Factors Affecting Groundwater Nitrate Concentration in Madinah Area, Saudi Arabia

Majed M. Almutairi
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Hydrology Commons, and the Physical and Environmental Geography Commons

Recommended Citation

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
High levels of nitrate in groundwater are a serious problem in Madinah Area, Saudi Arabia. Identifying factors affecting groundwater nitrate contamination is important for managing groundwater quality. This study examined factors that have significant impacts on the high level of groundwater nitrate in Madinah Area. Factors examined included well-depth and land cover. Relationships between variables were explored using three statistical approaches: the Kruskal-Wallis test and two types of regression (ordinary least squares [OLS] and geographically weighted regression [GWR]).

Nitrate concentration data show that 73% of obtained groundwater samples exceeded the United States Environmental Protection Agency (EPA) maximum contaminant level (MCL) of 10 mg/L in Madinah Area. Results indicate that well-depth and agricultural lands have negative significant impact on groundwater nitrate concentrations. Based on the adjusted $R^2$, OLS and GWR models explain 38% and 37% respectively, of the nitrate concentration on groundwater. GWR model shows better performance compared to the OLS model by having higher $R^2$ and lower Akaike Information Criterion (AIC) value.

This study suggests that there could be sources of contamination on the land surface of Madinah Area, other than built-up and agricultural lands. Since nitrate concentration in groundwater did not increase as the percentage of built-up and agricultural lands increased, as has been found in other studies. This suggestion is supported by the inverse relationship between the concentration of nitrate in groundwater and well-depth.
Due to the lack of available data, it was necessary to travel to Madinah Area, Saudi Arabia to collect data (primary data) in the field using special equipment (water sample bottles designed especially for collecting groundwater samples and a tool for measuring the depth of the wells) to collect data in the field. Samples were transferred to a laboratory to be analyzed. The Madinah Area has not been previously studied to determine the factors contributing to nitrate contamination in groundwater. This study is the first to focus on Madinah Area to examine the factors contributing to high levels of nitrate concentrations in groundwater.
FACTORS AFFECTING GROUNDWATER NITRATE CONCENTRATION
IN MADINAH AREA, SAUDI ARABIA

by

Majed Masoud Almutairi

A thesis submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Master of Science
Geography
Western Michigan University
April 2023

Thesis Committee:

Adam J. Mathews, Ph.D., Chair
Lisa M. DeChano-Cook, Ph.D.
Chansheng He, Ph.D.
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. iii

LIST OF FIGURES ................................................................................................................ iv

LIST OF EQUATIONS ......................................................................................................... v

INTRODUCTION .................................................................................................................... 1

Research Objectives .......................................................................................................... 2

LITERATURE REVIEW ......................................................................................................... 4

Agricultural Practices ........................................................................................................ 4

Built-up Land ...................................................................................................................... 5

Well-depth .......................................................................................................................... 6

Statistical and Spatial Analyses ........................................................................................ 7

DATA AND METHODS ........................................................................................................ 9

Study Area ........................................................................................................................ 9

Data .................................................................................................................................. 10

Data Processing and Statistical Analyses ......................................................................... 14

RESULTS .......................................................................................................................... 20

Kruskal-Wallis Test .......................................................................................................... 22

Ordinary Least Squares Regression ................................................................................ 23

Geographically Weighted Regression .............................................................................. 24

DISCUSSION ....................................................................................................................... 28

CONCLUSION ................................................................................................................... 30

REFERENCES .................................................................................................................... 31

APPENDIX .......................................................................................................................... 38
LIST OF TABLES

1. Groundwater nitrate contamination. ................................................................. 1
2. Confusion matrix table......................................................................................... 16
3. Descriptive statistics based on summarizing within .......................................... 22
4. Kruskal-Wallis test results ................................................................................ 22
5. Summary of OLS results.................................................................................... 24
6. GWR results....................................................................................................... 25
LIST OF FIGURES


2. Map of the location of the study area (Madinah Area) ........................................................................ 9

3. Map of the study area showing sampling points and built-up land ............................................11

4. The locations of well 17. Source: Google Earth .................................................................................. 12

5. The locations of well 18. Source: Google Earth .................................................................................. 12

6. Excluded well. Source: Google Earth ................................................................................................. 13

7. Measuring well-depth .......................................................................................................................... 14

8. Sentinel-2 satellite imagery of the study area, after compositing, masking, and mosaicking .. 15

9. Land cover map of the study area ..................................................................................................... 16

10. Majority land cover and percentages of built-up and agricultural lands using 200 meters radius..................................................................................................................... 17

11. The spatial distribution of nitrate concentration in Madinah Area ............................................... 21

12. The spatial distribution of well-depth in Madinah Area .................................................................. 21

13. Residual map of OLS model ............................................................................................................... 24

14. Spatial distribution of GWR coefficients and R2. (Note: green indicates strength and red indicates weakness) ........................................................................................................... 27
LIST OF EQUATIONS

Eqn. 1 ............................................................................................................................ 18
Eqn. 2 ............................................................................................................................ 18
Eqn. 3 ............................................................................................................................ 19
Eqn. 4 ............................................................................................................................ 19
INTRODUCTION

Groundwater is the single most significant source of drinking water in many regions of the world, particularly in areas where surface water sources are scarce or polluted (Howard et al., 2006). Nitrate contamination of groundwater impacts human health and the environment (Zhou, 2015). In drinking water, the maximum contaminant level (MCL) of nitrate permitted by the United States Environmental Protection Agency (EPA) is 10 mg/L. While nitrate does occur naturally in groundwater, concentrations greater than 3 mg/L generally indicate contamination from non-natural sources (EPA, n.d.). However, nitrate concentrations in drinking water more than 2 mg/L have been linked to negative health impacts. Table 1 shows groundwater guidelines for water with known nitrate concentrations (Gallagher & Gergel, 2017; Lotfata & Ambinakudige, 2019; Nolan et al., 2014).

