9-2018

Assessing the Relationship Between Groundwater Nitrate Concentrations and Environmental Variables Through Repeat Sampling and Statistical Machine Learning: Dutch Flats, Nebraska

Martin Wells

ASSESSING THE RELATIONSHIP BETWEEN GROUNDWATER NITRATE CONCENTRATIONS AND ENVIRONMENTAL VARIABLES THROUGH REPEAT SAMPLING AND STATISTICAL MACHINE LEARNING: DUTCH FLATS, NEBRASKA

by

Martin J. Wells

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Environmental Engineering

Under the Supervision of Professor Troy E. Gilmore

Lincoln, Nebraska

September, 2018

ASSESSING THE RELATIONSHIP BETWEEN GROUNDWATER NITRATE CONCENTRATIONS AND ENVIRONMENTAL VARIABLES THROUGH REPEAT SAMPLING AND STATISTICAL MACHINE LEARNING: DUTCH FLATS, NEBRASKA

Martin J. Wells, M.S.

University of Nebraska, 2018

Advisor: Troy E. Gilmore

Nitrate-contaminated aquifers are common in landscapes dominated by agricultural land use. Health concerns related to consuming nitrate-contaminated groundwater are well documented and continued research aimed at decreasing concentrations is critical. A 1990s U.S. Geological Survey (USGS) study focused on groundwater characteristics in the Dutch Flats area of western Nebraska. Agricultural-related practices were determined to largely influence groundwater recharge and nitrate concentrations ($[NO_3^-]$). Since the conclusion of the USGS study, a transition to more efficient irrigation technology has been observed in this region. The emphasis of this 2016 study was to resample several well nests examined in 1998 to determine whether shifts in water resources management have (1) reduced groundwater recharge rates, (2) increased biogeochemical processes, and (3) reduced groundwater $[NO_3^-]$. Though 2016 $^3H/^3He$ age-dating indicated an increase in groundwater age and decreased recharge rate (19.3 years; 0.35 m/year; n = 8), the mean values were not statistically different from 1998 (15.6 years; R = 0.5 m/year). Samples of $\delta^{15}N\text{-}NO_3^-$ and dissolved oxygen (n=14) did not indicate major changes in biogeochemical processes, including denitrification.

Long-term [NO_3^-] data from the North Platte Natural Resources District (NPNRD) showed 60% of wells sampled in both 1998 and/or 1999 and 2016 (n = 87) had decreased in [NO_3^-], though median concentrations were not statistically different. Given the extensive long-term NPNRD nitrate dataset (n=1,049), we also applied statistical machine learning to (1) evaluate the method as a means to estimate groundwater lag time, (2) assess the influence of 15 predictor variables on Dutch Flats groundwater [NO_3^-], and (3) evaluate the validity of the model through comparisons with field investigations. Overall, Random Forest displayed promising results for evaluating Dutch Flats groundwater [NO_3^-], though limitations were discovered when modeling temporal data.

ACKNOWLEDGMENTS

There are several people I would like to recognize for making this accomplishment possible. First and foremost, I would like to thank my advisor, Dr. Troy Gilmore, for the endless amount of support, guidance, and knowledge along the way. Dr. Gilmore has been an invaluable mentor over the past two years, always going above and beyond my hopes and expectations for an advisor. Dr. Aaron Mittelstet, thank you for assisting in several essential discussions leading to the formulation of numerous ideas applied to this research. I would also like to extend my sincere gratitude to Dr. Natalie Nelson, of North Carolina State University, for her willingness, and what is more important, time, when it came to conceptualizing certain concepts of this project. A special thank you to my committee members, Dr. Gilmore, Dr. Mittelstet, Dr. Nelson, and Dr. Derrel Martin, who all played a role in the success of this project.

Field work was made possible thanks to the exhaustive efforts of Mason Johnson, a fellow graduate student, and Steve Sibray of the University of Nebraska Conservation and Survey Division. I would also like to express gratitude to the Departments of Biological Systems Engineering and Civil Engineering for this experience, and Dr. Bruce Dvorak for his wisdom and help along the way. I will always reflect positively upon the University of Nebraska for the resources it has provided to complete this achievement. A final thank you is necessary to those not mentioned, including family and friends, that have helped me get to where I am today.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION TO STUDY

Nebraska is historically known for agricultural production, where corn, among other crops, requires high inputs of water. Unpredictable precipitation from one year to the next has led to the development of extensive irrigation networks throughout Nebraska. Nebraska has benefited from easily accessible sources of water, including the High Plains Aquifer. Therefore, irrigation is largely supplied from groundwater (Johnson et al., 2011). Diverted streams and rivers provide additional irrigation sources across the landscape. However, water is a resource utilized in several ways. Agriculture, domestic water supply, industry, and wildlife, for example, all depend on water. Finding a balance for supplying these beneficial uses is vital for conserving the valuable resource.

In the early 1970s, Nebraska established Natural Resources Districts throughout the State aimed at sustaining groundwater quantity and quality. Districts have constructed monitoring well networks or utilized irrigation wells from which tens of thousands of samples have been collected, mostly to monitor groundwater nitrate levels (NEDNR, 2016). Nitrate is a concern because of its prevalence in groundwater and associated health effects (van Berk and Fu, 2017, Comly, 1945, Nolan and Weber, 2015, Walton, 1951, Ward et al., 2005, 2018). With 80% of domestic drinking water in Nebraska coming from wells, protecting groundwater quality is essential (Skipton et al., 2011). Nitrate is the most common groundwater contaminant in Nebraska and has been the focus of numerous in-state studies (Böhlke et al., 2007, Exner et al., 2010, 2014, Gormly and Spalding, 1979, Helgesen et al., 1994, Spalding et al., 2001, Verstraeten et al., 2001a, 2001b). Elevated nitrate concentrations ($[NO_3^-]$) >2 mg N L^{-1} (Mueller and Helsel, 1996) are commonly observed in agriculturally dominated watersheds of Nebraska (Gormly and

Spalding, 1979), including many locations where [NO_3^-] exceed the maximum contaminant level (MCL) of 10 mg N L^{-1} in drinking water.

With numerous rural economies dependent on agriculture, concerns over nitrate-contaminated aquifers continue to increase, especially in areas where [NO_3^-] above the MCL may require costly investments in drinking water treatment facilities and/or limit availability of potable water for domestic wells. The semi-arid North Platte Natural Resources District (NPNRD), in western Nebraska, is no exception to this trend. Groundwater characteristics in the NPNRD have been the emphasis of several studies dating back to the 1950s in a small sub-region known as the Dutch Flats area. The U.S. Geological Survey (USGS) intensely sampled the Dutch Flats area from 1995 to 1999 (Böhlke et al., 2007, Verstraeten et al., 2001a, 2001b). Groundwater age-dating, hydrogen and oxygen isotopes, and [NO_3^-], among other groundwater constituents, were evaluated and results showed that the groundwater quantity and quality were greatly influenced by irrigation and canal systems.

In Chapter 2, we evaluate the Dutch Flats area by directly comparing samples we collected in 2016 to those of the USGS study in 1998. Over the nearly two-decade span, irrigation practice largely shifted from furrow irrigation to center pivot sprinkler systems, with the largest rate of change from 2000 to 2003. Given that center pivot irrigation is a more efficient irrigation method, our interest was in evaluating potential changes in the groundwater system related to this and other environmental variables in the area. More specifically, we investigated potential changes in recharge rates (through tritium-helium groundwater age-dating), biogeochemical processes (through dissolved oxygen and nitrogen isotopes of nitrate), and aquifer [NO_3^-]. Long-term nitrate data collected from

the NPNRD within the Dutch Flats, in addition to four groundwater-relevant variables, were also investigated. The objective of this study was to determine spatial and temporal trends in Dutch Flats groundwater as variables found to previously impact groundwater [NO_3^-] have changed.

Chapter 3 evaluates applying advanced statistical analysis to characterize groundwater [NO_3^-] in Dutch Flats. While often considered a black box, Random Forest is a popular machine-learning tool capable of assessing large nonparametric datasets and was used as a complementary analysis to Chapter 2. Random Forest was implemented in a relatively novel manner, perhaps applicable to future studies estimating groundwater transport rates through the vadose and saturated zones. Random Forest also calculated variable importance (i.e., ranking of best to worst predictors) and partial dependencies (i.e., graphical relationship between predictors and modeled [NO_3^-]) for each predictor. Results from variable importance and partial dependencies were then compared to findings by Böhlke et al. (2007) and Chapter 2.

References

Berk, W. van; Fu, Y. (2017) Redox Roll-Front Mobilization of Geogenic Uranium by Nitrate Input into Aquifers: Risks for Groundwater Resources. *Environmental Science & Technology*, **51** (1), 337–345.

Böhlke, J. K.; Verstraeten, I. M.; Kraemer, T. F. (2007) Effects of surface-water irrigation on sources, fluxes, and residence times of water, nitrate, and uranium in an alluvial aquifer. *Applied Geochemistry*, **22** (1), 152–174.

Comly, H. H. (1945) Cyanosis in infants caused by nitrates in well water. *Journal of the American Medical Association*, **129** (2), 112.

Exner, M. E.; Hirsh, A. J.; Spalding, R. F. (2014) Nebraska's groundwater legacy: Nitrate contamination beneath irrigated cropland. *Water Resources Research*, **50** (5), 4474–4489.

Exner, M. E.; Perea-Estrada, H.; Spalding, R. F. (2010) Long-Term Response of Groundwater Nitrate Concentrations to Management Regulations in Nebraska's Central Platte Valley. *The Scientific World Journal*, **10**, 286–297.

Gormly, J. R.; Spalding, R. F. (1979) Sources and Concentrations of Nitrate-Nitrogen in Ground Water of the Central Platte Region, Nebraskaa. *Ground Water*, **17** (3), 291–301.

Helgesen, J. O.; Zelt, R. B.; Stamer, J. K. (1994) Nitrogen and phosphorus in water as related to environmental sevi'ing in nebraska. *Journal of the American Water Resources Association*, **30** (5), 809–822.

Johnson, B.; Thompson, C.; Giri, A.; NewKirk, S. V. (2011) Nebraska Irrigation Fact Sheet (Fact Sheet No. 190), 6; Lincoln, NE: University of Nebraska - Lincoln, Department of Agricultural Economics.

Mueller, D. K.; Helsel, D. R. (1996) Nutrients in the Nation's Waters--Too Much of a Good Thing? (Report No. 1136). Circular; United States Geological Survey.

NEDNR. (2016) Quality-Assessed Agrichemical Contaminant Database for Nebraska Ground Water; Nebraska Department of Natural Resources. Retrieved from https://clearinghouse.nebraska.gov/Clearinghouse.aspx

Nolan, J.; Weber, K. A. (2015) Natural Uranium Contamination in Major U.S. Aquifers Linked to Nitrate. *Environmental Science & Technology Letters*, **2** (8), 215–220.

Skipton, S. O.; Dvorak, B. I.; Woldt, W. E. (2011) An Introduction To Drinking Water (NebGuide No. G1539); Lincoln, NE: University of Nebraska – Lincoln. Retrieved from http://extensionpubs.unl.edu/

Spalding, R. F.; Watts, D. G.; Schepers, J. S.; Burbach, M. E.; Exner, M. E.; Poreda, R. J.; Martin, G. E. (2001) Controlling Nitrate Leaching in Irrigated Agriculture. *Journal of Environmental Quality*, **30** (4), 1184.

Verstraeten, I.M.; Steele, G. V.; Cannia, J. C.; Hitch, D. E.; Scripter, K. G.; Böhlke, J. K.; Kraemer, T. F.; et al. (2001a) Interaction of surface water and ground water in the Dutch Flats area, western Nebraska, 1995-99 (Report No. 2001–4070). Water-Resources Investigations Report, 56; United States Geological Survey.

Verstraeten, I.M.; Steele, G. V.; Cannia, J. C.; Böhlke, J. K.; Kraemer, T. E.; Hitch, D. E.; Wilson, K. E.; et al. (2001b) Selected field and analytical methods and analytical results in the Dutch Flats area, western Nebraska, 1995-99 (Report No. 2000–413). Open-File Report, 53; Reston, VA: United States Geological Survey.

Walton, G. (1951) Survey of Literature Relating to Infant Methemoglobinemia Due to Nitrate-Contaminated Water. *American Journal of Public Health and the Nation's Health*, **41** (8 Pt 1), 986–996.

Ward, M. H.; deKok, T. M.; Levallois, P.; Brender, J.; Gulis, G.; Nolan, B. T.; VanDerslice, J. (2005) Workgroup Report: Drinking-Water Nitrate and Health—Recent Findings and Research Needs. *Environmental Health Perspectives*, **113** (11), 1607–1614.

Ward, M.; Jones, R.; Brender, J.; Kok, T. de; Weyer, P.; Nolan, B.; Villanueva, C.; et al. (2018) Drinking Water Nitrate and Human Health: An Updated Review. *International Journal of Environmental Research and Public Health*, **15** (7), 1557.

CHAPTER 2: ASSESSING DECADAL TRENDS OF A NITRATE-CONTAMINATED SHALLOW AQUIFER IN WESTERN NEBRASKA USING GROUNDWATER ISOTOPES, AGE-DATING, AND MONITORING[1]

2.1. Abstract and Keywords

Abstract: Shallow aquifers are prone to nitrate contamination worldwide. In western Nebraska, high groundwater nitrate concentrations ([NO_3^-]) have resulted in the exploration of new groundwater and nitrogen management regulations in the North Platte Natural Resources District (NPNRD). A small region of NPNRD ("Dutch Flats") was the focus of intensive groundwater sampling by the U.S. Geological Survey from 1995 to 1999. Nearly two decades later, notable shifts have occurred in variables related to groundwater recharge and [NO_3^-], including irrigation methods. The objective of this study was to evaluate how changes in these variables, in part due to regulatory changes, have impacted nitrate-contaminated groundwater in the Dutch Flats area. Groundwater samples were collected to assess changes in: (1) recharge rates; (2) biogeochemical processes; and (3) [NO_3^-]. Groundwater age increased in 63% of wells and estimated recharge rates were lower for 88% of wells sampled (n = 8). However, mean age and recharge rate estimated in 2016 (19.3 years; R = 0.35 m/year) did not differ significantly from mean values determined in 1998 (15.6 years; R = 0.50 m/year). δ^{15}N-NO_3^- (n = 14) and dissolved oxygen data indicate no major changes in biogeochemical processes.

---

[1] Published in *Water*

Wells, M.; Gilmore, T.; Mittelstet, A.; Snow, D.; Sibray, S. Assessing Decadal Trends of a Nitrate-Contaminated Shallow Aquifer in Western Nebraska Using Groundwater Isotopes, Age-Dating, and Monitoring. *Water*. **2018**, *10*, 1047, doi:10.3390/w10081047.

Available long-term data suggest a downward trend in normalized [NO_3^-] from 1998 to 2016, and lower [NO_3^-] was observed in 60% of wells sampled in both years (n = 87), but median values were not significantly different. Collectively, results suggest the groundwater system is responding to environmental variables to a degree that is detectable (e.g., trends in [NO_3^-]), although more time and/or substantial changes may be required before it is possible to detect significantly different mean recharge.

Keywords: groundwater nitrate; groundwater age; groundwater transit time; groundwater recharge rates; non-point source pollution; groundwater monitoring; isotopes; $^3H/^3He$; surface irrigation; center pivot irrigation

2.2. Introduction

Elevated groundwater nitrate concentrations ([NO_3^-]) in shallow aquifers are often linked to a combination of high groundwater recharge rates and intensive agricultural land use [1–6]. Greater recharge rates in areas with intense nitrogen fertilizer loading generally lead to higher [NO_3^-] in groundwater. For example, the central Wisconsin sand-plains region requires additional water and fertilizer inputs to sustain healthy crop yields, with irrigated agriculture having a governing influence on groundwater [NO_3^-] [7,8]. Similarly, high [NO_3^-] have been observed in groundwater in Nebraska, especially beneath areas with sandy soils and/or sand and gravel aquifers [9–13].

Growing concerns over changes in the state's water quality and quantity led to the creation of what are now 23 Natural Resources Districts (NRD) across Nebraska. Established in 1972, NRDs develop management plans and regulations to protect groundwater [14–16]. Regulations aimed at decreasing [NO_3^-] in groundwater have

shown some potential for success [4,17], though the exact impacts are not always clear [13,18]. Due to the tendency of nitrate to be transported with recharge water, agricultural water management (i.e., irrigation technology and practices) and groundwater [NO_3^-] are likely to have a direct relationship. In some areas, water allocations and/or moratoriums on new well drilling can incentivize greater irrigation efficiency, which has been found to decrease groundwater [NO_3^-] [12,19–21]. For instance, replacing furrow irrigated fields with sprinkler systems (i.e., center pivot) is one method believed to reduce [NO_3^-] leaching to groundwater [12,19]. Such changes in irrigation practice, driven in part by regulatory changes and economic drivers, have occurred in the Dutch Flats area of western Nebraska.

Groundwater age-dating has been used widely to determine historical trends in groundwater [NO_3^-] [17,18,22–25]. The U.S. Geological Survey's (USGS) National Water-Quality Assessment (NAWQA) program has emphasized the importance of implementing groundwater age-dating to evaluate long-term trends in groundwater characteristics and its contaminants [24]. Relatively few studies, however, have used groundwater age-dating to directly evaluate the impact of water and/or nutrient management regulations, or major shifts in irrigation management, on groundwater recharge rates and nitrate contamination. Visser et al. [17] used $^3H/^3He$ age-dating in the Netherlands to examine impacts of legislation aimed at decreasing groundwater [NO_3^-] in areas characterized by alluvial sand and gravel deposits. Groundwater age-dating showed trend reversal in groundwater nitrate, with old groundwater increasing in [NO_3^-], and young groundwater decreasing in [NO_3^-]. Further, groundwater age-dating provides a method for evaluating impacts of land use change on groundwater quality [26–28]. As a

result, coupling groundwater [NO_3^-] trends with apparent age may be useful in assessing how changing groundwater recharge and quality respond to irrigation management changes, in the context of other environmental variables (e.g., precipitation).

