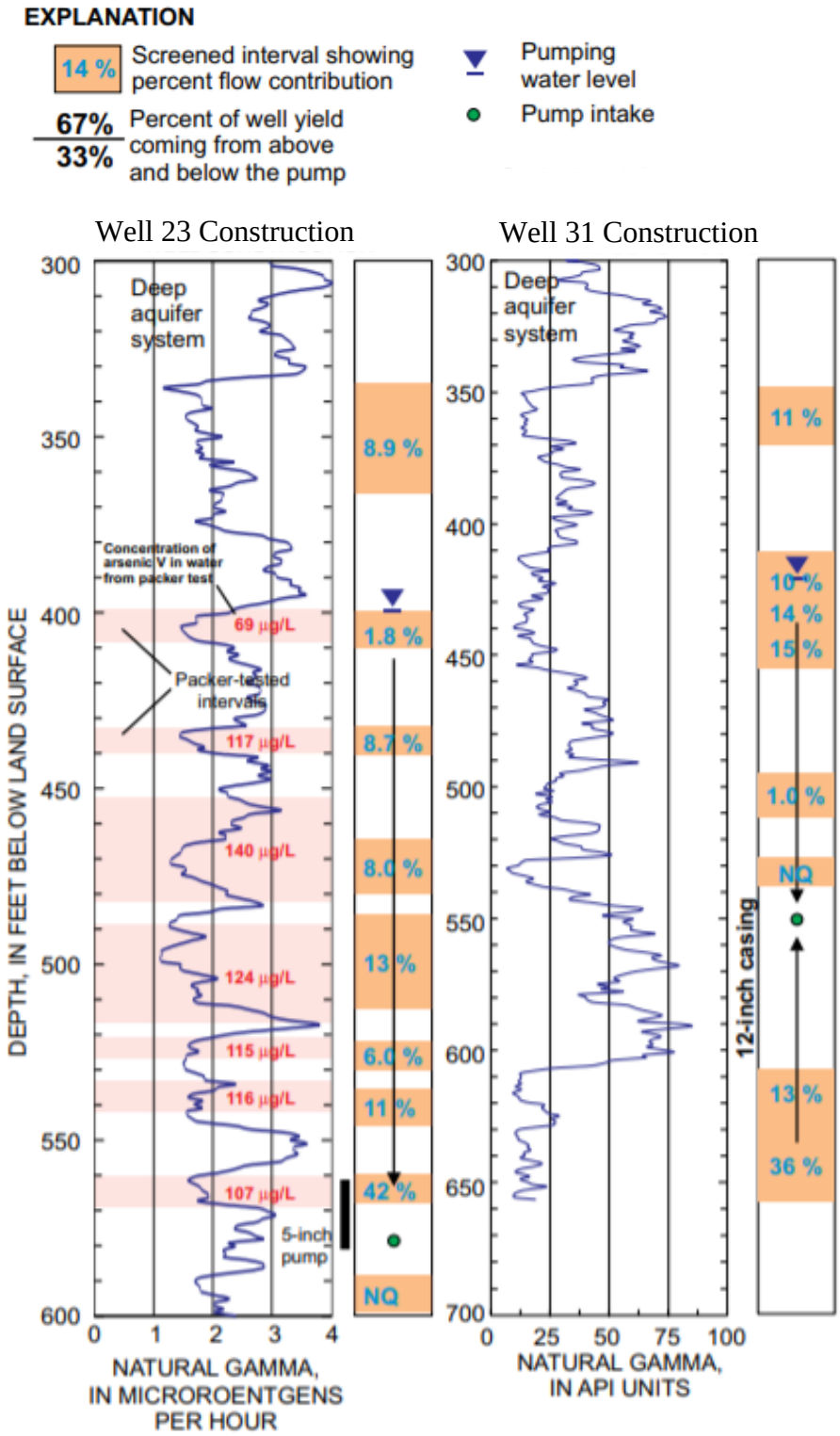


Figure II-3: Well 23 and well 31 screen, percent flow contribution, and well-log information from Smith et al. (2009). Small natural gamma indicates sandier units. Packer-testing was conducted on well 23 to isolate the intervals shown in red; resulting arsenic V concentrations in samples from those isolated intervals are also shown in red (Smith et al., 2009).