Table 1. *Groundwater nitrate contamination*

<table>
<thead>
<tr>
<th>Nitrate levels</th>
<th>Water type</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 2 mg/L</td>
<td>Natural water</td>
</tr>
<tr>
<td>&gt; 2 mg/L</td>
<td>Degradation of water quality</td>
</tr>
<tr>
<td>&gt; 5 mg/L</td>
<td>A severe degradation of groundwater</td>
</tr>
<tr>
<td>&gt;10 mg/L</td>
<td>EPA maximum contaminant level (MCL)</td>
</tr>
</tbody>
</table>

High levels of groundwater nitrate have been observed in many countries around the world, such as Belgium, Denmark, France, Germany, Luxembourg, Netherlands, the United Kingdom, China, India, and the United States (Burow et al., 2009; Fried, 1991; Rao et al., 2017; Zhang et al., 1996). Alabdula'aly et al. (2010) indicated that nitrate levels exceeded the maximum contaminant limits for drinking water in 213 of 1060 wells in different regions of Saudi Arabia. Importantly, there are no rivers, lakes, or other perennial surface water sources in Saudi Arabia. Saudi Arabia primarily relies on groundwater for municipal and agricultural consumption, at a rate of 82% in 2018 (General Authority for Statistics [GAS], 2018).
Research Objectives

Few hydrochemical and quality assessments of groundwater studies have been undertaken in the Madinah Area of Saudi Arabia. El Maghraby (2015) found that the nitrate concentrations exceeded the recommended values in parts of the Madinah Area. Alghamdi et al. (2020) report that 90.7% of water samples showed high amounts of nitrate. Elevated nitrate levels in groundwater are a serious problem in Madinah Area due its potential impact on human health. Identifying factors affecting the groundwater nitrate concentration is imperative to controlling groundwater quality. Though similar work has been done in other places, the Madinah Area has never been specifically examined to see what factors are contributing to nitrate contamination.

The overall aim of this study is to analyze the extent of nitrate concentration in groundwater in the Madinah Area, Saudi Arabia and identify factors that have significant impacts on nitrate concentrations of groundwater using multivariate statistical analysis with linear Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) models. It is important to identify areas within the study area that have high nitrate levels. Once identified, proper management techniques can be applied to stem further contamination of the groundwater. Specifically, this study addressed four research objectives:

1) Measure and model the spatial distribution of nitrate in groundwater throughout Madinah Area,

2) Identify primary factors leading to the high-level nitrate contamination of groundwater in the Madinah Area,

3) Examine how land cover types are associated with nitrate contamination of groundwater,
4) Explore the relationship between nitrate concentration and well-depth in the Madinah Area.
LITERATURE REVIEW

The scope and methods used to study nitrate contamination in groundwater vary. Land use and land cover are major factors that have received attention from researchers. Also, well-depth in many studies has been investigated in its contributions to nitrate contamination of groundwater. Statistics have been widely used to evaluate relationships between variables and nitrate contamination levels.

Agricultural Practices

Agricultural practices are often linked to high nitrate concentrations in groundwater (El Maghraby, 2015; Lotfata & Ambinakudige, 2019; Maas, 2006; McLay et al., 2001; Wick et al., 2012; Zhang et al., 2018; Zhang et al., 2020). Nitrogen, phosphate, and potash are essential for agriculture used for food for humans, animals and fuel. Most of these nutrients are absorbed by the crop, but when applied in excess, they can be lost to the environment through leaching into groundwater (United States Department of Agriculture [USDA], 2019; Lotfata & Ambinakudige, 2019). Lotfata (2019) found that nitrate concentrations of groundwater in Texas was significantly correlated with percent of the cotton grown in the area. Wick et al. (2012) found that the percentage of cropland in Austrian municipalities correlates positively with nitrate concentration in groundwater. Maas (2006) found that agricultural land had a significant relationship to nitrate levels. McLay et al. (2001) suggested that the region's non-point source groundwater nitrate contamination is a result of intensive agricultural practices. Zhang et al. (2018) indicated that groundwater in the Jinghui canal irrigation area in Shaanxi Province of China where agricultural activities are intense is seriously polluted by nitrate. Zhang et al. (2020) found that nitrate concentrations in groundwater were often quite high in agricultural lands. El Maghraby (2015) suggested that it might be one of the reasons for high concentration of nitrate in groundwater in
the Madinah Area, Saudi Arabia, due to the application of fertilizers for agricultural uses.

According to the UN Comtrade Database, Saudi Arabia's imports of fertilizers totaled US $120.7 million in 2021 (Figure 1).


Built-up Land

In industrial and urban land, sources of nitrate are sewage generated from domestic and industrial activities and septic tanks, where urea and ammonium prevail over other nitrogen compounds (Esmaeili et al., 2014). Studies have also linked high concentrations of nitrate in groundwater to urban areas (Groen et al., 1988; Eun-Hee et al., 2020; Bilgehan & Ali, 2006; Wakida & Lerner, 2005; Wang et al., 2017). Groen et al. (1988) showed that elevated nitrate concentrations were more prominent in villages with a high building density, and the disposal of organic waste in and around villages. Wang et al. (2017) found manure and septic waste were dominant sources for most groundwater with high nitrate concentrations in both farmland and residential areas. Eun-Hee et al. (2020) report that the urban variable significantly contributed to
nitrate enrichment in groundwater in the certain parts of the island. Wakida & Lerner (2005) focused on non-agricultural urban areas and found that the main source of nitrate in urban areas is sewage due to the poor sewage system in cities. Bilgehan and Ali (2006) found that nitrate concentrations generally tend to increase in the city center of Konya, Turkey.