Within North Platte Natural Resources District (NPNRD) in western Nebraska, irrigation canals provide a source of artificial recharge [11,29–33]. Locations of highest and lowest recharge potential in canals were captured using capacitively coupled and direct-current resistivity methods to profile lithology of two major canals in NPNRD [32]. Estimates suggest canals leak between 40% and 50% of their water within this region [34]. The Interstate Canal, with a water right of 44.5 m^3/s, operates during irrigation season and is the largest canal delivering water to the region [35]. Other large canals in the region include the Mitchell-Gering and Tri-State Canals. An extensive analysis of the relationship between surface irrigation and groundwater quantity and quality in this area was provided by Böhlke et al. [11].

Böhlke et al. [11] also summarized USGS reports from a five-year study beginning in the mid-1990s [30,31]. The investigation was conducted from 1995 to 1999 (referred to in this text as the 1990s study) in the Dutch Flats area, a region comprising roughly four percent of NPNRD. Crop production in this area historically depends on surface water with low [NO_3^-] (e.g., [NO_3^-] < 0.06 mg N L^{-1} in 1997) delivered via canals for irrigation supply. Groundwater age estimates ($^3H/^3He$), isotopes, nitrate, and other analytes were used to evaluate trends in groundwater recharge and nitrate contamination. For example, groundwater recharge rates and temporal changes in [NO_3^-] demonstrated the influence of canal seepage on nearby wells. Wells far from canals were more influenced by local irrigation practices. Relatively young groundwater ages (mean =

8.8 years) indicated that recharge was occurring from more than just regional precipitation (i.e., groundwater would be expected to reside in the aquifer much longer if recharge rates based on precipitation were assumed). As a result, Böhlke et al. [11] theorized that groundwater residence times and [NO_3^-] may be impacted if recharge from canals and/or irrigation were significantly reduced. Further, if groundwater residence times were to increase, then potential for biogeochemical activity such as denitrification might also increase, resulting in a decrease in groundwater [NO_3^-].

Since the 1990s USGS study, several variables related to groundwater recharge have changed in the extensively sampled Dutch Flats area. For example, a shift in irrigation practice and canal management have been noted in the region [36], with the largest changes in irrigation practice occurring during approximately 2000–2003. The timing of these changes relative to the USGS study, combined with the relatively young groundwater ages in the aquifer, provides a unique opportunity to evaluate the potential impact of changing water management on the overall timescale of groundwater movement through the aquifer, and subsequent impacts on groundwater quantity (recharge rate) and quality ([NO_3^-]). Other variables we considered were annual precipitation, volume of water diverted into the Interstate Canal, planted corn area, and fertilizer loads.

In this study, we evaluated how changes to water resources management, with respect to numerous underlying variables, have affected leaching and groundwater transport of nitrate nitrogen. More specifically, the objective of this study was to compare the composition of recently collected groundwater samples to those reported by Böhlke et al. [11] for changes in: (1) groundwater recharge rates; (2) biogeochemical processes

(i.e., denitrification) affecting [NO_3^-]; and (3) groundwater [NO_3^-] in the Dutch Flats area.

2.3. Materials and Methods

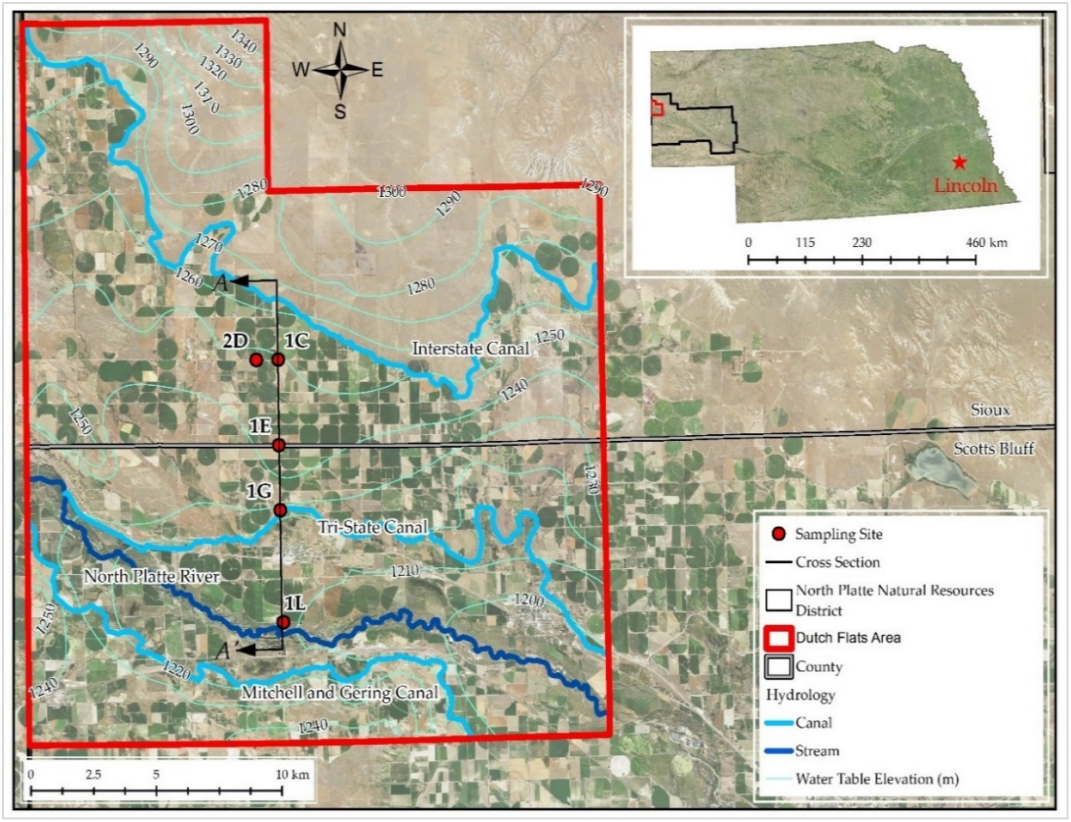
2.3.1. Site Description

The study area is within NPNRD in western Nebraska (Figure 2.1), where climate is classified as semi-arid [37]. Climate data retrieved from Western Regional Airport in Scottsbluff, Nebraska display long-term average annual rainfall and snow of 390 mm and 1021 mm, respectively (1908–2016) [38]. The average annual maximum and minimum temperatures from 1908 to 2016 were 17.6 °C and 1.1 °C, respectively. Growing season rainfall is typically insufficient to support high crop yields; therefore, irrigation is used extensively with 86% previously estimated to originate from surface water [39]. In 2002, a moratorium was implemented to restrict drilling of additional irrigation wells in NPNRD. The state legislature passed Legislative Bill 962 in 2004, allowing the district to declare areas either fully or over-appropriated and led to development of an integrated management plan intended to protect both groundwater and surface water. Regulations on water and soil include groundwater allocations and flow meters on wells in over-appropriated areas, requirements for irrigators using chemigation systems, well registration, and irrigation runoff controls.

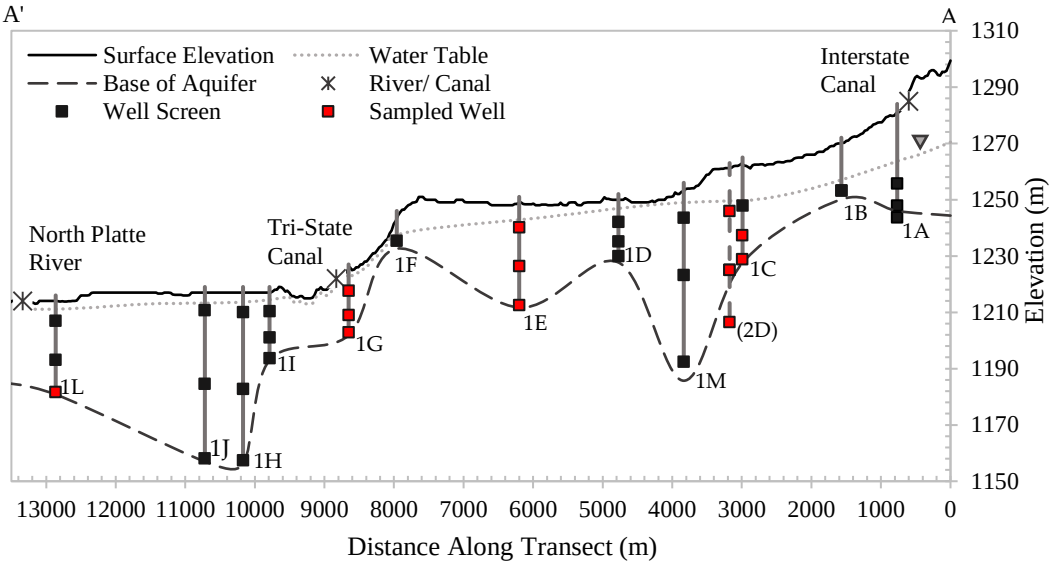
This study is focused within the Dutch Flats area in NPNRD [11,30,31,40]. The study area is in the North Platte River Valley, along the Nebraska-Wyoming border [39]. The Dutch Flats area is about 540 km^2 and located in Scotts Bluff and Sioux Counties (Figure 2.1). Approximately 48% of the study area is in Scotts Bluff County, while 52% is in Sioux County. Based on the 2011 National Land Cover Database (NLCD), 53.5% of

the Dutch Flats area is agriculture, while Scotts Bluff and Sioux Counties are 47.0 and 4.3, respectively [41]. Due to similarities in land use between Dutch Flats area and Scotts Bluff County, Scotts Bluff County was used as a proxy when data could not be determined directly for the study area. While surface water is the most common source for irrigation in this region, accessible groundwater offers alternative methods. Irrigation withdrawal estimates in Scotts Bluff County suggest surface water has remained the dominant source of irrigation water, ranging from 84.4% to 98.6% from 1985 to 2010 [42].

An extensive monitoring well network in NPNRD has been used to measure and record changing groundwater levels and [NO_3^-] over several decades. The Dutch Flats area varies in both vadose and saturated zone thickness, and is characterized as a sand and gravel alluvial aquifer, with limited areas of silt and clay [39] (Figure 2.1). The alluvial aquifer is underlain by the Brule Formation, made up of siltstone, mudstone, volcanic ash beds, gravel, and fine-grained sand. Groundwater for irrigation is typically pumped from Quaternary-aged alluvial deposits or water-bearing units of the Brule Formation. The direction of groundwater flow is generally southeast from canals toward the North Platte River, though flow in some locations is redirected by what is referred to as the Brule High [30].



(**a**)



(**b**)

Figure 2.1. Site description: (**a**) location of Dutch Flats area within Nebraska's North Platte Natural Resources District, including water table elevation contours [43]; and (**b**) representative cross-section along well transect with mid-screen elevations at each well nest. Elevations were derived from previous Dutch Flats studies [11,31]. Red well screens indicate locations where the current study collected groundwater samples. Site 2D, shown with a dashed line and in parenthesis, is situated behind 1C, or into the page, and is located above the base of aquifer in its respective location.

2.3.2. Sample Sites

Within the Dutch Flats area, five well nests were selected for sampling in 2016. Wells were selected based on completeness of data from the previous study [11], so direct comparisons could be made to our results. Samples for noble gases, tritium, nitrogen and oxygen isotopes of nitrate, nitrate nitrogen, ammonium nitrogen, and dissolved organic carbon (DOC) analysis were collected following standard sampling procedures. Groundwater parameters logged and recorded at these five well nests were pH, temperature (°C), dissolved oxygen (mg O_2 L^{-1}), percent saturation of dissolved oxygen, specific conductivity (µS cm^{-1}), and total dissolved gas (g L^{-1}). Each well was measured for depth to groundwater and depth to well bottom relative to surface elevation. At least one well bore volume was pumped from each well prior to sampling (Geotech SS Geosub Controller and Pump). Groundwater parameters were monitored by a Hydrolab MS5 Multiparameter Sonde. After parameters stabilized, pump speed was decreased, and samples were collected. Sampling occurred three times over the course of 2016: spring, summer, and fall. Spring sampling included collection of groundwater from shallow, intermediate, and deep wells from nest 1G (Figure 2.1). Groundwater beneath irrigated fields, and away from major canals, was collected in summer for assessing spatial patterns in groundwater. Shallow, intermediate, and deep samples were collected from Well Nests 2D and 1E, intermediate and deep samples from nest 1C, and deep samples at Well 1L. To capture temporal influences, Well Nest 1G (Figure 2.1) was sampled again at a shallow and intermediate depth during fall.

2.3.3. $^{3}H/^{3}He$ Sampling and Noble Gas Modeling

After rinsing with sample water, 0.9 m copper refrigeration tubes for noble gas analysis were filled and crimped [44]. Samples for analysis of tritium was collected in 0.5-liter HDPE plastic bottles. Noble gases (Ar, He, Kr, Ne, and Xe) were analyzed via mass spectrometry at the University of Utah's Dissolved Gas Service Center in Salt Lake City, UT. Tritium activities were determined using the helium ingrowth method [45].

Excess air and recharge temperatures were modeled using iNoble Version 2.2 workbook developed by the International Atomic Energy Agency (IAEA; Appendix A). We assumed a terrigenic $^{4}He/^{3}He$ (R_{terr}) of 2.88×10^{-8}, and ^{3}H half-life of 12.32 years [46]. Apparent groundwater age was calculated from

$$\tau = \lambda^{-1} \ln\left(1 + \frac{^{3}He_{trit}}{^{3}H}\right) \quad (1)$$

where τ is the apparent groundwater age in years, $^{3}He_{trit}$ is the modeled tritiogenic helium, and ^{3}H is tritium at the time of sampling, determined from helium ingrowth. The decay constant, λ, is a function of the half-life of tritium and is determined as $\lambda = \frac{\ln 2}{12.32 \text{ years}}$.

2.3.4. Recharge Estimates

Groundwater recharge, defined as the rate at which water moves vertically across the water table, was calculated for each well. Shallow wells typically have screen lengths of 6.1 m with the midpoint originally designed to be located at the water table. Intermediate and deep wells both have screens of approximately 1.5 m. To maintain consistency, calculations were performed using methods and assumptions made by Böhlke et al. [11]. Groundwater age was assumed to follow a linear gradient as a slug of water traveling from the water table to the midpoint between the upper and lower well

screen (Equation (2)). For shallow wells, recharge was determined at the midpoint between the bottom screen and water table. Recharge was calculated as follows:

$$R = \frac{z\theta}{\tau} \quad (2)$$

where R is recharge in m/year, θ is porosity, z is depth below water table to the screen midpoint (m), and τ is the groundwater age, in years, determined via apparent $^{3}H/^{3}He$ ages. Porosity was assumed to be 0.35. Intermediate and deep wells are screened below the water table. Therefore, z for these calculations was the distance from the water table to the midpoint of the screen. Again, recharge was estimated using Equation (2). Böhlke et al. [11] also used an exponential equation [47] to estimate recharge rates of intermediate wells, though this equation was not applied to shallow or deep wells (Equation (3)).

$$R = \frac{z\theta}{\tau} \ln\left(\frac{L}{L - z}\right) \quad (3)$$

where L is unconfined aquifer saturated thickness (m), and z is the distance from the water table to the screened midpoint (m). Applying this equation near the water table (shallow wells) or bottom of the aquifer (deep wells) may lead to uncertainties associated with groundwater mixing. While the simplistic modeling of Equations (2) and (3) fit the data reasonably well, it is acknowledged uncertainties related to $^{3}H/^{3}He$ age-dating and aquifer heterogeneity could affect calculated recharge rates. For instance, uncertainty in $^{3}H/^{3}He$ ages can be relatively large on a percentage basis [48]. Further, aquifer heterogeneity, namely porosity, directly influence Equations (2) and (3) calculations. However, assumptions by Böhlke et al. [11] were maintained in both studies, yielding

similar calculated recharge uncertainties. The appropriate applications and limitations of these equations have been addressed in previous literature [49,50].

2.3.5. Nitrate Isotopes, Nitrate, Ammonium, and DOC Concentrations

Samples for isotope analysis were collected in 1-liter HDPE plastic bottles, placed on ice immediately after collection, frozen within 48 h, and analyzed at the University of Nebraska's Water Science Laboratory. The oxygen isotope composition of nitrate was measured according to methods described in Chang et al. [51] and Silva et al. [52]. A measured volume of sample containing 0.25 mg NO_3-N was then treated with 1 M barium chloride to precipitate sulfate and phosphate. The solution was filtered, passed through a cation exchange column to remove excess Ba^{2+}, and then through an anion exchange column to concentrate nitrate. Nitrate was eluted using 3 M hydrochloric acid, neutralized with Ag_2O, filtered to remove the AgCl precipitate, and then dried to produce purified $AgNO_3$. The $AgNO_3$ was dissolved in 1 mL of reagent water and 100 µL (25 µg N) aliquots were transferred to three silver cups and dried for analysis of oxygen isotope composition using high temperature pyrolysis on nickelized graphite in a closed tube to produce carbon monoxide (CO) on a Eurovector EA coupled to an Isoprime continuous flow isotope ratio mass spectrometer. The final result was averaged from the triplicate instrumental results and converted to the standard oxygen isotope reference (VSMOW = 0.00 ‰).

A reagent grade potassium nitrate (KNO_3) was used as a working standard, and reference sucrose oxygen isotope standards were analyzed with every sample batch (up to 20 samples) both for calibration and for drift correction. USGS 34 and USGS 35 reference standards are analyzed at least monthly to compare and convert working

standards to a $\delta^{18}O$ isotope value with respect to Vienna Standard Mean Ocean Water (VSMOW). The 1σ measured analytical precision of $\delta^{18}O\text{-}NO_3^-$ is ±0.5‰ for solutions of KNO_3 standard processed through the entire procedure. In addition to triplicate instrumental average measurement, laboratory duplicates were carried through the preparation process and analyzed at a rate of 5%.

The nitrogen isotope composition of nitrate ($\delta^{15}N\text{-}NO_3^-$) was measured according to methods previously described [53,54]. Ammonia-N was quantified after addition of MgO on a steam distillation line and titrated with standardized sulfuric acid [55]. Nitrate was then reduced to ammonia with Devarda's alloy, distilled separately into a boric acid indicator solution, and then quantified titrimetrically with standardized sulfuric acid. Distillates were acidified with sulfuric acid immediately after titration and evaporated to 1 to 2 mL on a hot plate. They were then reacted with lithium hypobromite on a high-vacuum preparation line and the ammonium quantitatively reduced to nitrogen gas, purified by passage through two liquid nitrogen cryotraps and a 400 °C copper oven, and collected in a gas sample bulb. Atmospheric nitrogen standards were prepared on the same high-vacuum preparation line. Ultrapure tank nitrogen was used as the working standard and was calibrated against the atmospheric nitrogen standard. All nitrogen isotope measurements were performed on either a Micromass OPTIMA or a GVI Isoprime dual inlet stable isotope ratio mass spectrometer (IRMS). The $\delta^{15}N\text{-}NO_3^-$ of the sample was measured and expressed relative to the atmospheric standard expressed in parts per thousand (‰). Quality control was monitored through the analysis of replicate standards to determine the accuracy and repeatability of the method.