Well 36

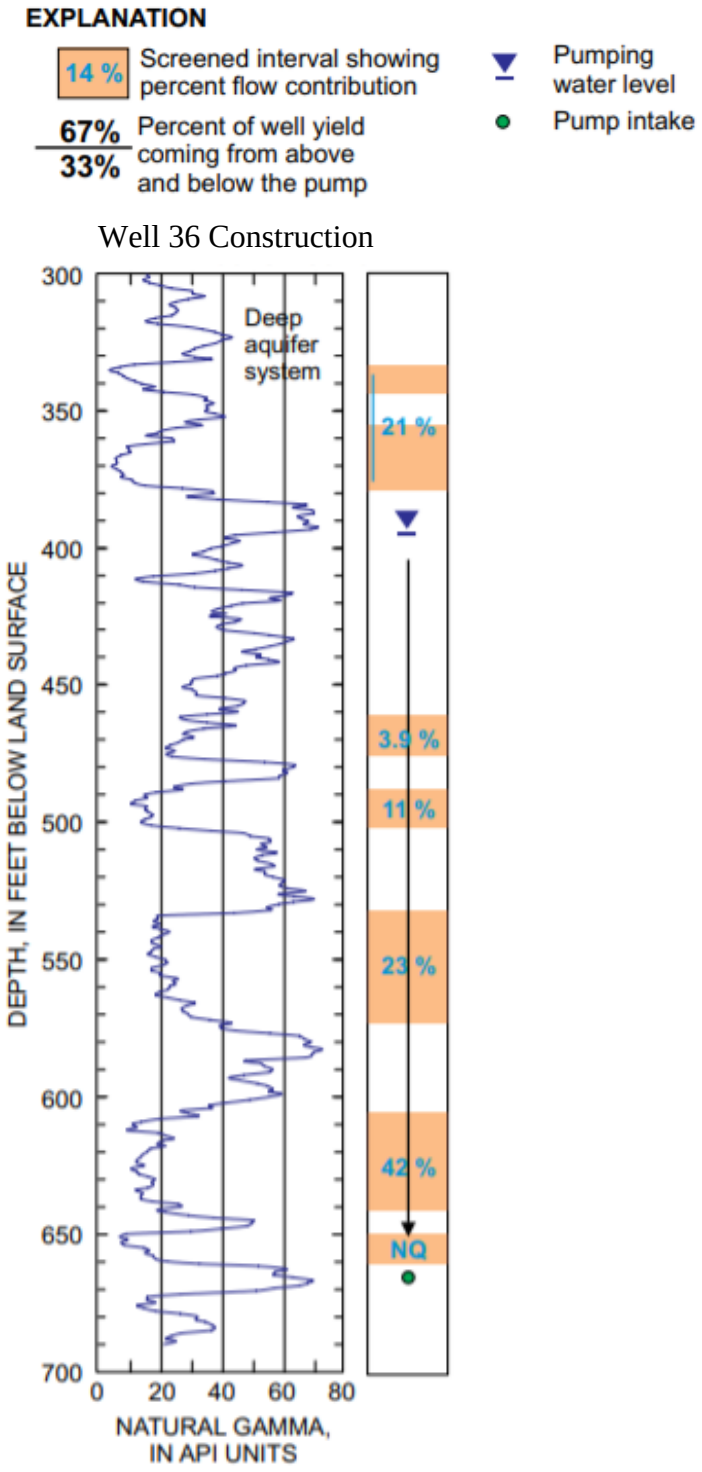


Figure II-4: Well 36 screen, percent flow contribution, and well-log information from Smith et al. (2009). Small natural gamma indicates sandier units.