Other studies have linked high concentrations of nitrate in groundwater to industrial area (Capri et al., 2009; Gu et al., 2013; Esmaeili et al., 2014). Gu et al. (2013) found that groundwater nitrate concentrations were the highest beneath industrial land, followed by urban land. Esmaeili et al. (2014) believed the agricultural operations, coupled with home sewage and industrial wastewaters, in inhabited areas are to blame for the high nitrate level. Capri et al. (2009) noted that groundwater contamination by nitrate was caused by both linear sources such as highways, and leaky sewage systems and by point sources such as surface impoundments, septic tanks, pits, lakes, and landfills in numerous urban and industrial regions. According to the International Trade Administration (ITA), every year, Saudi Arabia generates over 53 million tons of waste, damaging its soil and groundwater (ITA, 2022).

Well-depth

Many studies found a negative relationship between well-depth and nitrate concentration in groundwater (Liu et al., 2005; Lotfata, 2019; Uhlman & Artiola., 2011; MacLeod et al., 1995). According to MacLeod et al. (1995), as the depth of groundwater increases, nitrate contamination normally decreases. Median nitrate concentration and percent of wells from which water exceeds the EPA drinking-water standard for nitrate (10 mg/L) are highest for shallow groundwater (up to 100 feet deep). Because the water table in shallow wells is closer to the land surface and potential sources of contamination, such as fertilizers and septic systems. Lotfata (2019) used well-depth as an independent variable and nitrate concentration as the dependent
variable. Results indicated that the nitrate concentrations tended to increase as the well-depth decreased in both the Southern High-Plains and the Edwards-Trinity aquifers, Texas, USA.

Uhlman and Artiola (2011) conducted a study on Arizona groundwater findings that well-depth is an important factor, with shallower wells being more likely to be contaminated by nitrate. Liu et al. (2005) predicted the nitrate concentration in groundwater in Alabama, USA using data on land use, well-depth, septic tank use, distance from the well, and animal and farming activity around wells. Results indicated that well-depth and cropping activity were factors of statistical significance in influencing nitrate concentration in these wells.

In contrast, Bilgehan and Ali (2006) applied a statistical correlation procedure to well-depths and nitrate concentrations. Correlation coefficients of 0.259 and 0.261 were found for data gathered in 1998 and 2001, respectively. It is determined that, within the study area, there is no correlation between the distribution of nitrate concentrations and well-depths.

Statistical and Spatial Analyses

Ordinary Least Squares Regression (OLS) and Geographically Weighted Regression (GWR) models have been used in many studies to examine well-depth and land use activities’ impacts on groundwater nitrate concentration (Eun-Hee et al., 2020; Lotfata, 2019; Shrestha et al., 2017). Eun-Hee et al. (2020) used OLS and GWR to investigate the relationships between nitrate concentrations and various parameters (e.g., topography, hydrology, land use) on Jeju Island, South Korea. Results showed that the GWR models outperformed OLS regression models, with a higher $R^2$ and a lower corrected Akaike Information Criterion (AIC) value. The GWR model was able to reveal previously unknown information that was not revealed by the OLS regression models. For example, the GWR model discovered that orchards and urban variables significantly contributed to nitrate enrichment in certain parts of the island, whereas the
OLS regression model ignored these variables as statistically insignificant factors. These results attest to the importance of analyzing these data using different spatial approaches (i.e., GWR). Lotfata (2019) used OLS and GWR models to identify factors that have a significant impact on elevated groundwater nitrate concentrations in the Southern High Plains and the Edwards-Trinity aquifers. The OLS and GWR models' values were compared, and results showed lower AIC scores and higher adjusted $R^2$ values compared to the OLS models, demonstrating its robustness over the OLS model. Shrestha et al. (2017) used GWR and OLS models to determine which method is more successful at uncovering environmental variables that influence groundwater nitrate concentration. These results showed that the GWR model outperformed the OLS model, with lower AIC value, and captured the geographical variability of fertilizer, precipitation, and elevation for the percent of wells with concentrations above 5 mg/L in California's Central Valley.

Liu et al. (2009) used a Kruskal-Wallis test to determine whether there were any significant differences between land-use groups (i.e. urban, agriculture, grassland, forest, water, wetland, barren and shrubland) and water quality variables (i.e. total phosphorus, lead, and fecal coliform) finding that water quality variables showed significant differences (with $p$ values $\leq 0.0005$) among the land-use groups.
DATA AND METHODS

Study Area

The Madinah Area is in the western region of the Kingdom of Saudi Arabia (Figure 2). The administrative region of Madinah consists of nine provinces: Hunakiyah, Al-Mahd, Al-'Ula, Badr, Yanbu, Khaybar, Al-Ais, Al-Fara, and Al-Madinah Al-Munawwarah Area. Al-Madinah Al-Munawwarah Area (hereafter, Madinah Area) is the study area covering area of approximately 19,050 square kilometers.

Figure 2. Map of the location of the study area (Madinah Area)

The Madinah Area is famous for its agriculture, especially palm trees. The Madinah Area ranked third in the number of palm trees in the Kingdom of Saudi Arabia, where the number of palm trees reached 4,751,040 palm trees, representing 15% of the total number of palm trees in
the Kingdom in 2019 (General Authority for Statistics [GAS], 2019). Due to the lack of data, the amount of fertilizer import for palm trees in Madinah Area is unknown. However, (Figure 1) shows the imports of fertilizer to Saudi Arabia in general, which are increasing. The average maximum temperature in the Madinah Area reaches more than 40°C in some seasons. Also, the Madinah Area is in the dry subtropical zone, so the Madinah Area receives little rainfall. The average annual total amount of rain in the Madinah Area is 60 mm (Hafez & Meriz, 2013).

Madinah's elevation ranges between 600 and 610 meters. Lava plateaus cover some parts of the area (El Maghraby, 2015). According to GAS, the Madinah Area population reached 2,239,923 in mid-2019. In 2010 the population of the city was 1,694,749. This marked increase in population has resulted in an increase in a built-up land (GAS, 2019).