The nitrogen ($^{15}N/^{14}N$) and oxygen isotope ($^{18}O/^{16}O$) composition of nitrate was expressed as the difference (‰) of the sample ratio relative to each international standard ratio using Equation (4).

$$\delta(‰) = \frac{(Ratio)_{Sample} - (Ratio)_{Standard}}{(Ratio)_{Standard}} \times 1000 \quad (4)$$

Nitrate concentration was determined for the five well nests selected for detailed sampling in 2016 using the Cd-reduction method [56] on a Seal AQ2 autoanalyzer. Dissolved organic carbon (DOC) samples were collected in 40 mL glass vials and preserved with sulfuric acid. Each sample was field filtered through a 0.45-micron filter attached to a syringe. Samples were analyzed following heated persulfate SM 5310 protocol using an OI Analytical model 1010 TOC Analyzer [56].

2.3.6. Evaluating Long-Term Trends in Nitrate Concentrations

In addition to nitrate evaluated in 1998 and this current study, long-term data collected and/or maintained by the NPNRD and Nebraska Agricultural Contaminant Database [57] were used for further analysis. These samples were collected with a low-volume pump after monitored parameters (temperature, pH, specific conductivity, DO, and total dissolved solids) stabilized. Samples were placed on ice and preserved with sulfuric acid prior to analysis at Midwest Laboratories, Inc in Omaha, NE, USA [58]. Sporadically sampled data from 1998 to 2016 were normalized about the maximum and minimum nitrate value over the sampling period, as

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (5)$$

where x' is the normalized nitrate value, x is the observed (or average if multiple samples were collected from a well within the same year) [NO_3^-] for a specific year, and $\max(x)$

and $\min(x)$ are the maximum and minimum respective [NO_3^-] over all the sample years for each well.

2.4. Results and Discussion

Groundwater samples analyzed for $^3H/^3He$, NO_3^-, $\delta^{15}N\text{-}NO_3^-$, and $\delta^{18}O\text{-}NO_3^-$, among other groundwater parameters, were used to evaluate the hypothesis that changes in environmental variables since the previous study would: (1) decrease recharge rates; (2) increase biogeochemical activity; and (3) result in lower groundwater [NO_3^-]. Unless otherwise noted, data collected in 2016 were analyzed and compared to 1998 groundwater data collected by the USGS [31] in August 1998 and limited to the five well nests sampled in 2016.

2.4.1. Groundwater Age-Dating

An increase in groundwater age between the 1998 and 2016 studies would indicate reduced rates of water movement through the aquifer over that period. Apparent groundwater ages from five of eight samples (63%) collected in 2016 were greater than groundwater ages estimated from samples collected in 1998 (Table 2.1; Figure 2.2). Mean groundwater age in the sampled wells increased from 15.6 years in 1998 to 19.3 years in 2016, but the difference was not statistically significant ($p = 0.53$; two-sample *t*-test assuming unequal variances).

Recent depth to groundwater data (2017; $n = 162$) vary throughout Dutch Flats, ranging from less than 1 m to over 30 m, with a mean of 10.6 m (±10.3 m) and median of 7.7 m. Vadose zone thickness of wells sampled in August 2016 were between 1.2 and 13.3 m. Well nests constructed for the 1990s study (and re-sampled in this study) had screen intervals designed to intercept groundwater at the water table, mid-aquifer, and at

or near the base of the unconfined aquifer. In Dutch Flats, shallow wells typically have 6.1 m screens and were designed originally with roughly 3 m above and below the water table. Long well screens across the water table can increase error in groundwater age, due to mixing of a range of groundwater ages, and because fluctuations in the water table can lead to a loss of tritium-derived $^{3}He_{trit}$ escaping to the atmosphere. Among samples from shallow wells, apparent groundwater age sampled from Well 2D-S increased by approximately 2.5 years, or 93%, while 1E-S decreased by 0.5 years, or 9.3%.

Table 2.1. Apparent groundwater (GW) age from both 1998 and 2016 based on $^{3}H/^{3}He$ age estimates. Recharge rates were estimated with a linear equation in all cases, and with an exponential (Exp.) equation for intermediate wells.

Well ID	Böhlke et al. [11]				Current Study			
	Depth (m) *	GW Age (years)	Recharge -Linear (m/year)	Recharge -Exp. (m/year)	Depth (m)*	GW Age (years)	Recharge -Linear (m/year)	Recharge -Exp. (m/year)
1E-S	2.4	5.4	0.16	n.d.	1.9	4.9	0.13	n.d.
2D-S	4.9	2.8	0.61	n.d.	2.4	5.4	0.15	n.d.
1E-I	15.7	20.2	0.27	0.38	14.5	20.9	0.24	0.34
2D-I	25.8	11.3	0.80	1.2	22.6	20.5	0.39	0.56
1C-D	20.3	12.0	0.59	n.d.	19.4	12.0	0.57	n.d.
1E-D	29.4	31.5	0.33	n.d.	28.3	47.0	0.21	n.d.
1L-D	28.5	20.5	0.49	n.d.	30.3	20.2	0.53	n.d.
2D-D	44.3	20.9	0.74	n.d.	41.1	23.9	0.60	n.d.
Mean:		15.6	0.50			19.3	0.35	
Std. Dev.:		9.5	0.23			13.3	0.19	

Note: * Depth given as depth to mid-screen below water table; n.d., no data; S, Shallow well; I, Intermediate well; D, Deep well; Std. Dev., Standard Deviation.

Because of shorter (1.5 m) screened intervals, intermediate and deep wells may provide a better estimate of groundwater age and are subject to fewer uncertainties impacting shallow wells. Wells 1E-I and 2D-I had comparable groundwater ages of 20.9 and 20.5 years, respectively, in 2016. Well 1E-I groundwater age stayed similar between the two sampling periods (1998 = 20.2 years), while 2D-I increased by 81% (1998 = 11.3 years). Groundwater samples from 1G-I were collected in both the spring and fall of 2016 to explore temporal trends in groundwater age at a site near a canal. Although there is no

direct comparison with the previous study, apparent groundwater age at the well found little variation in groundwater age, with spring and fall ages of just 5.9 and 5.3 years, respectively. Results from Well 1C-I displayed nearly modern groundwater age (i.e., age ≈ 0 years). Results from this well appear erroneous, and were excluded from the comparisons in Table 2.1, since it is unlikely modern groundwater would be observed 11 m below the water table.

Groundwater age of samples collected in 1998 from deep wells ranged from 12.0 to 31.5 years, while in 2016 apparent groundwater ages were between 12.0 to 47.0 years. The largest change in groundwater age was for groundwater sampled from Well 1E-D. While apparent groundwater age stayed similar in Well 1E shallow and intermediate depths, groundwater age from the deep well increased from 31.5 to 47.0 years, or 49%. This increase would suggest groundwater was nearly unaffected by recharge and water sampled in 1998 was essentially the same water collected in 2016. The groundwater age trend in Well Nests 2D and 1E is consistent with non-uniform recharge in the region, where screens at different depths in the aquifer are influenced by different recharge sources (i.e., localized irrigation and/or canals).

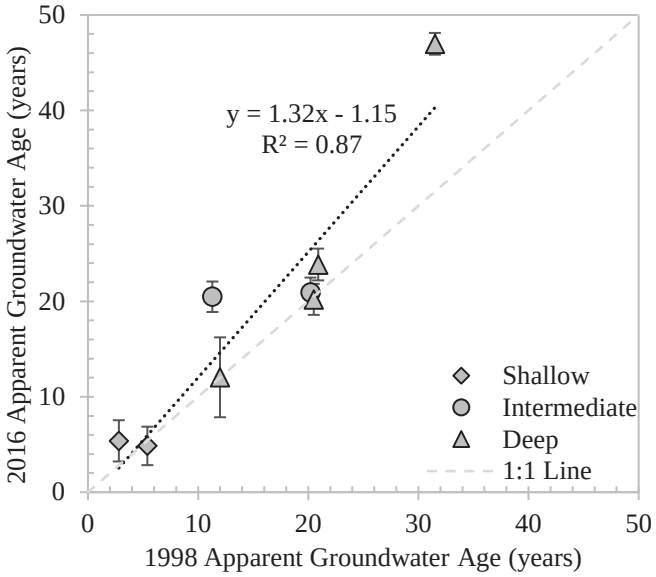


Figure 2.2. Apparent groundwater age determined in 2016 compared to apparent groundwater ages determined in 1998. Error bars are ±1σ from 2016 analysis using the International Atomic Energy Agency (IAEA) model.

2.4.2. Recharge Rates

Recharge rates (*R*) were estimated as a function of the vertical distance water travels below the water table, porosity, and apparent groundwater age (see Equations (2) and (3)). Figure 2.3 and Table 2.1 compare 1998 and 2016 recharge rates. Over nearly two decades, recharge rates decreased in each well, with exception to Well 1L, which had a minor increase. Recharge rates ranged from 0.16 to 0.80 m/year in 1998, with a mean rate of 0.50 m/year (±0.23). Rates in 2016 were not significantly different ($p = 0.19$; two-sample *t*-test assuming unequal variances) and varied between 0.13 and 0.60 m/year, averaging 0.35 m/year (±0.19). From 1998 to 2016, mean water depth increased from 7.9 m to 9.4 m.

Shallow wells far from canals are believed to reflect localized recharge, while intermediate and deep wells are more likely to represent recharge sources from both localized irrigation and canal leakage [11]. Shallow wells had varying results, with a comparable recharge rate in Well 1E-S in both studies, and Well 2D-S less than 25% the

1998 recharge rate. The recharge rate in 1E-I, again, was similar between the two studies, while the 2016 recharge rate in 2D-I was less than half that of 1998. The 2016 mean recharge rate in deep wells was nearly double the mean from shallow wells. Deep wells, typically associated with greater groundwater ages, may take time before they reflect changes to environmental variables related to groundwater quantity. The lowest 1998 (R = 0.33 m/year) and 2016 (R = 0.21 m/year) recharge rates from deep wells were both in 1E-D, which is located far from the larger regional canals. The largest 1998 (R = 0.74 m/year) and 2016 (R = 0.60 m/year) deep well recharge rates were in Well 2D-D.

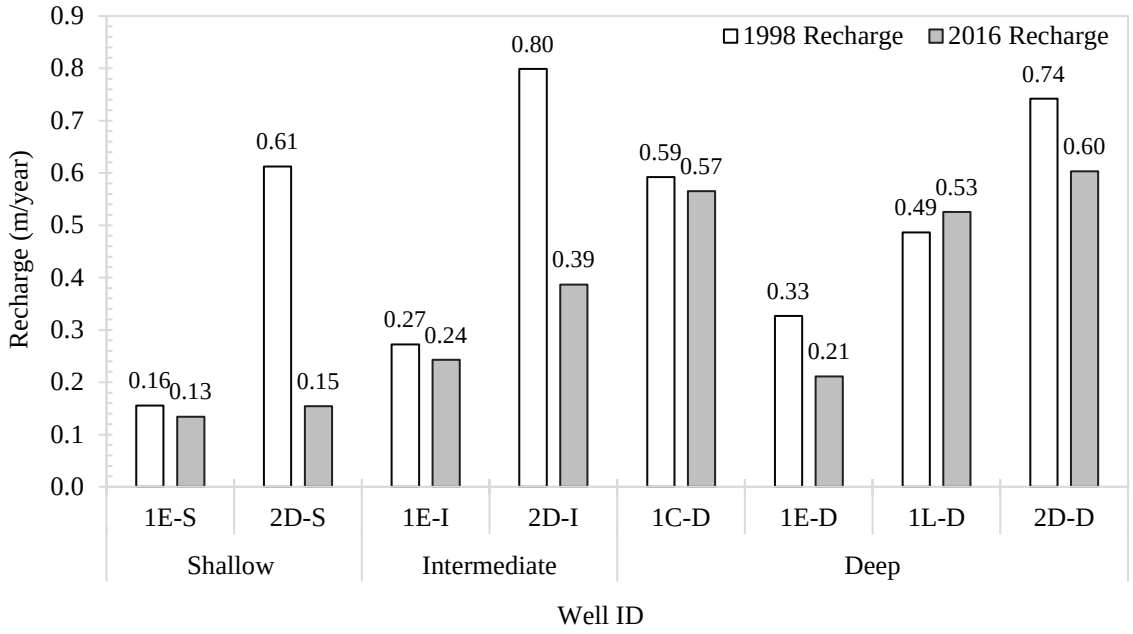


Figure 2.3. Comparison of 1998 and 2016 recharge rates categorized by shallow, intermediate, and deep well depths.

2.4.3. Nitrate Analysis

A combination of data collected in the 1990 and 2016 studies, in addition to long-term groundwater monitoring, provided two approaches to analyze nitrate concentration trends:

I. Comparison of data collected in 1998 [11,31] to data collected in 2016 at the same well nests where groundwater age-dating was conducted (Table 2.2 and Figure 2.4); and

II. Analysis of a long-term dataset from much broader groundwater collection efforts in the Dutch Flats area, including sporadic sampling between 1998 and 2016, and intensive sampling in 1998, 1999, 2008, and 2016 (Table 2.3, Figures 2.5 and 2.6, and Appendix B (Table B1)).

Focusing first on comparisons from Approach (1), nitrate samples collected in 1998 varied from 1.4 mg N L^{-1} to 15.8 mg N L^{-1}, while 2016 ranged from 1.1 mg N L^{-1} to 46.8 mg N L^{-1} (Table 2.2), and 6 out of 14 samples collected in 2016 had lower [NO_3^-] compared to samples collected in 1998 (Figure 2.4). Apart from Well 2D-I (1.3 mg N L^{-1}) and 2D-D (1.4 mg N L^{-1}), groundwater from well nests sampled in 2016 had lower [NO_3^-] with greater depth in the aquifer. Concentrations of ammonium in groundwater were below detection (<0.1 NH_4-N mg L^{-1}) in samples from the eleven wells where $\delta^{18}O$-NO_3^- was also determined (see wells listed in Table 2.2).

Prior to irrigation season, the 2016 Well 1G-S spring nitrate was 46.8 mg N L^{-1}, while post-irrigation season was 22.1 mg N L^{-1}. This trend is consistent with canal leakage diluting concentrations near canals, as suggested by Böhlke et al. [11]. The August 1998 [NO_3^-] was 8.8 mg N L^{-1}. Proximity to a nearby cattle feedlot could be influencing the high 2016 [NO_3^-] observed in 1G-S. It is unknown the exact date operations began at this feedlot, though it is believed between 1998 and 1999, and has increased over the past two decades.

Nitrate concentrations from Well 1E-S were high in both studies (1998 = 15.8 mg N L^{-1}; 2016 = 45.2 mg N L^{-1}), though 2016 was uncharacteristically high. Well 1E-I (3.6 mg N L^{-1}) and 1E-D (3.1 mg N L^{-1}), which had screens approximately 12.7 m and 26.5 m below 1E-S, respectively, had much lower [NO_3^-]. In addition to the 1998 and 2016 values reported above, data collected by NPNRD during 1996–1998 (n = 26) were used for supplementary analysis of Well 1E-S. Monthly averages over this period display increasing [NO_3^-] starting in June, peaking in August, and declining thereafter. If fertilizer is applied around growing season, Well 1E-S displays a very short transport rate through the vadose and saturated zones, in that groundwater [NO_3^-] quickly reflect surface activities. This is supported by a young groundwater age, and possibly suggests there is a preferential pathway to this well screen.

Table 2.2. Nitrate nitrogen and nitrogen isotopic ratio of nitrate from samples collected in 1998 and compared to 2016 samples, at well nests where age-dating was also conducted. Samples collected in 2016 were mostly analyzed for $\delta^{18}O\text{-}NO_3^-$, as shown.

Well ID	Böhlke et al. [11]			Current Study			
	Date Sampled	$\delta^{15}N\text{-}NO_3^-$ (‰)	[NO_3^-] (mg N L^{-1})	Date Sampled	$\delta^{15}N\text{-}NO_3^-$ (‰)	[NO_3^-] (mg N L^{-1})	$\delta^{18}O\text{-}NO_3^-$ (‰)
1G-S	27 August 1998	2.4	8.8	18 April 2016	17.0	46.8	n.d.
1G-I	27 August 1998	n.d.	10.6	18 April 2016	2.6	7.2	n.d.
1G-D	27 August 1998	2.5	8.0	18 April 2016	3.3	6.2	n.d.
2D-S	27 August 1998	5.7	8.3	16 August 2016	0.5	9.7	−4.5
2D-I	27 August 1998	5.6	5.1	16 August 2016	2.2	1.3	−9.16
2D-D	27 August 1998	4.9	1.4	16 August 2016	−2.9	1.4	−6.96
1E-S	24 August 1998	2.9	15.8	16 August 2016	−1.6	45.2	−5.45
1E-I	24 August 1998	2.7	10.8	16 August 2016	−1.3	3.6	−6.6
1E-D	24 August 1998	4.1	2.5	16 August 2016	−3.7	3.1	−5.37
1L-D	25 August 1998	10.2	2.4	17 August 2016	1.1	1.1	−3.67
1C-I	27 August 1998	4.2	2.5	17 August 2016	4.9	5.3	−7.27
1C-D	24 August 1998	4.5	2.5	17 August 2016	−2.0	3.8	−8.38
1G-S	27 August 1998	2.4	8.8	12 October 2016	18.4	22.1	4.08
1G-I	27 August 1998	n.d.	10.6	12 October 2016	9.5	6.9	0.33

Note: n.d., no data; S, Shallow well; I, Intermediate well; D, Deep well.