Appendix III: Aggregated hazard maps for the COA

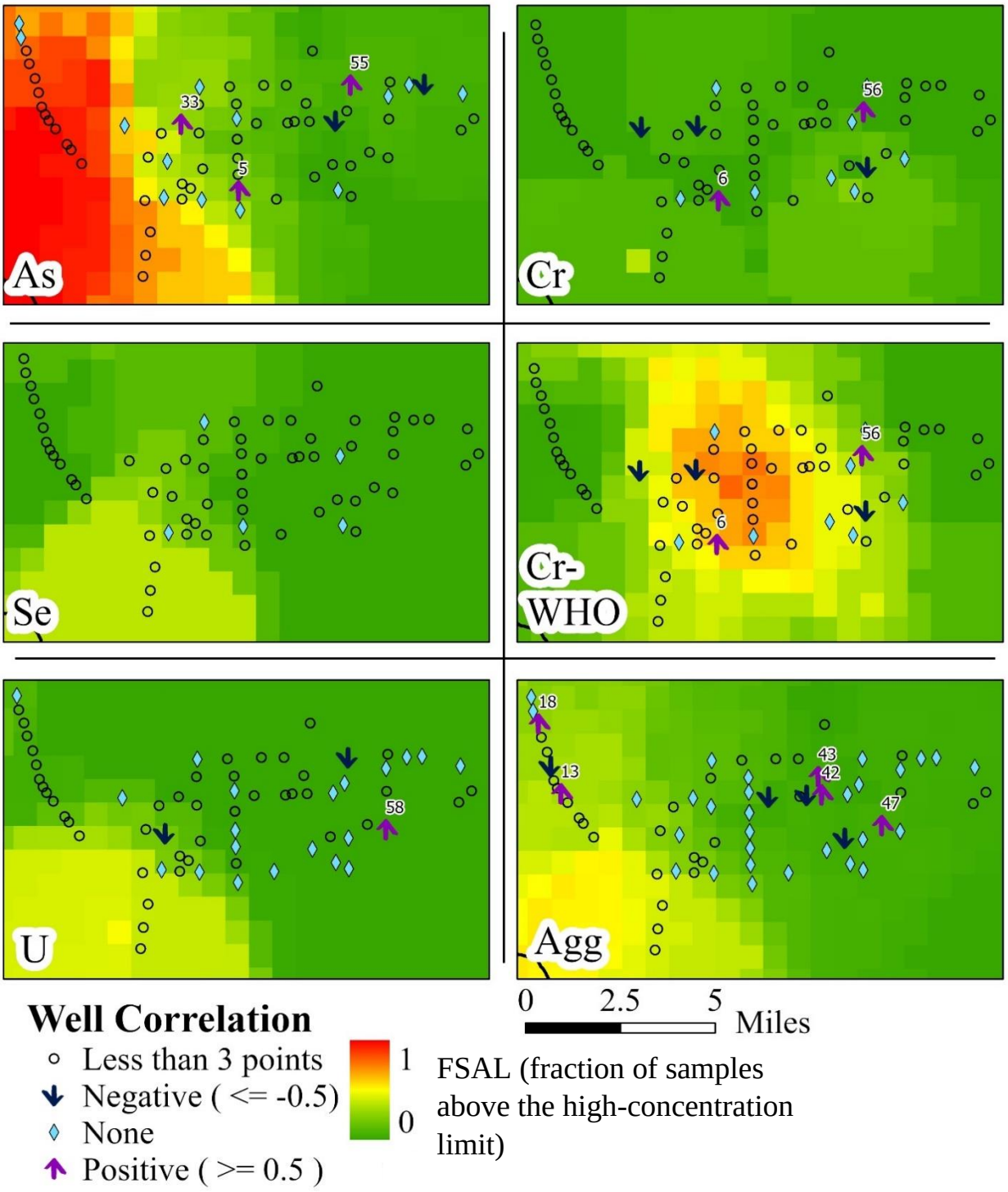


Figure III-1: Norman-specific trace metal hazard maps. The maps are identical to Figs. 24-26 but are shown together for better comparison.