Data

Groundwater samples were collected from 60 water wells in the study area (Figure 3). In choosing wells, several criteria were considered including ease of access to the well such as the presence of paved roads. Therefore, I excluded parts of the north-east and south-east of the study area, as there are no roads that lead to them. Wells with water pumps were included, while those operated using traditional methods (i.e., dug manually) were excluded due safety reasons, in which most of hand dug wells are open and it is dangerous to approach them to get both the measurement of well-depth and water samples. High spatial resolution satellite imagery from Google Earth was used to locate wells. Wells were numbered according to the order of visit (see Figures 4 and 5).
Figure 3. Map of the study area showing sampling points and built-up land.
Groundwater samples were collected in four-ounce glass containers designed to collect and store water samples. One sample was taken at each well (60 samples total). At the time of sampling, sampling bottles were thoroughly rinsed two to three times with the groundwater to be sampled. Water samples were collected after pumping the water for at least 10 minutes. This was
done to avoid impurities suspended inside the water pumps (El Maghraby, 2015). Sample bottles were placed in a container after collected and labeled with the water well number. Ice cubes were placed next to the sample bottles, and the container was closed so that groundwater samples would not be affected by heat or light (Farrell-Poe, 2010). Groundwater samples were transferred to Musa Group Environmental Laboratory (www.musaenvlab.com) to test the percentage of nitrate concentration in each sample. Hydro-chemical analysis was conducted for a fee by Musa Group Environmental Laboratory (www.musaenvlab.com). Results of hydro chemical analyses are presented in Appendix A. Nitrate concentration is reported in mg/L.

One water sample (sample No. 3) was excluded because the well from which the sample was taken is used by a sewage plant. The smell and color of the sample indicated that the water sample taken from the well is not groundwater but sewage water (Figure 6).

Figure 6. Excluded well. Source: Google Earth

Well-depth data groundwater in Madinah is not available and most of water wells in Madinah Area were drilled before the licensing system was issued. So well-depth was collected in the field using water level meter with fixed probe and steel tape for depth 100m. This tool can measure water level and well-depth. I used this measuring tool to get the well-depth only not the
water level. This is accomplished by lowering the weighted tape until the weighted end is felt resting on the bottom of the well (Figure 7; EPA, 2020). I collected 54 well-depth data, but I was not able to get the measurements of six water wells, which are 1, 33, 46, 47, 48, and 60. This is because these water wells have a depth that more than 100 meters, and that exceeds the length of the measuring tool so, I assigned these water wells a value of 100. Results of well-depth are in meters (Appendix A).

Remotely sensed imagery from Sentinel-2 for the study area was freely downloaded from the Copernicus Open Access Hub (scihub.copernicus.eu). Images have a spatial resolution of 10 meters and were collected on June 6th, 2022.

Figure 7. Measuring well-depth

Data Processing and Statistical Analyses

Twelve individual satellite images were used and mosaicked to create a single image for classification using four bands (2,3,4, and 8). After 12 satellite images were downloaded, they were brought into ArcGIS Pro 3.0.0 and analyzed using the Geostatistical Analyst extension. A
shapefile of the study area was downloaded from *ArcGIS Online* (esri.maps.arcgis.com) and used to extract the study area using spatial analyst tools (extract by mask) (Figure 8).

**Figure 8. Sentinel-2 satellite imagery of the study area, after compositing, masking, and mosaicking**

![Sentinel-2 satellite imagery of the study area](image)

Image classification was needed to transform the remotely sensed imagery into categorical data representing land cover types. Land cover categories are as follows: (1) built-up land; (2) agricultural land; (3) barren land; (4) volcanic land. Pixel-based supervised image classification was conducted using the Maximum Likelihood Classifier (MLC) approach. A training sample manager tool was used and 40 polygons were defined for each category. Heads-up digitizing of desired land cover classes guided the classification algorithm (Figure 9). After classification, the accuracy of the output categorical map was assessed using a confusion matrix and kappa coefficient. The kappa index value was 0.73 which indicate 73% level of agreement (Table 2).
Figure 9. Land cover map of the study area

Table 2. Confusion matrix table

<table>
<thead>
<tr>
<th>Class Value</th>
<th>Built-up</th>
<th>Agriculture</th>
<th>Barren</th>
<th>Volcanic</th>
<th>Total</th>
<th>Accuracy</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built-up</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>40%</td>
<td>0</td>
</tr>
<tr>
<td>Agriculture</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>Barren</td>
<td>5</td>
<td>5</td>
<td>473</td>
<td>2</td>
<td>485</td>
<td>98%</td>
<td>0</td>
</tr>
<tr>
<td>Volcanic</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>20</td>
<td>474</td>
<td>12</td>
<td>515</td>
<td>0%</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy</td>
<td>44%</td>
<td>50%</td>
<td>100%</td>
<td>83%</td>
<td>0</td>
<td>97%</td>
<td>0</td>
</tr>
<tr>
<td>Kappa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>73%</td>
</tr>
</tbody>
</table>

To determine the dominant type of land cover around these water wells, 200-meter buffers (e.g., 200m radius) were created for each well point, after that, the summarize within tool was used. The 200 meters radius was used as the input polygon, and I used the land cover map after converting it from raster to vector in the input summary features. This tool allowed me to
know the type of dominant land cover on each well by determining the type of dominant land cover and its percentage (Figure 10) (Lotfata, 2019; MacLeod et al., 1995). Dominant land cover for each water well was used in the Kruskal-Wallis test. The percentage of each land cover types around these water wells were used in OLS and GWR. Volcanic lands were excluded in both methods because they do not dominate any of wells’ land and they do not have any percentages in the 200m radius. Since there were only three land cover types, barren lands were excluded from OLS and GWR to avoid independent variables becoming dependent on each other.