Well 1G-I was sampled twice in 2016 to evaluate temporal trends in both groundwater age and [NO_3^-] near a canal. Apparent groundwater age was similar (spring

= 5.9 years; fall = 5.3 years). Interestingly, [NO_3^-] in this well decreased from 46.8 to 22.1 mg N L^{-1} in 2016. Similarity in groundwater ages between spring and fall sampling suggest groundwater [NO_3^-] in this well are not diluted from a large percentage of 2016 canal water. That is, a seasonal pattern in [NO_3^-] was observed, but if a significant fraction of canal water that infiltrated during the 2016 growing season was arriving at the well screen by fall 2016, then the groundwater age from fall sampling should be much less than groundwater age from spring sampling. Apparently, the mass flux of water leaking from canals drives groundwater deeper into the aquifer during irrigation season and dilutes groundwater nitrate with older (pre-2016) canal water with low [NO_3^-].

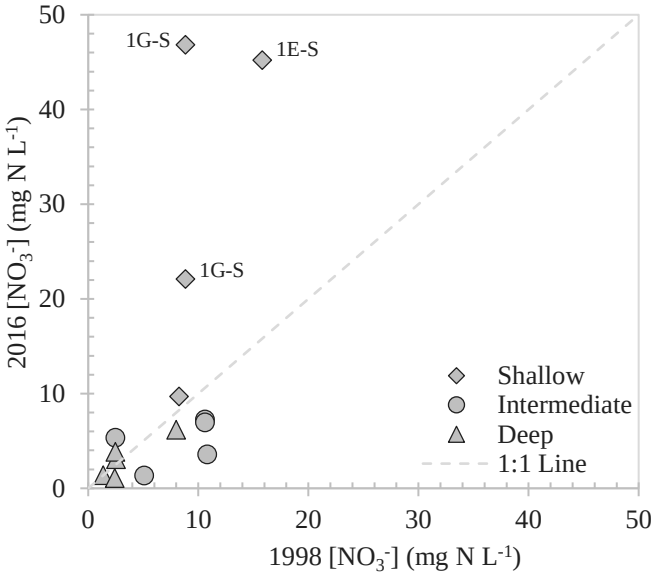


Figure 2.4. Comparison of nitrate concentrations from five well nests sampled in 2016 and 1998 (*n* = 14). Labels indicate the three wells with high 2016 [NO_3^-].

Based on two relatively small datasets (Table 2.2 and Figure 2.4), it is difficult to identify trends in groundwater nitrate due to large variations in concentrations, and additional sources of nitrogen influencing Well 1G. As a result, additional long-term nitrate data collected by NPNRD were analyzed, as described in Approach (2) at the

beginning of this section. In total, 2918 nitrate samples were collected in the Dutch Flats area between 1979 and 2016. However, wells are not consistently sampled, making it difficult to compare overall annual medians from one year to the next. Thus, only data from wells with two or more samples collected between 1998 and 2016 were used (n = 987 samples from a total of 160 wells; Figure 2.5a). If multiple samples were collected from a well within the same year, annual concentrations were averaged. The annual median and mean normalized values (Equation (5)) were then used to evaluate groundwater [NO_3^-] trends in Dutch Flats (Figure 2.5b). Both mean and median of normalized annual nitrate concentrations suggest a decrease in groundwater [NO_3^-] from 1998 to 2016, with statistically significant regression slopes (p-values of 0.04 in both cases).

Given the suggested decrease in [NO_3^-] from the normalized data, we further explored trends in nitrate by evaluating three time periods. Time periods were selected based on the number of samples collected during a given year, as well as with respect to the time elapsed since the previous study. To characterize groundwater [NO_3^-] during the 1990s study, data from 1998 and 1999 were used as a base comparison. Two additional years, 2008 and 2016, were compared directly to [NO_3^-] in the late 1990s. These two years were selected because many samples were collected during 2008 and 2016. A total of 87 wells were sampled during all three time periods, and samples were further split based on screen depth (i.e., shallow, intermediate, and deep wells had n = 44, 16, and 27, respectively).

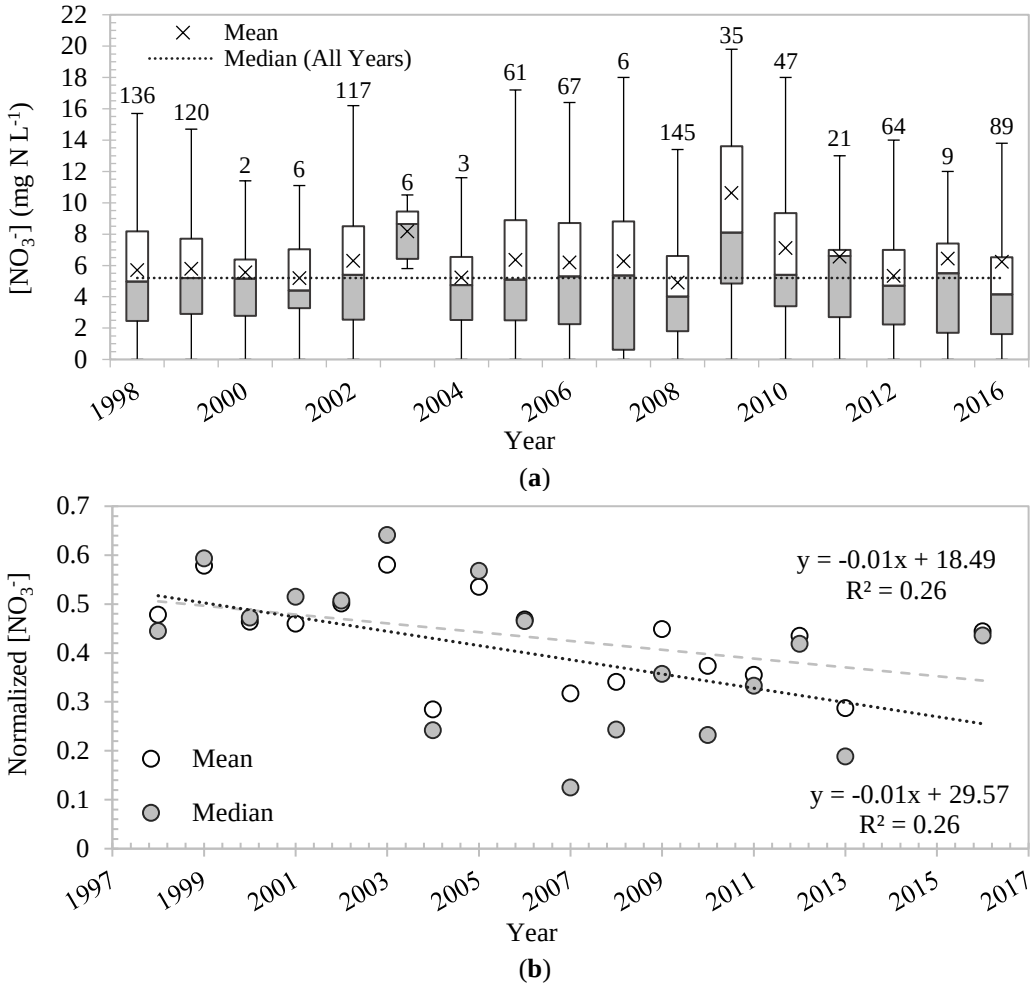


Figure 2.5. Nitrate data from 1998 to 2016 (*n* = 987) collected and/or maintained by the North Platte Natural Resources District, Nebraska Agricultural Contaminant Database, and the current study: (**a**) a Box-and-Whisker plot of all nitrate data, including the number of samples collected each year (referenced above the maxima) and long-term median of each annual median; and (**b**) mean and median of normalized annual Dutch Flats groundwater nitrate.

Overall, 52 out of 87 wells (60%) showed a decrease between 1998 and/or 1999 and 2016. The mean and median [NO_3^-] are reported in Table 2.3 for each well depth, and results from individual samples are shown in Appendix B and Figure 2.6. Since the data were not normally distributed and data transformation did not help, the Mann–Whitney test was used to determine if the median [NO_3^-] were different between the three periods (Table 2.3).

Table 2.3. The mean and median groundwater [NO_3^-] (mg N L^{-1}) for 1998 and/or 1999, 2008 and 2016. The calculated *p*-values are from Mann–Whitney tests comparing the medians between the three time periods.

Well Depth	**Shallow (*n* = 44)**			**Intermediate (*n* = 16)**			**Deep (*n* = 27)**		
	1990s *	**2008**	**2016**	**1990s ***	**2008**	**2016**	**1990s ***	**2008**	**2016**
Mean	7.2	6.2	8.0	6.6	5.6	5.1	4.0	3.5	3.5
Median	5.4	4.7	5.3	5.6	4.8	3.7	3.6	3.3	3.4
Mann–Whitney Test	Years		*p*-value	Years		*p*-value	Years		*p*-value
	1990s *–2008		0.15	1990s *–2008		0.66	1990s *–2008		0.68
	1990s *–2016		0.49	1990s *–2016		0.17	1990s *–2016		0.62
	2008–2016		0.70	2008–2016		0.38	2008–2016		0.94

Note: * Value shown is from 1998 or 1999, or the average from the two years.

The median [NO_3^-] for the 44 shallow wells decreased by 13% from 1998 to 2008, but increased by 13% from 2008 to 2016. These fluctuations may be attributed to the variation in precipitation, with high precipitation resulting in increased leaching rates. From 1996 to 1998 the average precipitation was 459 mm, compared to 303 mm and 497 mm from 2006 to 2008 and 2014 to 2016, respectively.

Though median concentrations were not significantly different (*p*-value 0.17; 5.6 versus 3.7 mg N L^{-1} for 1998 and 2016, respectively), 69% of the intermediate wells sampled had a reduction in [NO_3^-] from 1998 to 2008 and 75% had a reduction from 1998 to 2016. The median [NO_3^-] also decreased in the deep wells, but only by 8% from 1998 to 2008 and 6% from 1998 to 2016, and the differences were not statistically significant.

Although overall [NO_3^-] trends are decreasing in many individual wells, there is substantial variability and uncertainty in overall results, and a lack of statistical significance in median values. Other variables such as vadose zone depth, fertilizer application rates, percent cropland (specifically corn) should be considered.

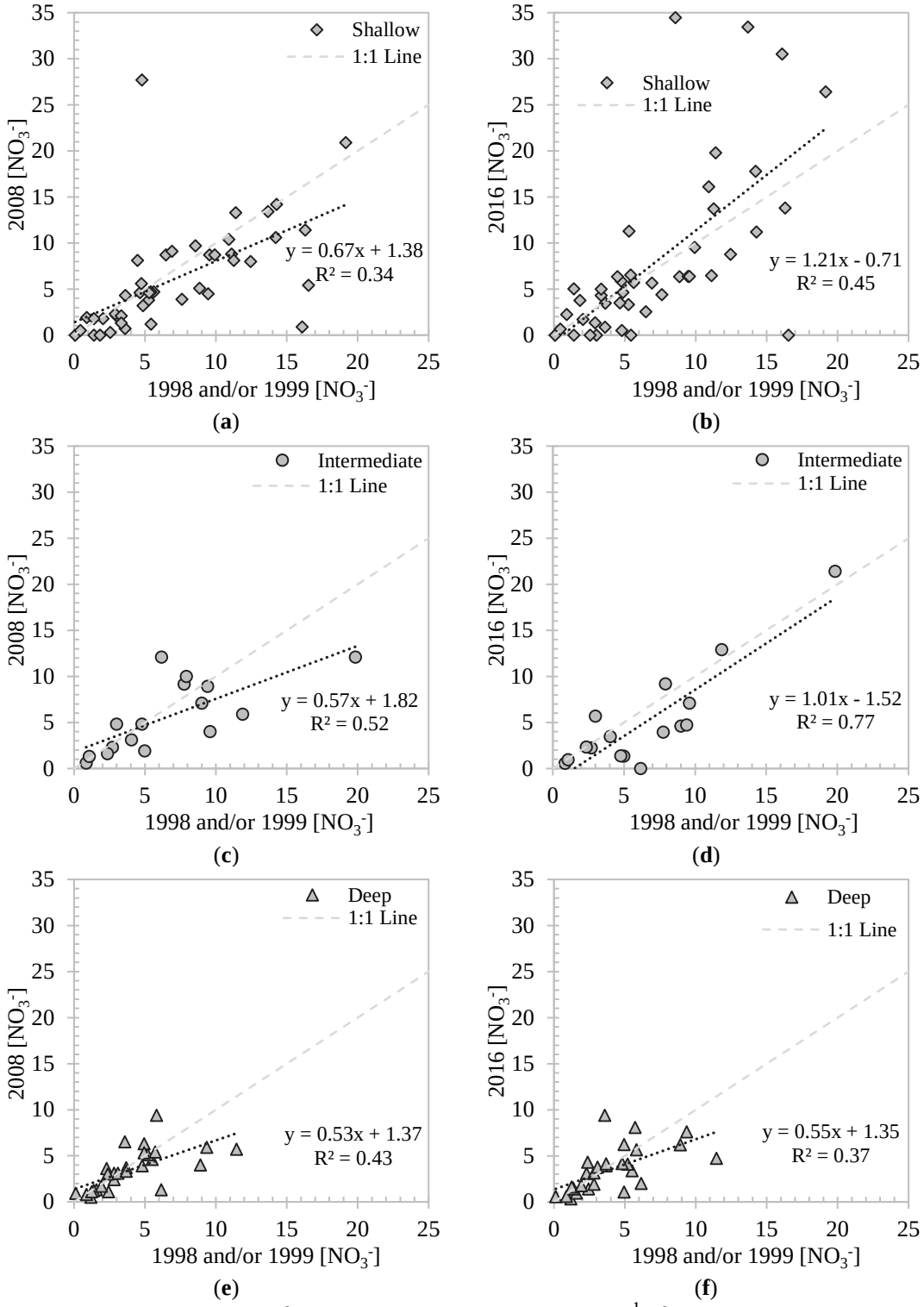


Figure 2.6. Comparison of groundwater [NO_3^-] (mg N L^{-1}) from samples collected in 1998 and/or 1999 in shallow, intermediate, and deep wells to: (**a**) 2008 [NO_3^-] in shallow wells (n = 44); (**b**) 2016 [NO_3^-] in shallow wells (n = 44); (**c**) 2008 [NO_3^-] in intermediate depth wells (n = 16); (**d**) 2016 [NO_3^-] in intermediate depth wells (n = 16); (**e**) 2008 [NO_3^-] in deep wells (n = 27); and (**f**) 2016 [NO_3^-] in deep wells (n = 27).

To develop a better understanding of the complexities of the system, a statistical model coupled with an increase in predictor variables may help explain the large fluctuations in nitrate trends. Continued long-term monitoring of groundwater [NO_3^-], perhaps with a sampling scheduled optimized by well characteristics and/or apparent vulnerability to groundwater nitrate contamination, will be critical for future studies.

2.4.4. Sources of Nitrate

Nitrogen-15 isotopes of nitrate were analyzed to characterize nitrogen sources (Table 2.2 and Figure 2.7). Oxygen-18 isotopes of nitrate were also analyzed in 2016. Nitrogen, of both ^{14}N and ^{15}N, is fixed from the atmosphere to produce fertilizer. Plants favor ^{14}N-NO_3^- to ^{15}N-NO_3^- during assimilation, increasing ^{15}N-NO_3^- observed in groundwater. It is believed this process also results in slightly elevated levels of ^{18}O-NO_3^- [59]. With relatively constant atmospheric ratios of stable oxygen (^{18}O:^{16}O) and nitrogen isotopes (^{15}N:^{14}N), Equation 4 may identify enrichment or depletion of ^{18}O-NO_3^- and ^{15}N-NO_3^- in groundwater. The combination of $\delta^{15}N$-NO_3^- and $\delta^{18}O$-NO_3^- may be used to determine sources of nitrate and potential for denitrification to have affected nitrate [60,61]. Different signatures suggest nitrogen sources may be naturally occurring or associated with anthropogenic activities.

Well Nest 1G was sampled in both spring and fall, but only fall samples were collected for oxygen isotopes. Nitrogen isotopes were collected during each survey (n = 14), but only samples with an oxygen isotope counterpart were used in this analysis (n = 11). Recently collected $\delta^{18}O$-NO_3^- varied between −9.2 and 4.1‰, while $\delta^{15}N$-NO_3^- spanned from −3.7‰ to 18.4‰. The majority of nitrate in Dutch Flats study area appears to be derived from nitrification of ammonium in fertilizer and precipitation. Values from

the 1G shallow well ($\delta^{18}O\text{-}NO_3^-$ = 4.08‰, $\delta^{15}N\text{-}NO_3^-$ = 18.4‰) and intermediate well ($\delta^{18}O\text{-}NO_3^-$ = 0.33‰, $\delta^{15}N\text{-}NO_3^-$ = 9.52‰) had higher nitrogen isotope composition in comparison to the other nine samples. Isotopically heavy $\delta^{15}N\text{-}NO_3^-$ in Well 1G is consistent with nitrate from organic sources (i.e., manure), and is consistent with its proximity to an adjacent feedlot (Figure 2.7).

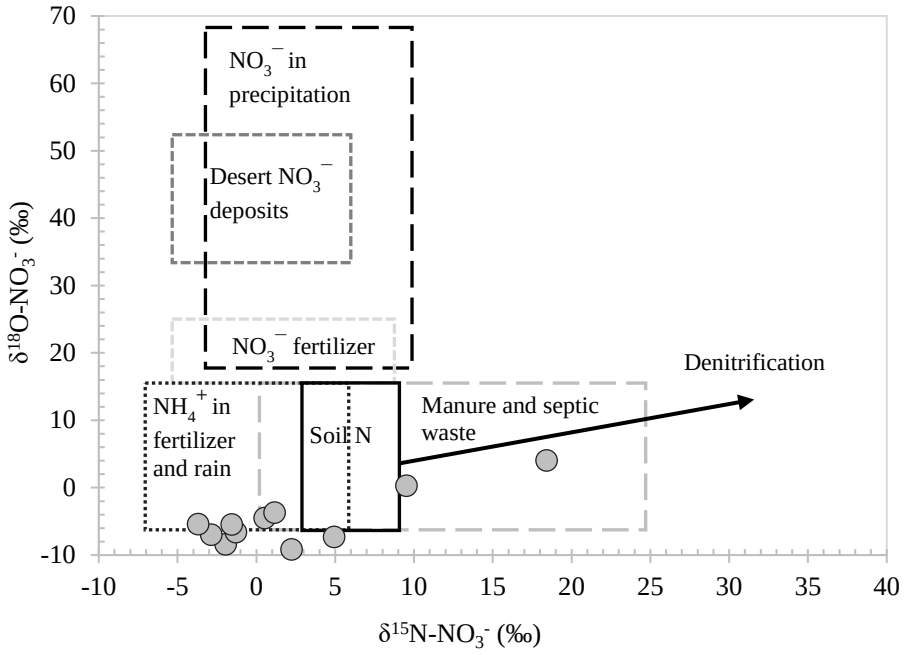


Figure 2.7. Determining sources of nitrogen in Dutch Flats area from oxygen and nitrogen isotopic ratios in nitrate. Figure labels are modified from Kendall [62].