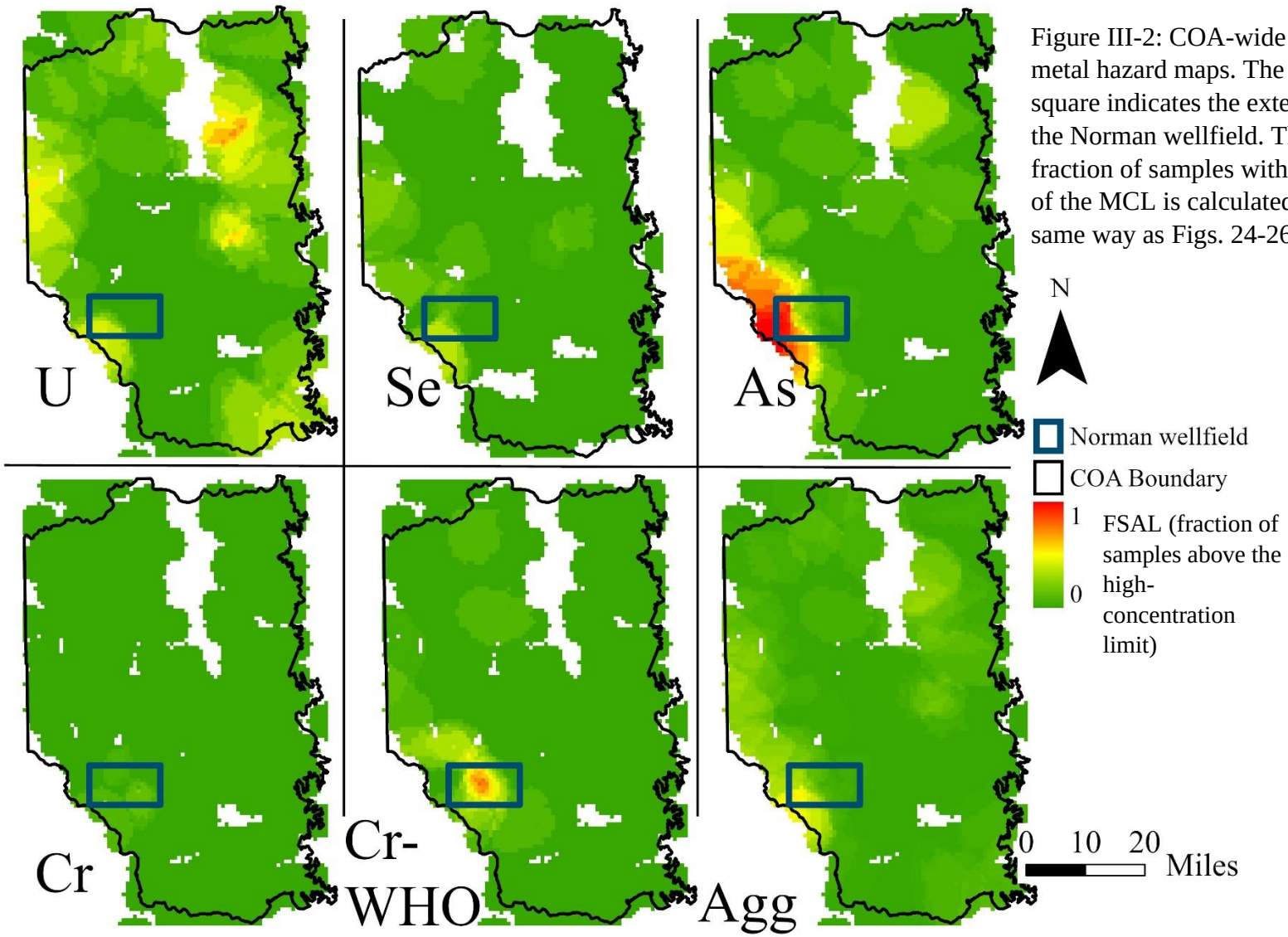


Figure III-2: COA-wide trace metal hazard maps. The blue square indicates the extent of the Norman wellfield. The fraction of samples within 10% of the MCL is calculated the same way as Figs. 24-26.