Figure 10. Majority land cover and percentages of built-up and agricultural lands using 200 meters radius

Relationships between variables were explored using three statistical approaches: the Kruskal-Wallis test and regression (two types: OLS and GWR). Kruskal–Wallis test for k independent groups is the nonparametric version of one-way Analysis of Variance (ANOVA) when the normality assumption of ANOVA is not met (Liu et al., 2009). The Kruskal–Wallis test compared groups based on land cover types to examine differences in nitrate concentration. Groups were determined based on the three different land cover types (e.g. agriculture, built-up, and barren). All statistical analyses were conducted in IBM SPSS Statistics 28.

OLS was used to investigate the relationship between the dependent variable, which was groundwater nitrate concentration, and the independent variables, which were the well-depth and
percentages of built-up agricultural lands in 200 meters radius that had been created around each water well. OLS is noted as follows (Equation 1):

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_3 \ldots \]

where \((Y)\) is a dependent variable that denotes nitrate concentration and \((X_1, X_2, \text{and} X_3)\) are independent variables that indicate well-depth and land cover type (built-up and agriculture), respectively (Lotfata, 2019).

The spatially varying relationship between well-depth, groundwater nitrate concentration, and land cover type were investigated using the GWR model. The GWR model extends the traditional multiple linear regression model by estimating local parameters. GWR is computed as follows (Equation 2):

\[ y_i = \beta_{i0} + \sum_{k=1}^{m} \beta_{ik} x_{ik} + \varepsilon_i \]

where \((y_i)\) is the dependent variable that denotes nitrate concentration at location \((i)\). \((x_{ik})\) is the \((kth)\) independent variables, which indicate well-depth, and land cover type (Bulit-up and Agriculture) at location \((i)\); \((m)\) is the number of independent variables; \((\beta_{i0})\) is the intercept parameter at location \((i)\); \((\beta_{ik})\) is the local regression coefficient for the \((kth)\) independent variable at location \((i)\); and \((\varepsilon_i)\) is the random error at location \((i)\) (Fotheringham et al., 2002).

When compared to the traditional multilinear regression model, GWR allows coefficients to vary across the study area, and a set of coefficients can be estimated at any location, allowing a coefficient surface to be visualized and questions about relationship heterogeneity to be asked.
The Gaussian kernel was used in GWR, and the weight for the Gaussian kernel function was assessed as follows (Equation 3):

\[ W_{ij} = \exp \left[ -\frac{1}{2} \left( \frac{d_{ij}}{b} \right)^2 \right] \]

where \( d_{ij} \) represents the Euclidean distance between point \( i \) and data point \( j \). A distance metric measure defines \( b \) as a fixed bandwidth size. The Gaussian kernel weight declines constantly and regularly from the kernel's center but never reaches zero. The Gaussian kernel is appropriate for fixed kernels because it can reduce the probability of a kernel having no data (Fotheringham et al., 2002; Lotfata, 2019). The Akaike Information Criterion (AIC) was used to evaluate model goodness of fit (Equation 4):

\[ \text{AIC} = \text{Deviance} + 2k \left[ \frac{n}{n - k - 1} \right] \]

\( n \) indicates the number of data points; \( k \) is the number of parameters in the model. A lower value for \( \text{AIC} \) implies better model fits (Lotfata, 2019).
RESULTS

The nitrate concentrations in groundwater exceeded 2 mg/L in 58 wells out of 59 wells, exceeded 5 mg/L in 54 wells out of 59 wells, and maximum contaminant level of 10 mg/L in 43 wells out of 59 wells. The maximum nitrate concentration was 96.5 mg/L, and the minimum nitrate concentration was 0.10 mg/L. The total average nitrate concentration was 22 mg/L. Based on Table 1, only one sample out of 59 samples was natural water which is below 2 mg/L and 98 percent of the sample was considered to be degraded water quality which is above 2 mg/L. 91% of the sample was considered to be severely degraded groundwater which is above 5 mg/L and 73% of the sample exceeded EPA maximum contaminant level. Based on my 100m measurement tool, the well-depth in the study area ranged between 8 to more than 100 meters.

Figure 11 shows the spatial distribution of nitrate concentration in groundwater in Madinah Area. The concentration of nitrate, which did not exceed the standards of EPA, is in dark and light green areas and distributed in separate wells throughout Madinah Area. The high nitrate concentration is represented in yellow, orange, and red and it is distributed in most parts in Madinah Area, but more concentrated in the southern part of Madinah Area.

Figure 12 shows the spatial distribution of well-depth across Madinah Area. The well-depth in the center and southern parts of Madinah Area ranged between 8 to 23 meters. These two parts which is in dark green is the shallowest well-depth in Madinah Area. Deepest wells are in dark red, they ranged between 70 to more than 100 meters. Most of these wells located in northeast parts of Madinah Area.
Figure 11. *The spatial distribution of nitrate concentration in Madinah Area*

![Nitrate Concentration Map](image1)

Figure 12. *The spatial distribution of well-depth in Madinah Area*

![Well-depth Map](image2)
Before completing the Kruskal-Wallis test, a descriptive test was carried out to determine the number of wells in each category based on results of summarizing within in *ArcGIS*. The result of this test indicated that the number of wells located within the built-up land category is 3 water wells; 13 water wells were classified as agricultural wells and 43 water wells were classified as wells located in the barren area (Table 3).

Again, Kruskal-Wallis does not assume normally distributed data, but one of the assumptions is the sample size. Each group must have a sample size of 5 or more. Kruskal-Wallis is normally applied to three or more independent groups, but it can be applied to two, with each group having a sample size of 5 or more (Lomuscio, 2021). Since the built-up group does not meet the assumption of sample size, I have excluded it. The Kruskal-Wallis test shows that there is no statistically significant difference between the nitrate concentration on wells that dominated by barren land and wells that dominated agricultural land (Table 4).