2.4.5. Biogeochemical Processes

Nitrate isotope composition, dissolved oxygen (DO), and DOC were used to indicate changes in biogeochemical activity affecting nitrate, such as denitrification. Results from DO and DOC may be referenced in Appendix B (Figures B1 and B2). Nitrogen isotopes were collected and analyzed for their signatures in both 1998 and 2016 (Table 2.2). Samples from Well 1G-I were not available from the 1990s study. Denitrification could lead to increased $\delta^{15}N\text{-}NO_3^-$. However, data collected in 2016 generally displayed a leftward shift in $\delta^{15}N\text{-}NO_3^-$ compared to 1998. Overall, results do

not suggest an increase in denitrification rates throughout the region (Figure 2.8). Values from 1998 ranged 2.4 to 10.2‰, while 2016 ranged from −3.7 to 18.4‰. Of the 14 samples collected in 2016 for $\delta^{15}N\text{-}NO_3^-$, 12 samples had a 1998 counterpart. Eight samples show stable or even a slight decrease in $\delta^{15}N\text{-}NO_3^-$ (average change = −6.0‰), while two, excluding two outliers at 1G-S, increased (average change = 0.8‰).

Two nitrate isotope sample results stand out, both from Well 1G. From 1998 to 2016, $\delta^{15}N\text{-}NO_3^-$ in Well 1G-S increased from 2.4‰, to 17‰ and 18.4‰ in the spring and fall, respectively. It was noted during collection of groundwater in spring of 2016 that Well 1G-S had a yellow color, and spring-time sample collection occurred immediately after a large precipitation event. Thus, it is possible the organic nitrogen source of nitrate was a result of runoff. Increased nitrogen isotope composition in 1G-S indicates an increased influence of organic nitrogen sources, but could also suggest more prevalent microbiological activity coinciding with a decrease in dissolved oxygen (1998 = 5.6 mg O_2 L^{-1}, fall 2016 = 1.4 mg O_2 L^{-1}), although the groundwater was not strictly anoxic when sampled in 2016. Dissolved organic carbon can serve as a mechanism driving microbial activity and denitrification, and within Well 1G-S, DOC increased from 1998 (2.9 mg C L^{-1}) to 2016 (10.2 mg C L^{-1}). Such increases indicate nitrogen from feedlot manure had little influence on samples collected in 1998, or yet to be in operation. Under anoxic conditions (DO < 0.5 mg O_2 L^{-1}), reduction of nitrate becomes favorable when concentrations are above 0.5 mg N L^{-1} [63]. It is possible recharge from high dissolved oxygen in canal water may prevent conditions from becoming anoxic at this site, and without canal leakage, denitrification could be higher.

Well 1L-D, located near North Platte River, had a water table approximately 2 m below the surface. This well had the largest 1998 (10.2‰) to 2016 (1.1‰) decrease in $\delta^{15}N\text{-}NO_3^-$, although DO (1998 = 0.1 mg O_2 L^{-1}, 2016 = 0.4 mg O_2 L^{-1}) indicated consistent anoxic conditions. Slightly higher 1998 concentrations of NO_3-N and DOC, coupled with lower DO in 1998, may explain the larger $\delta^{15}N\text{-}NO_3^-$ value. From further examination of $\delta^{15}N\text{-}NO_3^-$ versus $1/[NO_3^-]$ and $\delta^{15}N\text{-}NO_3^-$ versus $\ln[NO_3^-]$, mixing analyses (Figure B3) were inconclusive as to whether mixing of high- and low-$[NO_3^-]$ groundwater or denitrification were factors influencing groundwater nitrate in the Dutch Flats area [64].

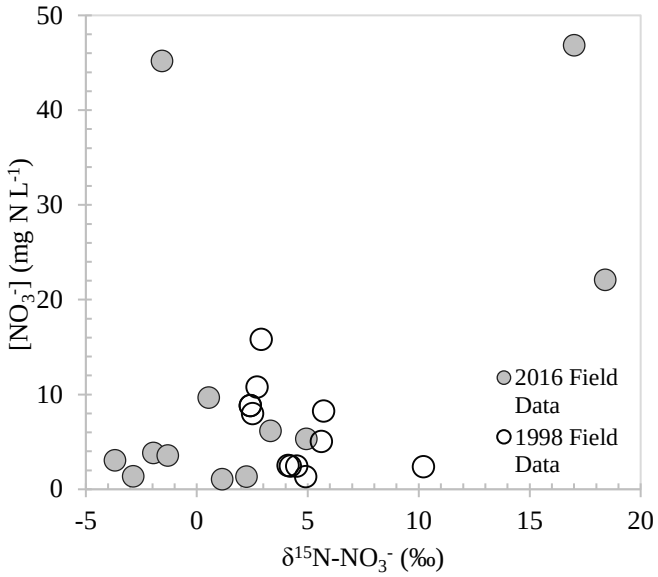


Figure 2.8. Comparison of $[NO_3^-]$ and $\delta^{15}N\text{-}NO_3^-$ for 1998 and 2016 field data. Far right data points are from Well 1G-S and suggest organic nitrogen source.

2.4.6. Analysis of Other Relevant Environmental Variables

Potential changes in relevant environmental variables were evaluated for time periods prior to the 1998 and 2016 studies (i.e., before and after the USGS study). Variables associated with recharge (precipitation and Interstate Canal discharge) and nitrate (planted corn area and fertilizer loads) were analyzed for statistically significant

differences between the two time periods. Precipitation records used were from the Western Regional Airport in Scottsbluff, NE [38]. Annual volume of water diverted into the Interstate Canal was from a gage station approximately 1.6 km downstream of Whalen Diversion Dam in Wyoming [65]. Planted corn area and fertilizer loads are both estimates for Scotts Bluff County [66,67]. Due to limited fertilizer application data, two time periods, each 13 years, were compared from 1987 to 1999, and 2000 to 2012. All other variables were compared over 17-year periods prior to and after the completion of the 1990s USGS study. Datasets were determined to follow a normal distribution, and a two-sample *t*-test was used to evaluate statistically significant differences between two time periods for each variable (Table 2.4).

Besides precipitation ($p = 0.11$), each environmental variable was determined to be significantly different when comparing the two time periods. Both Interstate Canal discharge and fertilizer loads exhibited a reduction, while planted corn area increased between the time periods. The inverse relationship between planted corn area and fertilizer loads is interesting and perhaps suggests an improvement in fertilizer application management, or possibly higher uncertainties associated with county level fertilizer estimates. The reduction in discharge from the canal may be attributed to the change in irrigation management practices.

Table 2.4. Summary of statistical analysis evaluating variables potentially influencing groundwater quantity and quality. *p*-values were determined from two-sample *t*-tests.

Variable	Mean (±std)	Mean (±std)	*p*-value
	Year: 1983–1999	Year: 2000–2016	
Precipitation (mm)	431 (±97)	370 (±118)	0.11
Interstate Canal Discharge (km^3/year)	0.52 (±0.08)	0.44 (±0.08)	0.007 *
Planted Corn Area (hectares)	29,471 (±2568)	34,217 (±2608)	<0.001 *
	Year: 1987–1999	Year: 2000–2012	
Fertilizer Loads (kg)	11,503,061 (±1,150,187)	9,540,057 (±1,222,507)	<0.001 *

Note: * Statistically significant difference between two time periods ($\alpha = 0.05$).

2.4.7. Further Discussion

Since the previous 1990s study, numerous environmental variables related to groundwater nitrate contamination have undergone changes, including shifts in irrigation practices and canal management. Center pivot irrigated area has increased an estimated 270% from 1999 to 2017 within Dutch Flats. Much of the increase has occurred on fields previously irrigated by furrow systems, although there has likely been an increase in overall irrigated acres as well. Scotts Bluff County total irrigated area statistics from the National Agricultural Statistics Service [68,69] estimated an increase in irrigated area over a similar time period (1997 = 70,075 hectares; 2012 = 80,611 hectares). A significant difference in the means was found when comparing volumes diverted into the Interstate Canal from 1983–1999 to 2000–2016. While precipitation also displayed decreasing trends, statistical analysis did not find a significant difference over the same period. With more efficient irrigation methods, less precipitation, and decreased canal discharge, a decrease in recharge rates would be expected. Although seven of the eight wells did have decreased recharge rates, the mean recharge rates from the two studies

were not significantly different. This may be attributed to insufficient time passing between the two sampling periods.

Groundwater denitrification still does not appear to be a major process affecting nitrate attenuation in Dutch Flats since the previous study. Nitrate data collected in 2016 and compared to a small subset of results from the previous study identified a large scatter in data. Three 2016 samples (two from the same well) had high nitrate concentrations, making it difficult to compare trends from the two small datasets. Additional analysis of long-term nitrate data collected by the NPNRD suggests [NO_3^-] have decreased in most wells between 1998 and 2016. Further, two variables related to nitrogen inputs were evaluated during this study: hectares of planted corn and estimated fertilizer loads. The two variables showed significant differences in their means; planted corn area has increased, while fertilizer loads have decreased since the previous study.

Decreased recharge rates were hypothesized to have reduced nitrate concentrations in the Dutch Flats area. However, trends from this study suggest that: (1) improved irrigation efficiency and changes to canal management have yet to significantly influence recharge rates in Dutch Flats; and (2) [NO_3^-] are currently decreasing as a result of a combination of variables, perhaps including improved nitrogen management practices. The second point is further suggested by a significant increase in planted corn area, yet significant decrease in estimated fertilizer application in Scotts Bluff County. In other words, it is possible other variables beyond irrigation practice and canal management are currently driving the decreasing trends in groundwater [NO_3^-], consistent with other studies that emphasize the need to improve both water and nitrogen management in agricultural production [12,13,70,71]. It should be noted, however, that,

while this study was unable to detect a significant reduction in recharge rates, limitations and uncertainties associated with groundwater age-dating and recharge calculations may make it difficult to identify a small but meaningful decrease. If future research were performed in a similar manner and found decreased recharge rates compared to this study, an even greater reduction in groundwater [NO_3^-] is likely to be observed.

It is noteworthy that an estimated increase in irrigated area from 1997 to 2012 is being supplemented by a decrease in annual volume of water diverted into the Interstate Canal, with little indication of increased groundwater withdrawals. Without increased precipitation, these trends, nonetheless, serve as potential evidence of the extent at which irrigation efficiency has improved. Simply put, it is possible less water is being applied to fields regionally, even with increases in irrigated area.

An interesting dynamic to consider is the high leakage potential from canals in this region, and their association with diluting groundwater [NO_3^-]. If future efforts are made to improve irrigation efficiency through lining canals, less artificial recharge will be supplied to the region. Ultimately, this could result in a declining water table elevation, where it has been found artificial recharge is important in restoring aquifer storage and improving groundwater quality [72]. Further, it is unknown how the impact these water management improvements may have on groundwater [NO_3^-]. For instance, less nitrate might leach below the root zone with continued advancements in irrigation efficiency, however, less artificial recharge from low-[NO_3^-] canal water would be present to dilute groundwater [NO_3^-].

2.5. Conclusions

The study area, known locally as the Dutch Flats, has undergone changes in numerous variables influencing groundwater attributes, on timescales similar to those of groundwater movement through the aquifer. This study exemplifies the ability to use intensive snapshot sampling, coupled with long-term continuous data, to evaluate groundwater trends. Varying results in both recharge and [NO_3^-] from this study promote supplementary, and possibly more expansive investigations in the Dutch Flats area. It is possible more time is required to observe changes in groundwater recharge rates. Accounting for lag time through the vadose and saturated zones, some portions of the aquifer may have yet to reflect how changes in environmental variables will impact groundwater quantity and quality. Future resampling in the study area would be beneficial, though carefully-designed long-term monitoring and/or sampling from more wells would offer a better comparison to the more comprehensive 1990s survey. With a vast dataset of nitrate data available through the NPNRD, additional analysis could be beneficial in the identification of variables within Dutch Flats most strongly related to groundwater [NO_3^-]. Existence of the long-term nitrate dataset, coupled with groundwater age-dating to establish a range of lag times, makes for an opportune setting for additional analytics such as machine learning algorithms or classification techniques.

Author Contributions: Conceptualization, T.E.G.; Formal analysis, M.J.W., T.E.G. and A.R.M.; Funding acquisition, T.E.G., D.S. and S.S.S.; Investigation, M.J.W., T.E.G. and S.S.S.; Methodology, T.E.G. and D.S.; Project administration, T.E.G.; Resources, T.E.G.; Supervision, T.E.G.; Writing—original draft, M.J.W.; and Writing—review and editing, M.J.W., T.E.G., A.R.M., D.S. and S.S.S.

Funding: This work was supported by the U.S. Geological Survey 104b Program (Project 2016NE286B), U.S. Department of Agriculture—National Institute of Food and Agriculture (Hatch project NEB-21-177), and Daugherty Water for Food Global Institute Graduate Student Fellowship.

Acknowledgments: The authors acknowledge the North Platte Natural Resources District for providing technical assistance and resources, including long-term groundwater nitrate data accessed via the Quality-Assessed Agrichemical Contaminant Database for Nebraska Groundwater, and Mason Johnson (graduate research assistant) for his support in field sampling efforts. J.K. Böhlke (USGS) also provided insightful comments and suggestions during the project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Almasri, M.N.; Kaluarachchi, J.J. Assessment and management of long-term nitrate pollution of ground water in agriculture-dominated watersheds. *J. Hydrol.* **2004**, *295*, 225–245, doi:10.1016/j.jhydrol.2004.03.013.
2. Burkart, M.R.; Stoner, J.D. Nitrate in aquifers beneath agricultural systems. *Water Sci. Technol.* **2007**, *56*, 59–69, doi:10.2166/wst.2007.436.
3. Ritter, A.; Muñoz-Carpena, R.; Bosch, D.D.; Schaffer, B.; Potter, T.L. Agricultural land use and hydrology affect variability of shallow groundwater nitrate concentration in South Florida. *Hydrol. Process.* **2007**, *21*, 2464–2473, doi:10.1002/hyp.6483.
4. Derby, N.E.; Casey, F.X.M.; Knighton, R.E. Long-term observations of vadose zone and groundwater nitrate concentrations under irrigated agriculture. *Vadose Zone J.* **2009**, *8*, 290–300, doi:10.2136/vzj2007.0162.
5. Lockhart, K.M.; King, A.M.; Harter, T. Identifying sources of groundwater nitrate contamination in a large alluvial groundwater basin with highly diversified intensive agricultural production. *J. Contam. Hydrol.* **2013**, *151*, 140–154, doi:10.1016/j.jconhyd.2013.05.008.
6. Nolan, B.T.; Ruddy, B.C.; Hitt, K.J.; Helsel, D.R. Risk of nitrate in groundwaters of the united states—A national perspective. *Environ. Sci. Technol.* **1997**, *31*, 2229–2236, doi:10.1021/es960818d.
7. Saffigna, P.G.; Keeney, D.R.; Tanner, C.B. Nitrogen, chloride, and water balance with irrigated russet burbank potatoes in a sandy soil. *Agron. J.* **1977**, *69*, 251–257, doi:10.2134/agronj1977.00021962006900020014x.
8. Albertson, P.N. Agricultural Chemicals, Land Use, and Their Impacts on Stream and Ground Water Quality in the Little Plover River Watershed. Master's Thesis, University of Wisconsin—Stevens Point, Stevens Point, WI, USA, August 1998.
9. Piskin, R. Evaluation of nitrate content of ground water in Hall County, Nebraskaa. *Groundwater* **1973**, *11*, 4–13, doi:10.1111/j.1745-6584.1973.tb02987.x.
10. Gosselin, D.C. *Bazile Triangle Groundwater Quality Study*; University of Nebraska: Lincoln, NE, USA, 1991; p. 29.
11. Böhlke, J.K.; Verstraeten, I.M.; Kraemer, T.F. Effects of surface-water irrigation on sources, fluxes, and residence times of water, nitrate, and uranium in an alluvial aquifer. *Appl. Geochem.* **2007**, *22*, 152–174, doi:10.1016/j.apgeochem.2006.08.019.
12. Exner, M.E.; Perea-Estrada, H.; Spalding, R.F. Long-term response of groundwater nitrate concentrations to management regulations in Nebraska's Central Platte Valley. *Sci. World J.* **2010**, *10*, 286–297, doi:10.1100/tsw.2010.25.
13. Exner, M.E.; Hirsh, A.J.; Spalding, R.F. Nebraska's groundwater legacy: Nitrate contamination beneath irrigated cropland. *Water Resour. Res.* **2014**, *50*, 4474–4489, doi:10.1002/2013WR015073.
14. Jenkins, W.J. *A History of Nebraska's Natural Resources Districts*; Hyer, R.B., Ed.; Nebraska Department of Natural Resources: Lincoln, NE, USA, 1975.
15. Exner, M.E.; Spalding, R.F. Groundwater quality and policy options in Nebraska. In *Groundwater quality and policy options in Nebraska*; Smith, R., Ed.; Center for Applied Urban Research, University of Nebraska: Omaha, NE, USA, 1987.