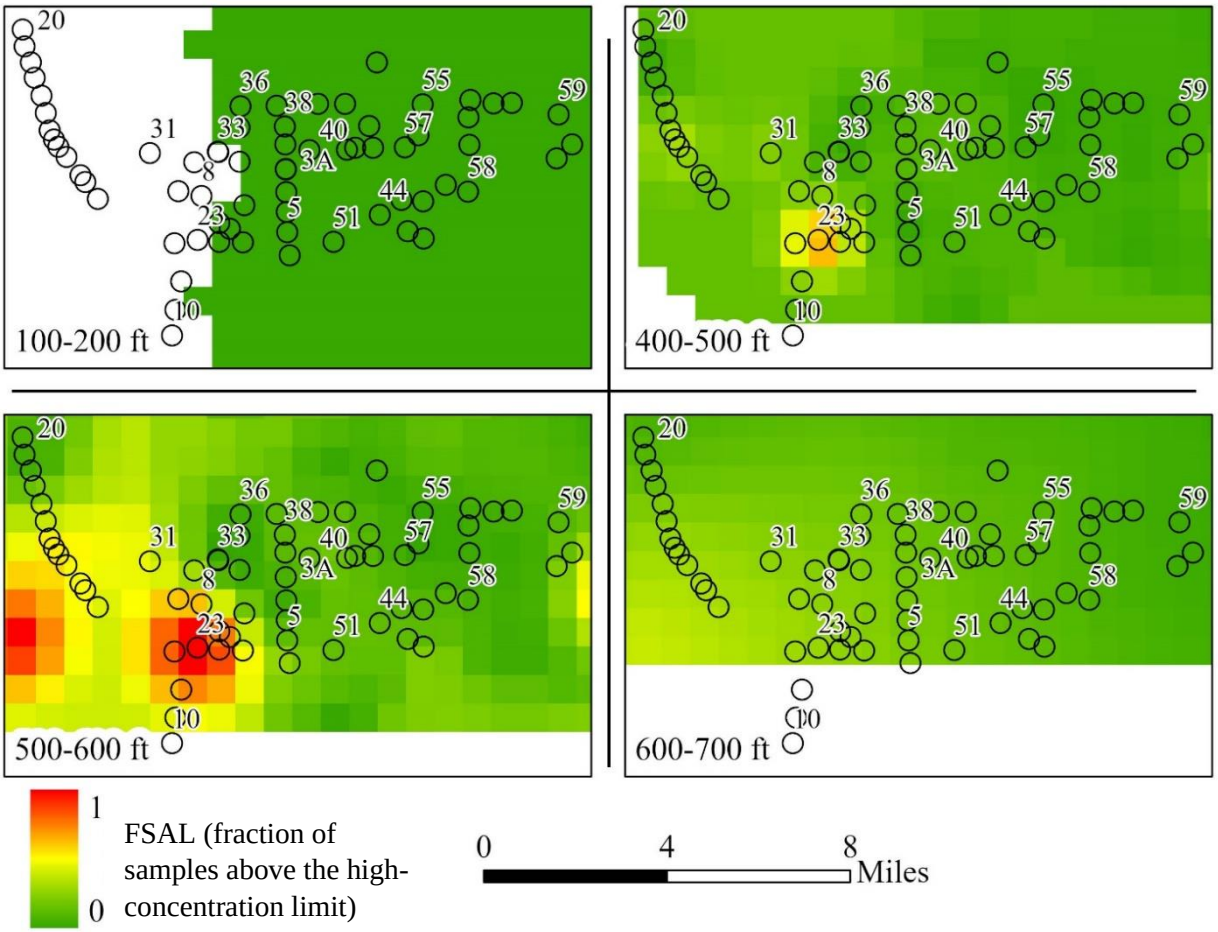


Figure III-3: Depth maps of aggregate trace metal hazard below the Norman wellfield. The EPA chromium limit was used in the hazard calculation. Depth layers 700-800 ft and 800-900 ft had one well apiece, both with no samples above the high-concentration limits. The other missing depth layers had 3-12 wells, all without samples near any of the MCLs. Kriging could therefore not be done on those layers. Labeled vs. unlabeled wells are for reference only and were chosen arbitrarily.