**Table 3. Descriptive statistics based on summarizing within**

<table>
<thead>
<tr>
<th>Land Type</th>
<th>N</th>
<th>Mean</th>
<th>Std. D.</th>
<th>Std. E.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built-up</td>
<td>3</td>
<td>21</td>
<td>20.9</td>
<td>12</td>
<td>.10</td>
<td>41.90</td>
</tr>
<tr>
<td>Agriculture</td>
<td>13</td>
<td>20</td>
<td>13.7</td>
<td>3.8</td>
<td>5.50</td>
<td>52.80</td>
</tr>
<tr>
<td>Barren</td>
<td>43</td>
<td>22</td>
<td>18.6</td>
<td>2.8</td>
<td>2.70</td>
<td>96.50</td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>22</td>
<td>17.5</td>
<td>2.2</td>
<td>.10</td>
<td>96.50</td>
</tr>
</tbody>
</table>

**Table 4. Kruskal-Wallis test results**

<table>
<thead>
<tr>
<th>Land Type</th>
<th>N</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO3 Agriculture</td>
<td>13</td>
<td>29.23</td>
</tr>
<tr>
<td>Barren</td>
<td>43</td>
<td>28.28</td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>0.034</td>
</tr>
<tr>
<td>Kruskal-Wallis H&lt;sup&gt;a&lt;/sup&gt;,&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>0.854</td>
</tr>
<tr>
<td>Asymp. Sig.</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>a</sup>. Kruskal Wallis Test  
<sup>b</sup>. Grouping Variable: Land Type
**Ordinary Least Squares Regression**

OLS results indicated that the nitrate concentration in groundwater in Madinah Area was negatively associated to well-depth and the percent of agriculture land but was not statistically significant in relation to the percent of built-up land. The coefficient of well-depth, built-up lands, and agricultural lands are -0.42, -0.16, and -0.23, respectively (Table 5).

The $R^2$ is 0.42 and adjusted $R^2$ is 0.38. An adjusted $R^2$ value of 0.38 would indicate that this model explains approximately 38 percent of the variation in the dependent variable. In other words, 38 percent of nitrate concentrations in Madinah Area can be explained by well-depth and percent of agricultural land (Table 5).

Residuals were mapped and presented in Figure 12 with red points represent where the actual values of nitrate are higher than the model estimated. Blue points are locations where the actual values of nitrate are lower than the model estimated. Most of the wells that are close to the center of the Area are perfectly estimated, while wells in the northeast are lower than estimated. Wells with actual values higher than estimated are located throughout the study area.

The Akaike Information Criterion (AIC) is an indicator of the relative information missed by the model during the estimation process (Eun-Hee et al., 2020). The model with the smaller AIC value is the better model (Fotheringham et al., 2002). In this model, AIC value is 483.4.
Figure 13. Residual map of OLS model

![Residual Map]

Table 5. Summary of OLS results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
<th>t-Statistic</th>
<th>Probability</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>48.14</td>
<td>4.65</td>
<td>10.34</td>
<td>0.000*</td>
<td>-</td>
</tr>
<tr>
<td>Well-depth</td>
<td>-0.42</td>
<td>0.06</td>
<td>-6.15</td>
<td>0.000*</td>
<td>1.15</td>
</tr>
<tr>
<td>Built-up</td>
<td>-0.16</td>
<td>0.10</td>
<td>-1.54</td>
<td>Not Sig.</td>
<td>1.00</td>
</tr>
<tr>
<td>Agriculture</td>
<td>-0.23</td>
<td>0.07</td>
<td>-3.21</td>
<td>0.002*</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Number of Observations 59
R² 0.42
Akaike's Information Criterion (AIC) 483.4
Adjusted R² 0.38

* An asterisk next to a number indicates a statistically significant p-value (p < 0.01).

**Geographically Weighted Regression**

Results of GWR model is presented in Table 6. The table shows the coefficient values of all the variables, R² value, Adjusted R² value, and AIC value within the study area. All variables have a negative correlation with nitrate concentrations, but well-depth has a more negative correlation than others. The coefficient of well-depth, built-up, and agriculture variables are -0.42, -0.20, -0.23, respectively. The R² is 0.46 and an adjusted R² is 0.37. An adjusted R² of 0.37
explains approximately 37 percent of nitrate concentrations in Madinah Area. AIC is 480.5 which is lower than OLS model. And the $R^2$ is 46 which is higher than the OLS model.

Table 6. GWR results

<table>
<thead>
<tr>
<th>Geographically Weighted Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well-depth (coefficient)</td>
</tr>
<tr>
<td>Built-up (coefficient)</td>
</tr>
<tr>
<td>Agriculture (coefficient)</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
</tr>
<tr>
<td>Akaike's Information Criterion (AIC)</td>
</tr>
</tbody>
</table>

Note: There is currently no consensus on how to assess confidence in the coefficients from a GWR model. (ArcGIS Pro).

Figure 14 shows the spatial distribution of the GWR coefficients of well-depth, built-up, and agriculture and the spatial distribution of $R^2$ in the study area. Green color indicates the strength and red color indicates weakness of coefficient. Throughout Madinah Area, there is an inverse relationship between nitrate concentration and well-depth. This indicates that as well-depth decreases, nitrate concentration increases. GWR model showed where the relationship is strong and where it is weak.

As shown in Figure 14a, the inverse relationship between nitrate concentration and well-depth were stronger in the southern part of Madinah Area compared to northern part. It reached -0.47, which indicate a moderate negative correlation. This relationship started to decrease towards the north. It reached -0.39 in the center of Madinah Area and reached its minimum, -0.34, on the northern part of Madinah Area.

As shown in Figure 14b, the negative relationship between nitrate and built-up land was observed to be weak in general but vary spatially. For example, it reached its strongest which is -
0.26 in the center and southern parts of Madinah Area, and this weak relationship was absent in the eastern and northern parts of Madinah Area, where it reached its minimum which is -0.1. This probably due to the lack of urban lands in the eastern and northern parts of Madinah Area, while it reached its strongest in the center of Madinah Area where built-up lands are existed. Again, this means that as the percent of built-up land increased, the nitrate concentration decreased.