16. Cash, D.W. Innovative natural resource management: Nebraska's model for linking science and decisionmaking. *Environ. Sci. Policy Sustain. Dev.* **2003**, *45*, 8–20, doi:10.1080/00139150309604573.
17. Visser, A.; Broers, H.P.; van der Grift, B.; Bierkens, M.F.P. Demonstrating trend reversal of groundwater quality in relation to time of recharge determined by $^{3}H/^{3}He$. *Environ. Pollut.* **2007**, *148*, 797–807, doi:10.1016/j.envpol.2007.01.027.
18. Wassenaar, L.I.; Hendry, M.J.; Harrington, N. Decadal geochemical and isotopic trends for nitrate in a transboundary aquifer and implications for agricultural beneficial management practices. *Environ. Sci. Technol.* **2006**, *40*, 4626–4632, doi:10.1021/es060724w.
19. Spalding, R.F.; Watts, D.G.; Schepers, J.S.; Burbach, M.E.; Exner, M.E.; Poreda, R.J.; Martin, G.E. Controlling nitrate leaching in irrigated agriculture. *J. Environ. Qual.* **2001**, *30*, 1184–1194, doi:10.2134/jeq2001.3041184x.
20. McMahon, P.B.; Plummer, L.N.; Böhlke, J.K.; Shapiro, S.D.; Hinkle, S.R. A comparison of recharge rates in aquifers of the United States based on groundwater-age data. *Hydrogeol. J.* **2011**, *19*, 779–800, doi:10.1007/s10040-011-0722-5.
21. Baudron, P.; Alonso-Sarría, F.; García-Aróstegui, J.L.; Cánovas-García, F.; Martínez-Vicente, D.; Moreno-Brotóns, J. Identifying the origin of groundwater samples in a multi-layer aquifer system with Random Forest classification. *J. Hydrol.* **2013**, *499*, 303–315, doi:10.1016/j.jhydrol.2013.07.009.
22. Johnston, C.T.; Cook, P.G.; Frape, S.K.; Plummer, L.N.; Busenberg, E.; Blackport, R. Ground water age and nitrate distribution within a glacial aquifer beneath a thick unsaturated zone. *Groundwater* **1998**, *36*, 171–180, doi:10.1111/j.1745-6584.1998.tb01078.x.
23. Katz, B.G.; Chelette, A.R.; Pratt, T.R. Use of chemical and isotopic tracers to assess nitrate contamination and ground-water age, Woodville Karst Plain, USA. *J. Hydrol.* **2004**, *289*, 36–61, doi:10.1016/j.jhydrol.2003.11.001.
24. Rosen, M.R.; Lapham, W.W. Introduction to the U.S. Geological Survey National Water-Quality Assessment (NAWQA) of ground-water quality trends and comparison to other national programs. *J. Environ. Qual.* **2008**, *37*, 190–198, doi:10.2134/jeq2008.0049.
25. Puckett, L.J.; Tesoriero, A.J.; Dubrovsky, N.M. Nitrogen contamination of surficial aquifers—A growing legacy. *Environ. Sci. Technol.* **2011**, *45*, 839–844, doi:10.1021/es1038358.
26. Böhlke, J.K. Groundwater recharge and agricultural contamination. *Hydrogeol. J.* **2002**, *10*, 153–179, doi:10.1007/s10040-001-0183-3.
27. Moore, K.B.; Ekwurzel, B.; Esser, B.K.; Hudson, G.B.; Moran, J.E. Sources of groundwater nitrate revealed using residence time and isotope methods. *Appl. Geochem.* **2006**, *21*, 1016–1029, doi:10.1016/j.apgeochem.2006.03.008.
28. Carlson, M.A.; Lohse, K.A.; McIntosh, J.C.; McLain, J.E.T. Impacts of urbanization on groundwater quality and recharge in a semi-arid alluvial basin. *J. Hydrol.* **2011**, *409*, 196–211, doi:10.1016/j.jhydrol.2011.08.020.
29. Harvey, F.E.; Sibray, S.S. Delineating ground water recharge from leaking irrigation canals using water chemistry and isotopes. *Groundwater* **2001**, *39*, 408–421, doi:10.1111/j.1745-6584.2001.tb02325.x.

30. Verstraeten, I.M.; Steele, G.V.; Cannia, J.C.; Hitch, D.E.; Scripter, K.G.; Böhlke, J.K.; Kraemer, T.F.; Stanton, J.S. *Interaction of Surface Water and Ground Water in the Dutch Flats Area, Western Nebraska, 1995–99*; Water-Resources Investigations Report; United States Geological Survey, United States Department of the Interior: Reston, VA, USA, 2001; p. 56.
31. Verstraeten, I.M.; Steele, G.V.; Cannia, J.C.; Böhlke, J.K.; Kraemer, T.E.; Hitch, D.E.; Wilson, K.E.; Carnes, A.E. *Selected Field and Analytical Methods and Analytical Results in the Dutch Flats Area, Western Nebraska, 1995-99*; United States Geological Survey: Reston, VA, USA, 2001; p. 53.
32. Ball, L.B.; Kress, W.H.; Steele, G.V.; Cannia, J.C.; Andersen, M.J. *Determination of Canal Leakage Potential Using Continuous Resistivity Profiling Techniques, Interstate and Tri-State Canals, Western Nebraska and Eastern Wyoming, 2004*; Scientific Investigations Report; United States Geological Survey, United States Department of the Interior: Reston, VA, USA, 2006; p. 59.
33. Hobza, C.M.; Andersen, M.J. *Quantifying Canal Leakage Rates Using A Mass-Balance Approach and Heat-Based Hydraulic Conductivity Estimates in Selected Irrigation Canals, Western Nebraska, 2007 through 2009*; Scientific Investigations Report; United States Geological Survey, United States Department of the Interior: Reston, VA, USA, 2010; p. 38.
34. Luckey, R.R.; Cannia, J.C. *Groundwater Flow Model of the Western Model Unit of the Nebraska Cooperative Hydrology Study (COHYST) Area*; Nebraska Department of Natural Resources: Lincoln, NE, USA, 2006; p. 63.
35. Nebraska Department of Natural Resources (NEDNR). *Fifty-fifth biennial report of the Department of Natural Resources*; Lincoln, NE, USA, 2009; p. 675.
36. Cannia, J.C. (Aqua Geo Frameworks, Mitchell, NE, USA); Gilmore, T.E. (University of Nebraska, Lincoln, NE, USA). Personal communication, 2016.
37. Steele, G.V.; Cannia, J.C. *Reconnaissance of Surface-Water Quality in the North Platte Natural Resources District, Western Nebraska, 1993*; Water-Resources Investigations Report; United States Geological Survey, United States Department of the Interior: Reston, VA, USA, 1997.
38. NOAA Data Tools | Climate Data Online (CDO) | National Climatic Data Center (NCDC). Available online: https://www.ncdc.noaa.gov/cdo-web/datatools (accessed on 4 August 2017).
39. Verstraeten, I.M.; Sibray, S.S.; Cannia, J.C.; Tanner, D.Q. *Reconnaissance of Ground-Water Quality in the North Platte Natural Resources District, Western Nebraska, June–July 1991*; Water-Resources Investigations Report; United States Geological Survey, United States Department of the Interior: Reston, VA, USA, 1995; p. 114.
40. Babcock, H.M.; Visher, F.N.; Durum, W.H. *Ground-Water Conditions in the Dutch Flats Area, Scotts Bluff and Sioux Counties, Nebraska, With a Section on Chemical Quality of the Ground Water*; United States Geological Survey, United States Department of the Interior: Reston, VA, USA, 1951; p. 51.
41. Homer, C.G.; Dewitz, J.; Yang, L.; Jin, S.; Danielson, P.; Xian, G.Z.; Coulston, J.; Herold, N.; Wickham, J.; Megown, K. Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.* **2015**, *81*, 345–354.

42. Maupin, M.A.; Kenny, J.F.; Hutson, S.S.; Lovelace, J.K.; Barber, N.L.; Linsey, K.S. *Estimated Use of Water in the United States in 2010*; United States Geological Survey: Reston, VA, USA, 2014; ISBN 978-1-4113-3862-3.
43. *Conservation and Survey Division 1995 Water Table Contours*; Institute of Agriculture and Natural Resources, University of Nebraska-Lincoln: Lincoln, NE, USA, 2003.
44. Aeschbach-Hertig, W.; Solomon, D.K. Noble gas thermometry in groundwater hydrology. In *The Noble Gases as Geochemical Tracers*; Burnard, P., Ed.; Springer: Berlin, Germany, 2013; pp. 81–122; ISBN 978-3-642-28835-7.
45. Clarke, W.B.; Jenkins, W.J.; Top, Z. Determination of tritium by mass spectrometric measurement of ^{3}He. *Int. J. Appl. Radiat. Isot.* **1976**, *27*, 515–522, doi:10.1016/0020-708X(76)90082-X.
46. Lucas, L.L.; Unterweger, M.P. Comprehensive review and critical evaluation of the half-life of tritium. *J. Res. Natl. Inst. Stand. Technol.* **2000**, *105*, 541–549, doi:10.6028/jres.105.043.
47. Vogel, J.C. Investigation of groundwater flow with radiocarbon. In *Isotopes in Hydrology*; International Atomic Energy Agency: Vienna, Austria, 1967; pp. 355–369.
48. Gilmore, T.E.; Genereux, D.P.; Solomon, D.K.; Solder, J.E. Groundwater transit time distribution and mean from streambed sampling in an agricultural coastal plain watershed, North Carolina, USA. *Water Resour. Res.* **2016**, *52*, 2025–2044, doi:10.1002/2015WR017600.
49. Broers, H.P. The spatial distribution of groundwater age for different geohydrological situations in the Netherlands: Implications for groundwater quality monitoring at the regional scale. *J. Hydrol.* **2004**, *299*, 84–106, doi:10.1016/j.jhydrol.2004.04.023.
50. Solomon, D.K.; Cook, P.G.; Plummer, L.N. Models of groundwater ages and residence times. In *Use of Chlorofluorocarbons in Hydrology: A Guidebook*; Busenberg, E., Ed.; IAEA: Vienna, Austria, 2006; pp. 73–88.
51. Chang, C.C.; Langston, J.; Riggs, M.; Campbell, D.H.; Silva, S.R.; Kendall, C. A method for nitrate collection for δ^{15}N and δ^{18}O analysis from waters with low nitrate concentrations. *Can. J. Fish. Aquat. Sci.* **1999**, *56*, 1856–1864, doi:10.1139/f99-126.
52. Silva, S.R.; Kendall, C.; Wilkison, D.H.; Ziegler, A.C.; Chang, C.C.Y.; Avanzino, R.J. A new method for collection of nitrate from fresh water and the analysis of nitrogen and oxygen isotope ratios. *J. Hydrol.* **2000**, *228*, 22–36, doi:10.1016/S0022-1694(99)00205-X.
53. Kreitler, C.W. Determining the Source of Nitrate in Ground Water by Nitrogen Isotope Studies. Ph.D. Dissertation, The University of Texas at Austin, Austin, TX, USA, August 1974; p. 57.
54. Gormly, J.R.; Spalding, R.F. Sources and concentrations of nitrate-nitrogen in ground water of the central platte region, Nebraskaa. *Groundwater* **1979**, *17*, 291–301, doi:10.1111/j.1745-6584.1979.tb03323.x.
55. Bremner, J.M.; Keeney, D.R. Steam distillation methods for determination of ammonium, nitrate and nitrite. *Anal. Chim. Acta* **1965**, *32*, 485–495, doi:10.1016/S0003-2670(00)88973-4.

56. American Public Health Association, American Water Works Association, Water Pollution Control Federation. Total Organic Carbon (TOC) Method #5310D, Wet-Oxidation. In *Standard Methods for the Examination of Water and Wastewater*; Washington, DC, USA, 1992.
57. NEDNR; University of Nebraska-Lincoln Quality-Assessed Agrichemical Contaminant Database for Nebraska Ground Water. Available online: https://clearinghouse.nebraska.gov/Clearinghouse.aspx (accessed on 22 September 2016).
58. Hudson, C. (North Platte Natural Resources District, Scottsbluff, NE, USA); Wells, M.J. (University of Nebraska, Lincoln, NE, USA). Personal communication, 2018.
59. Kendall, C.; Aravena, R. Nitrate isotopes in groundwater systems. In *Environmental Tracers in Subsurface Hydrology*; Cook, P.G., Herczeg, A.L., Eds.; Springer US: Boston, MA, USA, 2000; pp. 261–297; ISBN 978-1-4615-4557-6.
60. Amberger, A.; Schmidt, H.L. Natural isotope content of nitrate as an indicator of its origin. *Geochim. Cosmochim. Acta* **1987**, *51*, 2699–2705.
61. Durka, W.; Schulze, E.-D.; Gebauer, G.; Voerkeliust, S. Effects of forest decline on uptake and leaching of deposited nitrate determined from ^{15}N and ^{18}O measurements. *Nature* **1994**, *372*, 765–767, doi:10.1038/372765a0.
62. Kendall, C. Tracing nitrogen sources and cycling in catchments. In *Isotope Tracers in Catchment Hydrology*; Elsevier: Amsterdam, The Netherlands, 1998; pp. 519–576; ISBN 978-0-444-81546-0.
63. McMahon, P.B.; Chapelle, F.H. Redox processes and water quality of selected principal aquifer systems. *Groundwater* **2008**, *46*, 259–271, doi:10.1111/j.1745-6584.2007.00385.x.
64. Mariotti, A.; Landreau, A.; Simon, B. ^{15}N isotope biogeochemistry and natural denitrification process in groundwater: Application to the chalk aquifer of northern France. *Geochim. Cosmochim. Acta* **1988**, *52*, 1869–1878, doi:10.1016/0016-7037(88)90010-5.
65. USBR Hydromet: Archive Data Access. Available online: https://www.usbr.gov/gp/hydromet/hydromet_arcread.html (accessed on 22 May 2018).
66. Brakebill, J.W.; Gronberg, J.M. *County-Level Estimates of Nitrogen and Phosphorus from Commercial Fertilizer for the Conterminous United States, 1987–2012*; United States Geological Survey: Reston, VA, USA, 2017.
67. NASS USDA/NASS QuickStats Ad-hoc Query Tool. Available online: https://quickstats.nass.usda.gov/ (accessed on 15 February 2018).
68. *2012 Census of Agriculture, Nebraska State and County Data*; United States Department of Agriculture, National Agricultural Statistics Service: Washington, DC, USA, 2014.
69. *1997 Census of Agriculture, Nebraska State and County Data*; United States Department of Agriculture, National Agricultural Statistics Service: Washington, DC, USA, 1999.
70. Green, C.T.; Liao, L.; Nolan, B.T.; Juckem, P.F.; Shope, C.L.; Tesoriero, A.J.; Jurgens, B.C. Regional variability of nitrate fluxes in the unsaturated zone and groundwater, Wisconsin, USA. *Water Resour. Res.* **2018**, *54*, 301–322, doi:10.1002/2017WR022012.

71. Mas-Pla, J.; Menció, A. Groundwater nitrate pollution and climate change: Learnings from a water balance-based analysis of several aquifers in a western Mediterranean region (Catalonia). *Environ. Sci. Pollut. Res.* **2018**, *25*, 1–19, doi:10.1007/s11356-018-1859-8.
72. Ma, L.; Spalding, R.F. Effects of artificial recharge on ground water quality and aquifer storage recovery. *J. Am. Water Resour. Assoc.* **1997**, *33*, 561–572, doi:10.1111/j.1752-1688.1997.tb03532.x.

CHAPTER 3: APPLYING MACHINE LEARNING DATA ANALYTICS TO EVALUATE NITRATE-RELATED VARIABLES IN A SHALLOW AQUIFER

3.1. Introduction

Nitrate is the most common contaminant in groundwater, and in Nebraska there are many locations with nitrate concentrations ($[NO_3^-]$) near or above the 10 mg N L^{-1} primary drinking water standard set by the U.S. Environmental Protection Agency (USEPA). Groundwater is the primary drinking water source for 80% of Nebraskans, and as populations decline in rural communities, the cost to relocate wells or treat groundwater increases on a per household basis. The prevalence of aquifer impairment has resulted in widespread efforts to characterize and manage groundwater quality.

Growing concern over groundwater management in Nebraska led to the establishment of 23 Natural Resources Districts (NRD). Since their founding in 1972, NRDs have accumulated substantial datasets characterizing groundwater quality through extensive monitoring well networks. The North Platte Natural Resources District (NPNRD), in western Nebraska, has collected thousands of nitrate samples dating back to 1979. While these data are used to determine aquifer degradation and establish groundwater management areas, spatial and temporal variability of groundwater $[NO_3^-]$ has made it difficult to synthesize water-quality information, identify trends, and/or directly link groundwater $[NO_3^-]$ to governing environmental (e.g., management) variables. Determining variables with the greatest impact on groundwater $[NO_3^-]$ will assist watershed managers in identifying the most efficient management practices to implement.

For this study, data from the NPNRD was refined to focus on groundwater [NO_3^-] within the Dutch Flats area (Figure 1). Dutch Flats was the site of an intensive evaluation in a concurrent study (Wells et al., 2018), where the long-term NPNRD nitrate dataset and groundwater samples collected in 2016 were analyzed for a direct comparison to a previous U.S. Geological Survey (USGS) study from 1995 to 1999 (Böhlke et al., 2007, Verstraeten et al., 2001a, 2001b). References to the 1995 to 1999 study will be stated throughout this text as the "1990s study."

Wells et al. (2018) found that identification of groundwater [NO_3^-] trends was challenging when using basic statistical analyses. As with many groundwater datasets, available [NO_3^-] were not available over regular time-steps (e.g., annually) and/or had relatively few samples collected over the previous 2-3 decades, making it difficult to compare one year to the next and apply time series analysis. Additionally, groundwater [NO_3^-] is driven by a wide range of non-linear/interactive processes, which challenges the use of traditional statistical models that require the selection of a model's functional form. Thus, Wells et al. (2018) concluded that advanced statistical analysis approaches (e.g., statistical machine learning) that are well-suited to handle non-linear environmental data should be applied to further investigate this intensively-studied site.

Data mining may apply statistical machine learning by using algorithms to assess and identify complex relationships between variables. Random Forest, for example, is a machine learning data-analysis method that has been effective in analyzing large datasets, sometimes referred to as "big data" (Baudron et al., 2013, Grimm et al., 2008, Jones and Linder, 2015, Peters et al., 2007, Rahmati et al., 2016, Rodriguez-Galiano et al., 2012, 2014, Wiesmeier et al., 2011), where outliers and non-parametric distributions may

inhibit performance in other analyses. This method also has benefits of valuing predictors that may be otherwise overshadowed in less comprehensive modeling techniques. As Random Forest learns a dataset, arbitrary subsets of predictors are assessed throughout the process, where at times, highly correlated variables are left out. This condition coaxes Random Forest into assessing weaker predictors in the model.