Figure 14c shows that the spatial distribution of the negative correlation between nitrate concentration and the percent of agricultural land is mirrors that of well-depth. This relationship was observed stronger in the southern part of Madinah Area compared to northern part of Madinah Area. It reached -0.3, which indicate weak negative correlation. This relationship started to weaken towards the north. It reached -0.25 in the center of Madinah Area. This weak relationship was absent in the northern parts of Madinah Area, where it reached its minimum which is -0.1.

Local $R^2$ indicated that the GWR model is performing well in the southern part of the Madinah Area, but the performance of the model decreases as it goes north (Figure 14d).
Figure 14. Spatial distribution of GWR coefficients and R2. (Note: green indicates strength and red indicates weakness)
DISCUSSION

In this study, the Kruskal-Wallis test did not indicate significant differences between the two land cover groups (barren land and agricultural land). However, by looking at the mean of both land covers in the descriptive statistics (Table 3), the nitrate concentration exceeded recommended values in both. The highest nitrate concentration overall was found in barren area.

The results of OLS and GWR indicated that groundwater nitrate concentration in Madinah Area controlled by well-depth. As the well-depth increased nitrate concentration decreased. Many research results support this finding (MacLeod et al., 1995; Lotfata, 2019; Uhlman & Artiola, 2011; Liu et al., 2005). This is likely because the water in shallow wells is closer to the land surface and potential sources of contamination (MacLeod et al., 1995). The average well-depth in Madinah Area is 44 m. Other studies found that nitrate is likely to occurred in wells that are deeper than 100 m in average such as in Texas aquifer (Lotfata, 2019). So, there could be source of contamination on the land surface of Madinah Area, other than built-up and agricultural lands. Since the nitrate concentration in groundwater did not increase as the percentage of built-up and agricultural lands increased, as has been found in other studies (Lotfata, 2019; Eun-Hee et al., 2020). In fact, the nitrate concentration in groundwater in Madinah Area tended to increase as the percentage of agricultural land decreased.

In this study, the OLS model has a low adjusted R² which indicated that the OLS model has low goodness-of-fit. The goodness-of-fit of the GWR model was better than OLS model, having lower AIC value and a higher R² compared to the OLS model, which is similar to results reported by Eun-Hee et al. (2020) and Lotfata (2019).

There are some limitations in this study. For example, the lack of permanent monitoring well data for the concentration of nitrate in groundwater in the Medina Area, which leads to a decrease in the accuracy of these data since nitrate concentrations in groundwater could vary
spatially and temporally (Arslan et al., 2017). Also, the failure of the well-depth measuring tool to be able to measure depths greater than 100 meters led to less accuracy of well-depth data.

Because of the lack of available data other potential factors were not included here such as intensive livestock farming which is a potential source of nitrate contamination to the surrounding surface and groundwater in many countries (Sahoo et al., 2016; Kei et al., 2021). Also, soil, rain, and irrigation were not used in this study due to the lack of available data. These three variables contribute to high nitrate concentrations in groundwater. Soil converts many types of nitrogen into nitrates, while rain and irrigation lead to the infiltration of nitrates into the groundwater (Washington State Department of Ecology (WSDE), n.d.). However, including these variables in a future study would be beneficial.
CONCLUSION

This study used different statistical methods (Kruskal-Wallis, OLS, and GWR) to determine the potential factors (land cover and well-depth) that affect the concentration of nitrate in groundwater in Madinah Area.

By using OLS and GWR, two factors have identified affecting nitrate concentration of groundwater in Madinah Area which are well-depth and agricultural lands. Both negatively affected the nitrate concentration in Madinah Area, while built-up lands were not statistically significant in its relationship with the groundwater nitrate concentration. The relationship between the nitrate concentration of groundwater and well-depth was moderately negative while agricultural lands had weak negative relationship. The Kruskal-Wallis test did not find any significant differences between land cover types.

This study suggests that there could be source of contamination on the land surface of Madinah Area, other than built-up and agricultural lands. Nitrate concentration in groundwater did not increase as the percentage of built-up and agricultural lands increased, as has been found in other studies. This suggestion is supported by the inverse relationship between the concentration of nitrate in groundwater and well-depth.

Due to the lack of available data, it was necessary to travel to Madinah Area, Saudi Arabia to collect data (primary data) in the field using special equipment (water sample bottles designed especially for collecting groundwater samples and a tool for measuring the depth of the wells) to collect data in the field. Samples were transferred to a laboratory to be analyzed. The Madinah Area has not been previously studied to determine the factors contributing to nitrate contamination in groundwater. This study is the first to focus on Madinah Area to examine the factors contributing to high levels of nitrate concentrations in groundwater.
REFERENCES


https://doi.org/10.1007/s12517-014-1505-9


Lotfata, A. (2019). Geospatial analyses of groundwater depletion and contamination: Multiscale -
global, regional, and local analyses. (Publication No. 2418) [Doctoral dissertation,
Mississippi State University]. ProQuest Dissertations and Theses Global.

contamination in texas aquifers. [Factors affecting the spatial pattern] Management of

Maas, J. A. (2006). Spatial analysis of nitrate contamination of groundwater in Dodge County,
Wisconsin. (Publication No. 3912) [Master’s Thesis, Western Michigan University].
ProQuest Dissertations and Theses Global.

concentrations in ground water to well depth, well use, and land use in Franklin

nitrate concentrations in a region of mixed agricultural land use: A comparison of three
approaches, Environmental Pollution, 15(2), 191-204. https://doi.org/10.1016/S0269-
7491(01)00111-7.

recharged groundwaters in the conterminous United States. Environmental Science and
Technology, 36(10), 2138-2145.

Assessment of nitrate threat to water quality in India. In Y, P. Abrol., T. K. Adhya., V. P.
nitrogen assessment (pp. 323-333). Elsevier.