Since Random Forest is a relatively new method for data analysis, its potential applications in water-quality studies are still being explored. Recent studies have utilized Random Forest for modeling [NO_3^-] in aquifers (Anning et al., 2012, Nolan et al., 2014, Ouedraogo et al., 2017, Rodriguez-Galiano et al., 2014, Wheeler et al., 2015). Random Forest has been found to accurately predict groundwater [NO_3^-] in unconsolidated aquifers (Rodriguez-Galiano et al., 2014, Wheeler et al., 2015). Rodriguez-Galiano et al. (2014) used Random Forest to evaluate nitrate pollution in an agricultural region of southern Spain. The study investigated 24 predictors, identifying a relatively low-cost parameter (i.e., electrical conductivity) as the most important variable in predicting groundwater [NO_3^-]. Another comprehensive study in Iowa assembled a large dataset to evaluate the exposure of private well users to nitrate contamination throughout the state (Wheeler et al., 2015). A dataset of over 34,000 observations and nearly 300 predictors were compiled for the study, with well depth having the largest influence on groundwater [NO_3^-] in private wells. These observations are consistent with basic hydrogeological theory and observations, where specific conductance is generally higher (relative to background conditions) in agricultural areas, and nitrate contamination is commonly higher in shallow aquifers or shallow parts of aquifers than in deeper locations.

The Central Valley of California has also been the focus of several data-driven studies, where machine learning has been used to model [NO_3^-] and evaluate variable importance using various techniques (Nolan et al., 2014, 2015, Ransom et al., 2017). Random Forest was found capable for modeling groundwater [NO_3^-] (Nolan et al., 2014), though Nolan et al. (2015) suggested groundwater age would be a beneficial predictor for improving model performance in the region. While the importance of groundwater transit time has been acknowledged in Random Forest modeling of [NO_3^-], studies have primarily used alternative variables as a proxy for estimating groundwater lag. Wheeler et al. (2015) evaluated time-sensitive variables based on their importance in the model. These predictors were used to approximate groundwater lag, with a goal of estimating cancer risk from modeled historical drinking water [NO_3^-]. Lag was estimated in the study by relating the importance of temporally dependent variables, such as agricultural land use and county-level fertilizer, to the sample date. For instance, nitrogen inputs from 1978 to 1990 were found more important in the model compared to inputs from 1992 to 2006, suggesting decadal travel times were a realistic estimate. Depth to screen midpoint and perforated interval have also been used as a substitute for groundwater age (Nolan et al., 2014), where these variables were identified as fairly important predictors. Although Ransom et al. (2017) did not use Random Forest, their investigation used an alternative machine learning method in boosted regression tree modeling. Their study explored the use of groundwater age and reduction-oxidation models to create a three-dimensional representation of the Central Valley Aquifer. Results were unclear as to whether groundwater age or reduction-oxidation conditions played a larger role; however, it was

found these two predictors improved model performance over the previous Central Valley models.

This study is a unique application of the wide range of Random Forest capabilities. While other studies have recognized the importance of groundwater transit time through bulk lag times (i.e., applying a single lag time value to many data) or groundwater age proxies (e.g., using well depth as a proxy for travel time), none have utilized Random Forest to estimate a distinct travel time at each well. Additionally, studies often overlook the importance of considering vadose zone transport when modeling groundwater [NO_3^-]. However, Gerber et al. (2018) investigated the use of environmental tracers to estimate vadose zone transport and found this an important component for improving hydrogeologic models and nitrate predictions. The objective of this study is to implement a data-driven modeling-based approach to estimate vadose and saturated zone transport rates, in addition to interpreting long-term groundwater [NO_3^-] data sampled by the NPNRD (NEDNR, 2016). In other words, Random Forest will be used as a tool to (1) estimate lag time through the vadose and saturated zones by identifying optimal transport rates, and (2) determine variables with governing influences on Dutch Flats groundwater [NO_3^-]. This study will also (3) validate Random Forest results through direct comparisons to the Dutch Flats hydrogeologic understanding developed in previous groundwater studies (Böhlke et al., 2007, Wells et al., 2018). From the second and third objectives, this study will evaluate the legitimacy of applying Random Forest as a means of supplementary analysis to larger physical models, and/or substituting intensive hydrogeologic field investigations.

3.2. Methods

3.2.1. Site Description

The Dutch Flats area is situated on the Nebraska side of the Nebraska-Wyoming border (Figure 3.1). Several previous studies in the Dutch Flats area have investigated groundwater characteristics and provided thorough site descriptions of the semi-arid region (Babcock et al., 1951, Böhlke et al., 2007, Verstraeten et al., 2001a, 2001b). Located within the NPNRD, the roughly 540 km^2 area incorporates parts of Sioux (52%) and Scotts Bluff (48%) Counties. The Dutch Flats area sits upon an alluvial aquifer characterized by unconsolidated deposits of predominantly sand and gravel, with the base of aquifer largely consisting of consolidated deposits of the Brule Formation (Verstraeten et al., 1995). Supplemental irrigation water, or water not derived from the North Platte River, is typically pumped from the alluvial aquifer, or water-bearing units of the Brule Formation.

Approximately 53.5% of Dutch Flats is agriculture, with the majority concentrated south of the Interstate Canal (Homer et al., 2015). Agricultural land use represents 47% and 4% of Scotts Bluff and Sioux Counties, respectively. The combination of intense agriculture and low annual precipitation have producers in Dutch Flats relying on a network of irrigation canals to supply water to the region. From 1908 to 2016, mean rain and snowfall of 390 mm and 1021 mm, respectively, was observed at the nearby Western Regional Airport in Scottsbluff, NE (NOAA, 2017).

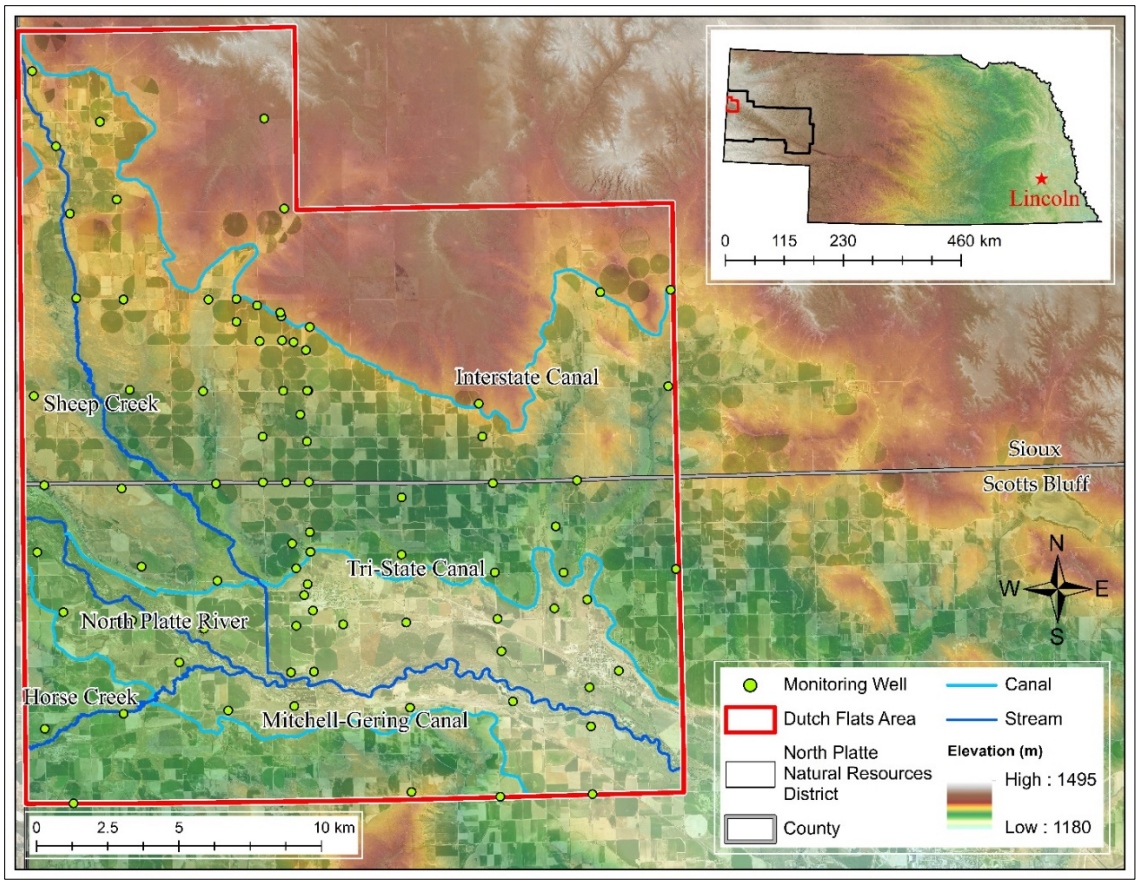


Figure 3.1. Dutch Flats study area overlain by surface elevation Digital Elevation Model. Depending on data availability, multiple wells (well nest) or single wells may be found at each monitoring well location.

While some groundwater is withdrawn, Scotts Bluff County irrigation is typically supplied from surface water sources and has been estimated to deliver between 84.4% to 98.6% of the water from 1985 to 2010 (Maupin et al., 2014). Canals carry water, often derived from the North Platte River, to fields throughout the study area. High leakage from these canals provide a source of artificial groundwater recharge to the region. Previous studies have investigated the leakage potential of canals in the region (Ball et al., 2006, Harvey and Sibray, 2001, Hobza and Andersen, 2010, Luckey and Cannia, 2006). A study by Luckey and Cannia (2006) estimated between 40% to 50% of canal water is lost during conveyance. Larger canals in Dutch Flats include the Mitchell-

Gering, Tri-State, and Interstate Canals, with the latter holding the largest water right of 44.5 m^3/s per year. (NEDNR, 2009). The 1990s study investigated both spatial and temporal influences from canals in the Dutch Flats (Verstraeten et al., 2001a, 2001b), with results later analyzed by Böhlke et al. (2007). Canals were found to dilute groundwater [NO_3^-] near canals with low-[NO_3^-] (e.g., [NO_3^-] < 0.06 mg N L^{-1} in 1997) canal water during irrigation season. Tritium-helium age-dating was used to determine apparent groundwater age and recharge rates. It was noted that wells near canals displayed recharge rates influenced by local canal leakage. Regional wells far from the canals suggested that shallow wells were mostly influenced by local irrigation practices, while deeper wells were impacted by both localized irrigation and canal leakage.

The NPNRD has a large monitoring well network consisting of 797 wells, with 327 of them nested. Nested wells are drilled and constructed such that screen intervals represent shallow groundwater at the water table (screen length = 6.1 m), intermediate aquifer depths (screen length = 1.5 m), and deep groundwater (screen length = 1.5 m) near the base of the unconfined aquifer. This study incorporated 162 monitoring wells sampled and maintained by the NPNRD within Dutch Flats.

Influenced by both regulatory and economic incentives, the Dutch Flats area has undergone a notable shift in irrigation practices in the last two decades. From 1999 to 2017, center pivot irrigated area has increased by approximately 270%. While some of the increase is, in part, due to an increase in the overall irrigated area, the majority has occurred on fields previously irrigated by furrow irrigation (Figure C1). Conventional furrow irrigation has an estimated application efficiency of 45% to 65%, compared to center pivot sprinklers at 75% to 85% (Irmak et al., 2011). As related to groundwater

recharge rates, a shift in irrigation practice was hypothesized to decrease groundwater [NO_3^-] in that less nitrate would leach beneath the root zone, as well as longer residence times could promote greater nitrate reduction. Other long-term changes to the landscape have included statistically significant reductions in mean fertilizer application rates (1987–1999 vs. 2000–2012) and volume of water diverted into the Interstate Canal (1983–1999 vs. 2000–2016), and an increase in area of planted corn (1983–1999 vs. 2000–2016). Precipitation was also evaluated, and though the mean has decreased over a similar time period, it was not statistically significant. The hypothesis was investigated by Wells et al. (2018), and while statistically insignificant, an estimated 88% of wells decreased in recharge rates. Nitrogen isotopes of nitrate found little change in biogeochemical processes, though groundwater [NO_3^-] exhibited promising decreasing trends from 1998 to 2016.

3.2.2. Random Forest Modeling Framework

Random Forests are created by combining hundreds of individual decision trees into one model ensemble, or "forest." In Random Forest, Classification and Regression Trees (CART) are used as the sub-model (Breiman, 1984). CART are grown through partitioning predictors (i.e., independent variables) with the goal of creating a homogenous outcome, or a grouping of similar observations, referred to as the response variable. Predictors represent site-specific variables (e.g., precipitation, vadose zone thickness, well depth, etc.) that may impact the response variable (referred to as [NO_3^-] hereafter). For each tree, a random bootstrapped sample is extracted from the dataset (Efron, 1979), as well as random subset of predictors to consider fitting at each split. Thus, each tree is grown from a bootstrap sample and random subset of predictors,

making trees random and grown independent of the others. Observations not used as bootstrap samples are termed out-of-bag data (OOB). When building a tree, all [NO_3^-] from the bootstrap sample are categorized into terminal nodes, such that each node is averaged and yields a predicted [NO_3^-]. The performance and mean squared error (MSE) of a Random Forest model is evaluated by comparing the observed [NO_3^-] of the OOB data to the average predicted [NO_3^-] from the forest.

3.2.3. Random Forest Application

Observations of [NO_3^-] and predictors were compiled into a single dataset. A Random Forest model was fit to these data using five-fold cross validation (Hastie et al., 2009). The full dataset was randomly shuffled and split into five-folds, where four-folds were used to build the model (training data), and one-fold was held out (testing data). The maximum and minimum of the [NO_3^-] and each predictor were determined and placed into each fold for training models to eliminate the potential for extrapolation during validation. Each fold was used as training data four times, and testing data once. This process was repeated five times to create a total of 25 models and is a similar approach to a study by Nelson et al. (2018).

The four-folds designated to build the model underwent a nested five-fold cross validation, as specified in the *trainControl* function within *caret* (Classification and Regression Training) package (Kuhn, 2008), and managed the methods used by the *train* function. Tuning parameter *mtry* is the number of variables the model randomly selects at each split and is chosen based upon the optimal number of variables resulting in the best model performance. The nested cross validation was used to optimize parameter *mtry*. Once *mtry* is selected, all training data are used to build the forest. The *train* function,

using the *rf* method, fit models and built the Random Forest, consisting of 500 trees by default.

To evaluate model performance, Nash-Sutcliffe Efficiency (NSE), permutation importance, and partial dependence were quantified. NSE was used to assess model performance (Equation (1)). NSE indicates the degree to which observed and predicted values deviate from a 1:1 line, and ranges from negative infinity to 1.

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{pred}\right)^2}{\sum_{i=1}^{n}\left(Y_i^{obs} - Y^{mean}\right)^2}\right] \quad (1)$$

where n is the number of observations, Y_i^{obs} is the ith observation ([NO_3^-]), Y_i^{pred} is the ith prediction from the Random Forest model, and Y^{mean} is the mean of observations i through n. Values from negative infinity to 0 suggest the mean of the observed [NO_3^-] serve as a better predictor than the model. When NSE = 0, model predictions are as accurate as the mean observed [NO_3^-]. From 0, larger NSE values indicate a model's predictive ability improves, until NSE = 1, where observations and predictions are equivalent, and a 1:1 line is achieved. NSE was calculated for both the training and testing data.

OOB data from the training dataset may be used to evaluate both permutation importance, referred to in the rest of this text as variable importance, and partial dependence. Variable importance uses percent increase in mean squared error (%$_{inc}$MSE) to describe predictive power of each predictor in the model. During this process, a single predictor is permuted, or shuffled in the dataset. Therefore, each observed [NO_3^-] has the same relationship between itself and all predictors, except one permuted variable. The

$\%_{inc}$MSE of a variable is determined by comparing the permuted OOB MSE to unpermuted OOB MSE. Important predictors will result in a large MSE increase, while permuting a variable of minor importance does little to impact a model's performance, resulting in only a slight variation in MSE. Partial dependence curves serve as a graphical representation of the relationship between [NO_3^-] and predictors. Each plot considers the significance of other variables, since predictions of [NO_3^-] are influenced by several predictors when building each tree. In these models, the y-axis of a partial dependence plot represents the average of the OOB predicted [NO_3^-] at a specific x-value of each predictor.

3.2.4. Variables and Project Setup

The data from 15 predictors were collected and analyzed (Table 3.1). Spatial variables were manipulated using ArcGIS 10.4. The nitrate dataset for the entire NPNRD had 10,676 observations from 1979 to 2014, and was downloaded from the Quality-Assessed Agrichemical Contaminant Database for Nebraska Groundwater (NEDNR, 2016). Spatial locations for each well were included in the original nitrate dataset, and imported into GIS. Wells were clipped to the Dutch Flats model area, resulting in 2,829 nitrate observations from 214 wells. In order to have an accurate vadose zone thickness, only wells with a corresponding depth to groundwater record, of which the most recent record was used, were selected (2,651 observations from 172 wells). Over this period, several wells were sampled more frequently than others (e.g., monthly sampling, over a short period of record), especially during a USGS National Water-Quality Assessment (NAQWA) study from 1995 to 1999. In order to prevent those wells from dominating the training and testing of the model, annual median [NO_3^-] was calculated for each well and

used in the dataset. The dataset was further manipulated such that each median nitrate observation had 15 complementary predictors (Table 3.1). After incorporating all data, including limited records of dissolved oxygen (DO), the final dataset was 1,049 [NO_3^-] observations from 162 wells sampled between 1993 and 2013.

Table 3.1. List of the 15 predictors used for Random Forest evaluation.

Predictor	Predictor Type	Source
Center Pivot Irrigated Area	Dynamic	NAIP; NAPP; Landsat-1,5, 7, 8;*
Interstate Canal Discharge	Dynamic	USBR (2018)
Area of Planted Corn	Dynamic	NASS (2018)
Precipitation	Dynamic	NOAA (2017)
Available Water Capacity	Static	NRCS (2018)
Dissolved Oxygen	Static	C. Hudson, Personal Communication (2018)
Distance from a Major Canal	Static	USGS (2012);*
Distance from a Minor Canal	Static	USGS (2012);*
Bottom Screen	Static	NEDNR (2016)
Saturated Hydraulic Conductivity	Static	NRCS (2018)
Saturated Thickness	Static	T. Preston, Personal Communication (2017);*
Saturated Zone Travel Distance	Static	NEDNR (2016),*
Surface Elevation (DEM)	Static	NEDNR (1997)
Total Travel Time	Static	NEDNR (2016);*
Vadose Zone Thickness	Static	T. Preston, Personal Communication (2017); A. Young, Personal Communication (2016)

*Signifies data requiring further analysis to yield calculated values.