APPENDIX

Appendix A: The results of nitrate concentration in mg/L and well-depth in the study area.

<table>
<thead>
<tr>
<th>Well Number</th>
<th>Well-depth</th>
<th>Nitrate (NO₃) mg/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&gt;100</td>
<td>7.9</td>
</tr>
<tr>
<td>2</td>
<td>23.1</td>
<td>18.4</td>
</tr>
<tr>
<td>3*</td>
<td>7.5*</td>
<td>291.2*</td>
</tr>
<tr>
<td>4</td>
<td>8.2</td>
<td>96.5</td>
</tr>
<tr>
<td>5</td>
<td>41.3</td>
<td>13.9</td>
</tr>
<tr>
<td>6</td>
<td>51.8</td>
<td>13.4</td>
</tr>
<tr>
<td>7</td>
<td>36.5</td>
<td>33.3</td>
</tr>
<tr>
<td>8</td>
<td>39.1</td>
<td>27.4</td>
</tr>
<tr>
<td>9</td>
<td>14.1</td>
<td>58.4</td>
</tr>
<tr>
<td>10</td>
<td>15.8</td>
<td>30</td>
</tr>
<tr>
<td>11</td>
<td>11.4</td>
<td>41.9</td>
</tr>
<tr>
<td>12</td>
<td>39.7</td>
<td>23.1</td>
</tr>
<tr>
<td>13</td>
<td>62.2</td>
<td>6.8</td>
</tr>
<tr>
<td>14</td>
<td>49.3</td>
<td>8.7</td>
</tr>
<tr>
<td>15</td>
<td>19.7</td>
<td>0.1</td>
</tr>
<tr>
<td>16</td>
<td>81.6</td>
<td>5.9</td>
</tr>
<tr>
<td>17</td>
<td>78.8</td>
<td>6.6</td>
</tr>
<tr>
<td>18</td>
<td>33</td>
<td>18.7</td>
</tr>
<tr>
<td>19</td>
<td>26.4</td>
<td>42.4</td>
</tr>
<tr>
<td>20</td>
<td>84</td>
<td>19.8</td>
</tr>
<tr>
<td>21</td>
<td>83</td>
<td>17.8</td>
</tr>
<tr>
<td>22</td>
<td>89</td>
<td>10.7</td>
</tr>
<tr>
<td>23</td>
<td>32.1</td>
<td>15.4</td>
</tr>
<tr>
<td>24</td>
<td>35.8</td>
<td>14</td>
</tr>
<tr>
<td>25</td>
<td>21.9</td>
<td>55.8</td>
</tr>
<tr>
<td>26</td>
<td>71</td>
<td>7.9</td>
</tr>
<tr>
<td>27</td>
<td>49.6</td>
<td>45.5</td>
</tr>
<tr>
<td>28</td>
<td>19.5</td>
<td>21.2</td>
</tr>
<tr>
<td>29</td>
<td>22.2</td>
<td>32.1</td>
</tr>
<tr>
<td>30</td>
<td>38.4</td>
<td>12.5</td>
</tr>
<tr>
<td>31</td>
<td>54</td>
<td>13.6</td>
</tr>
<tr>
<td>32</td>
<td>76.8</td>
<td>9.9</td>
</tr>
<tr>
<td>33</td>
<td>&gt;100</td>
<td>6.7</td>
</tr>
<tr>
<td>34</td>
<td>37</td>
<td>12.9</td>
</tr>
<tr>
<td>35</td>
<td>29.4</td>
<td>34</td>
</tr>
<tr>
<td>36</td>
<td>31.9</td>
<td>51.2</td>
</tr>
<tr>
<td>37</td>
<td>52.4</td>
<td>16.6</td>
</tr>
<tr>
<td>38</td>
<td>61.5</td>
<td>4.1</td>
</tr>
<tr>
<td>39</td>
<td>53.7</td>
<td>27.9</td>
</tr>
<tr>
<td>40</td>
<td>20.3</td>
<td>37</td>
</tr>
<tr>
<td>41</td>
<td>13.4</td>
<td>19.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>42</td>
<td>24</td>
<td>18.5</td>
</tr>
<tr>
<td>43</td>
<td>20.8</td>
<td>35.3</td>
</tr>
<tr>
<td>44</td>
<td>27.1</td>
<td>45.7</td>
</tr>
<tr>
<td>45</td>
<td>29.8</td>
<td>17.4</td>
</tr>
<tr>
<td>46</td>
<td>&gt;100</td>
<td>2.7</td>
</tr>
<tr>
<td>47</td>
<td>&gt;100</td>
<td>4.1</td>
</tr>
<tr>
<td>48</td>
<td>&gt;100</td>
<td>7.6</td>
</tr>
<tr>
<td>49</td>
<td>67.3</td>
<td>25.1</td>
</tr>
<tr>
<td>50</td>
<td>12.9</td>
<td>14.8</td>
</tr>
<tr>
<td>51</td>
<td>10.5</td>
<td>26</td>
</tr>
<tr>
<td>52</td>
<td>16.1</td>
<td>16</td>
</tr>
<tr>
<td>53</td>
<td>21.8</td>
<td>5.5</td>
</tr>
<tr>
<td>54</td>
<td>17</td>
<td>29.2</td>
</tr>
<tr>
<td>55</td>
<td>13.7</td>
<td>52.8</td>
</tr>
<tr>
<td>56</td>
<td>41.3</td>
<td>11.4</td>
</tr>
<tr>
<td>57</td>
<td>23.1</td>
<td>17.7</td>
</tr>
<tr>
<td>58</td>
<td>30.9</td>
<td>7.4</td>
</tr>
<tr>
<td>59</td>
<td>25</td>
<td>27.1</td>
</tr>
<tr>
<td>60</td>
<td>&gt;100</td>
<td>5</td>
</tr>
</tbody>
</table>

* This water well was omitted from analysis.