While Random Forest can be used as a predictive model (i.e., to predict historical/future [NO_3^-] or aquifer vulnerability), this study used it to evaluate variables driving [NO_3^-] in Dutch Flats groundwater. For instance, electrical conductivity has been found as a strong predictor of [NO_3^-] (Rodriguez-Galiano et al., 2014); however, conductivity does not inherently increase or decrease [NO_3^-]. Thus, variables such as this were neglected from the model. Predictors were divided into two categories; static and

dynamic (Table 3.1). Static predictors are those that either do not change over the period of record, or annual records were limited. While DO, for example, would generally experience slight annual variations, data was not available to assign each nitrate sample a unique value. Further, nitrate isotopic composition and DO collected in the 1990s and Wells et al. (2018) studies did not indicate any major changes to biogeochemical processes over nearly two decades. Total travel time was considered a static predictor in this dataset, and was used to link nitrate sampling year to a dynamic predictor value.

Dynamic predictors were defined in this study as data that change temporally. Therefore, each annual median [NO_3^-] was assigned a lagged dynamic value to represent the difference between the time of a particular surface activity (e.g., timing of a particular irrigation practice) and the time when groundwater sampling occurred. Dynamic predictors were available from 1946 to 2013 and included annual precipitation, Interstate Canal discharge, area under center pivot sprinklers, and area of planted corn (Figure 3.2). Dynamic predictors were included to assess their ability to optimize Random Forest groundwater modeling and determine an appropriate lag time. Lag times were based on the vertical travel distance through both the vadose and saturated zones. Area of planted corn was included as a proxy for fertilizer data, which was unavailable prior to 1987. However, analysis suggests there has been decreasing fertilizer application per planted hectare, while area of planted corn has increased in recent decades (Wells et al., 2018). There was a likely trade-off in using this proxy; we were able to extend the period of record back to 1946, allowing for analysis of a wider range of lag times in the model, but might have sacrificed some accuracy in recent decades when nitrogen management may have improved. Wells et al. (2018), also interested in how changes to water resources

management impacted Dutch Flats groundwater quantity, calculated recharge rates. Though mean recharge rate decreased from 1998 to 2016, it was determined their differences were not significantly different. Variable analysis and rationale for their inclusion in this model may be referenced in Appendix C.

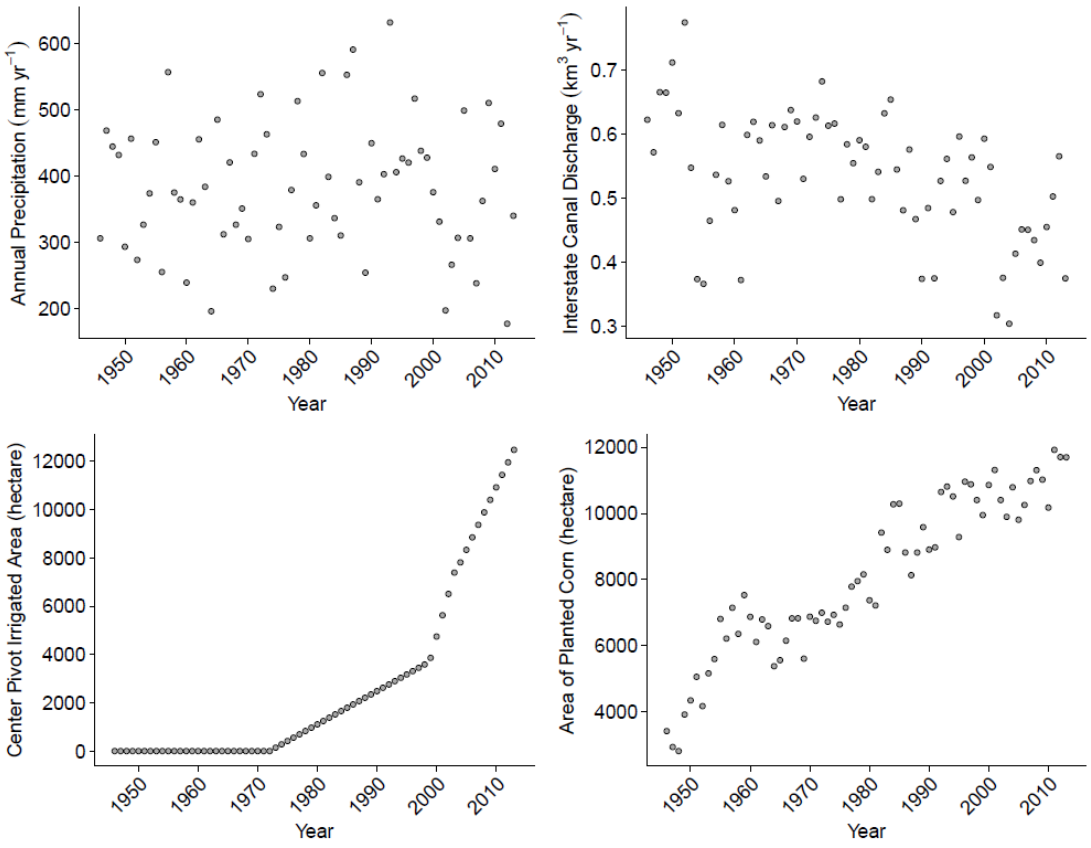


Figure 3.2. Time series plots of all four dynamic predictors from 1946 to 2013

3.2.5. Vadose and Saturated Zone Analysis

A range of transport rates through both the vadose and saturated zones were estimated from $^3H/^3He$ age-dating samples collected in the Dutch Flats area in both 1998 (Verstraeten et al., 2001b) and 2016 (Wells et al., 2018). Due to the influence of canals on both intermediate and deep samples (Böhlke et al., 2007), only shallow well recharge rates were used to estimate vadose zone travel time. The mean ($\bar{x}$ = 0.38 m/yr) and standard deviation (σ = ±0.23 m/yr) of both the 1998 (n=7) and 2016 (n=2) shallow

recharge rates were calculated. Using $\bar{x} \pm 1\sigma$, a range of recharge rates from 0.15 to 0.61 m/yr was converted to a transport rate (t_r) using Equation 2.

$$t_r = \frac{R}{\theta} \tag{2}$$

where *R* is the upper and lower range of recharge rates and θ for this case is the vadose zone mobile water content. Throughout the text, vadose zone transport rates will be abbreviated as t_{rv}, while saturated zone transport rates will be t_{rs}. A mobile water content of 0.13 was previously calibrated using a vertical transport model in the Dutch Flats area (Liao et al., 2012). Slightly expanding the range of vadose zone transport rate values, a range of 1.0 to 4.75 m/yr was used in the model. Vertical transport rates in the vadose zone were stepped by 0.25 m/yr from 1.0 to 4.75 m/yr, ultimately including 16 vadose zone transport rates to evaluate. Equation 3 is generically given below, and was used to calculate the travel time (τ) through both the vadose and saturated zones.

$$\tau = \frac{z}{t_r} \tag{3}$$

where τ is either vadose zone (τ_u) or saturated zone (τ_s) travel time, and z is the vadose zone thickness (z_u) or distance from the water table to well mid-screen (z_s). Though Equations 2 and 3 do not consider horizontal groundwater flow, they are believed to sufficiently model shallow groundwater ages, where groundwater is more likely to follow a linear age gradient. The use of these simple equations are also suggested to sufficiently estimate groundwater ages in wedge-shaped aquifers (Cook and Böhlke, 2000), where Böhlke et al. (2007) found a linear model adequately fit their data in the Dutch Flats area. Nonetheless, saturated zone transport rates and travel times calculated from Equations 2 and 3 should be considered “apparent” vertical transport rates.

Mean ($\bar{x}$ = 0.84 m/yr) and standard deviation (σ = ±0.73 m/yr) of all shallow, intermediate, and deep well recharge rates were included in identifying a range of saturated zone recharge rates from 0.10 to 1.57 m/yr. A total of 35 and 8 recharge rates were used from the 1990s and Wells et al. (2018) studies, respectively. Equation 2 was used to calculate saturated zone transport rates, where this time θ was porosity and set to 0.35 (Böhlke et al., 2007). From 0.25 to 4.5 m/yr, saturated zone transport rates were again stepped by 0.25 m/yr, thus including 18 unique saturated zone transport rates to analyze.

A travel time of τ_u and τ_s was calculated for each well based on z_u and z_s, respectively. A unique total travel time τ_t, was then calculated for each well, with respect to the selected transport rate:

$$\tau_t = \tau_u + \tau_s \quad (4)$$

The total travel time calculated above was used to lag dynamic predictors relative to the nitrate sample date. For instance, a nitrate sample collected in 2010 with a 20-year total travel time (e.g., τ_u = 10 yrs and τ_s = 10 yrs) would be assigned the 1990 values for precipitation (450 mm), Interstate Canal discharge (0.4 km^3/yr), center pivot irrigated area (2484 hectares), and area of planted corn (8905 hectares).

A total of 288 unique transport rate combinations (corresponding to different combinations of the 16 vadose and 18 saturated zone transport rates) joined into a dataset of 302,112 observations was evaluated to determine the optimal rate resulting in the maximum testing NSE from the model. To decrease runtime, Random Forest models were parallel processed through a Holland Computing Center (HCC) cluster at the University of Nebraska-Lincoln (Appendix D). With high uncertainty and small

differences in NSE values, ten transport rate combinations (the five top performing models, plus four transport rate combinations of high and low transport rates, and one intermediate transport rate combination) were then selected out for further evaluation of variable importance and partial dependence.

3.3. Results and Discussion:

This research addressed a relatively unexplored use of Random Forest, which was to identify optimal lag time based on varying transport rate combinations through the vadose and saturated zones. Random Forest was also used in a more traditional approach to relate the influence of several environmental variables on groundwater [NO_3^-], where results were compared and contrasted to previous hydrogeologic studies in the Dutch Flats area. Results and discussion are organized into four subsections, as follows:

I. *Optimizing Transport Rate Based on Nash-Sutcliff Efficiencies (NSE)*
    - The first approach to identify optimal lag times was to run Random Forest using a wide range of potential transport rate combinations (288 total combinations of vadose and saturated zone transport rates) and test whether certain transport rate combination(s) would have substantially higher NSE. It was theorized that if an optimal lag time was identified, Random Forest performance would improve as it could more accurately model the relationship between dynamic predictors and [NO_3^-].

II. *Evaluating Variable Importance of Ten Transport Rate Combinations Further Analyzed*
    - Based on ten selected transport rate combinations (i.e., results from ten Random Forest models) we evaluated the influence of dynamic predictors

in the models, and the potential value of dynamic predictors for identifying optimal transport rates, or lag times.

III. *Optimizing Transport Rate Based on the Relative Importance of the Total Travel Time Variable*
    - Results from the previous section found one predictor, total travel time, had wide-ranging outcomes for the ten selected transport rate combinations. Concluding dynamic predictors were ineffective in this model and using only static predictors, we evaluated the capability of using total travel time to select an optimal transport rate from the 288 vadose and saturated zone transport rate combinations.

IV. *Hydrogeologic Analysis of Model*
    - From former comprehensive site investigations, hydrogeologic conditions have been well defined in the Dutch Flats area. Utilizing variable importance and partial dependence plots, we compared modeled results to findings from the previous studies.

3.3.1. Optimizing Transport Rate Based on Nash-Sutcliff Efficiencies (NSE)

Nash-Sutcliffe Efficiency was used to assess model performance. With a five-fold cross validation, repeated five times, 25 models were built for each of the 288 transport rate combinations. The median training and testing NSE for the 25 models of each transport rate combination was calculated, with testing NSE used to quantify the top transport rate combination. Testing NSE was used in this analysis because compared to the NSE from training data, it is more rigorous and better represents model performance when applied broadly across the study area. Median training NSE ranged from 0.82 to

0.91, while testing NSE ranged from 0.54 to 0.62 (Figure 3.3). The transport rate combination with the highest NSE had a vadose zone transport rate of 4 m/yr and saturated zone transport rate of 0.5 m/yr. The top five performing models are presented in Table 3.2. Generally, vadose zone transport rates for the top performing models were on the upper end of the 1 m/yr to 4.75 m/yr input range, with a mean of 3.7 m/yr. The top performing saturated zone transport rates varied between the 0.25 m/yr to 4.5 m/yr limit. As displayed by Figure 3.3, maximum and minimum testing NSE values were dispersed, with no obvious pattern separating certain ranges of transport rate combinations from others.

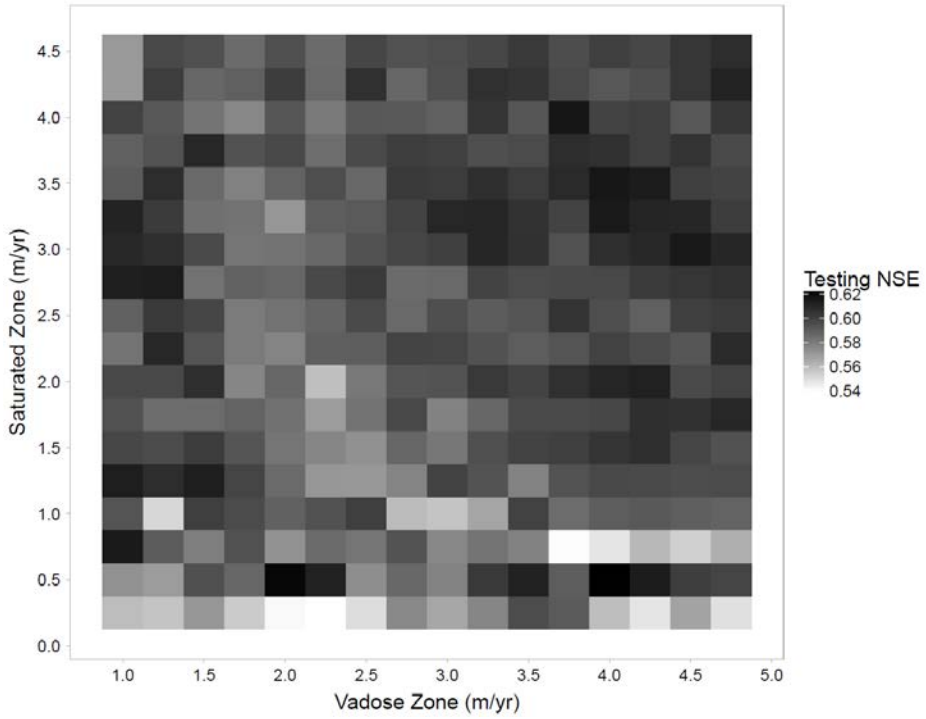


Figure 3.3. Heat map of testing NSE results from 288 vadose and saturated zone transport rate combinations. Testing NSE in this figure is the median of all 25 model outputs from each of the 288 transport rate combinations.

Median total travel times were also determined for each dataset. Three of the top performing transport rate combinations identified very short total travel times (Table 3.2). With a fairly small range in NSE between the highest and lowest performing models, it is

difficult to say with certainty that this approach offered conclusive results regarding the use of Random Forest to determine optimal groundwater lag time. Nonetheless, the top five performing transport rate combinations based on testing NSE were selected for further analysis of variable importance and partial dependence. Due to the variability of transport rate combination results, five additional transport rate combinations were evaluated; four focused on evaluating the range extremes, and a mid-range of both the vadose and saturated zone transport rates (Table 3.2).

Table 3.2. Summary of the ten vadose and saturated zone transport rates evaluated and sorted by testing NSE. Testing NSE values are from the top performing transport rates and range extremes, including range midpoints.

	Vadose Zone Transport Rate (m/yr)	Sat. Zone Transport Rate (m/yr)	Test NSE	[NO_3^-] Observations	Total Travel Time (yrs)**	
					Mean (±1σ)	Median
Five Top-Performing Transport Rates	4	0.5	0.623	878	19.9 (± 15.8)	11.3
	2	0.5	0.622	861	21.6 (± 15.0)	16.5
	3.75	4	0.617	1049	6 (± 3.7)	5.4
	4	3.5	0.617	1049	6.3 (±4.1)	5.7
	4.5	3	0.616	1049	6.7 (± 4.7)	5.7
Evaluating Range Extremes	4.75	4.5	0.608	1049	5.1 (± 3.2)	4.6
	2.75	2.25	0.599	1049	9.6 (± 6.3)	8.5
	1	4.5	0.570	1049	12.6 (± 7.7)	10.8
	1	0.25	0.559	607	26.7 (± 13.3)	20.6
	4.75	0.25	0.548	664	21.3 (± 15.0)	14.9

Variation in NSE amongst the 288 runs could be attributed to the number of observations available for the model to evaluate, rather than dynamic predictors improving the model. Due to data limitations prior to 1946, observations with lag times indicating a surface activity year before 1946 would be excluded from the model. Wells with a large vadose zone and/or depth to mid-screen (often associated with lower [NO_3^-]) and slow transport rates would be excluded from the dataset. For instance, transport rates at the lower end of the range (e.g., t_{rv}= 1 m/yr, t_{rs}= 0.25 m/yr) have 607 observations. High transport rates (e.g., t_{rv}= 4.75 m/yr, t_{rs}= 4.5 m/yr) have all 1,049 observations.

3.3.2. Evaluating Variable Importance of Ten Transport Rate Combinations Further Analyzed

Variable importance, expressed as $\%_{inc}$MSE, evaluates the individual relationship between a predictor and the response variable. Predictors with the largest $\%_{inc}$MSE are suggested by the model to have the largest influence on Dutch Flats groundwater [NO_3^-]. Figure 3.4 plots variable importance for the ten transport rate combinations shown in Table 3.2. Negative values for $\%_{inc}$MSE signify variables where [NO_3^-] were better predicted after randomly permuting the variable. For the ten transport rate combinations evaluated, surface elevation had the highest median $\%_{inc}$MSE (27.5%), with this predictor perhaps associated with locations of agricultural activity throughout the Dutch Flats. Precipitation was indicated as the least important predictor, with a median $\%_{inc}$MSE of -1.5% (Figure 3.4). Overall, results indicate that selecting the top performing transport rate combinations did not have a large influence on variable importance, where, with exception to total travel time, variable importance did not heavily fluctuate with respect to its associated transport rate combination (Figure 3.4).

Analysis of variable importance suggests Random Forest had difficulty identifying dynamic data as effective predictors. In each of the ten transport rate combinations evaluated, the four dynamic predictors consistently ranked among the four lowest predictors, and there is a distinct cutoff between the static and dynamic predictors. While all four dynamic predictors represent processes that influence groundwater [NO_3^-] differently, precipitation, for example, likely has a much larger influence on [NO_3^-] than supported by the model.

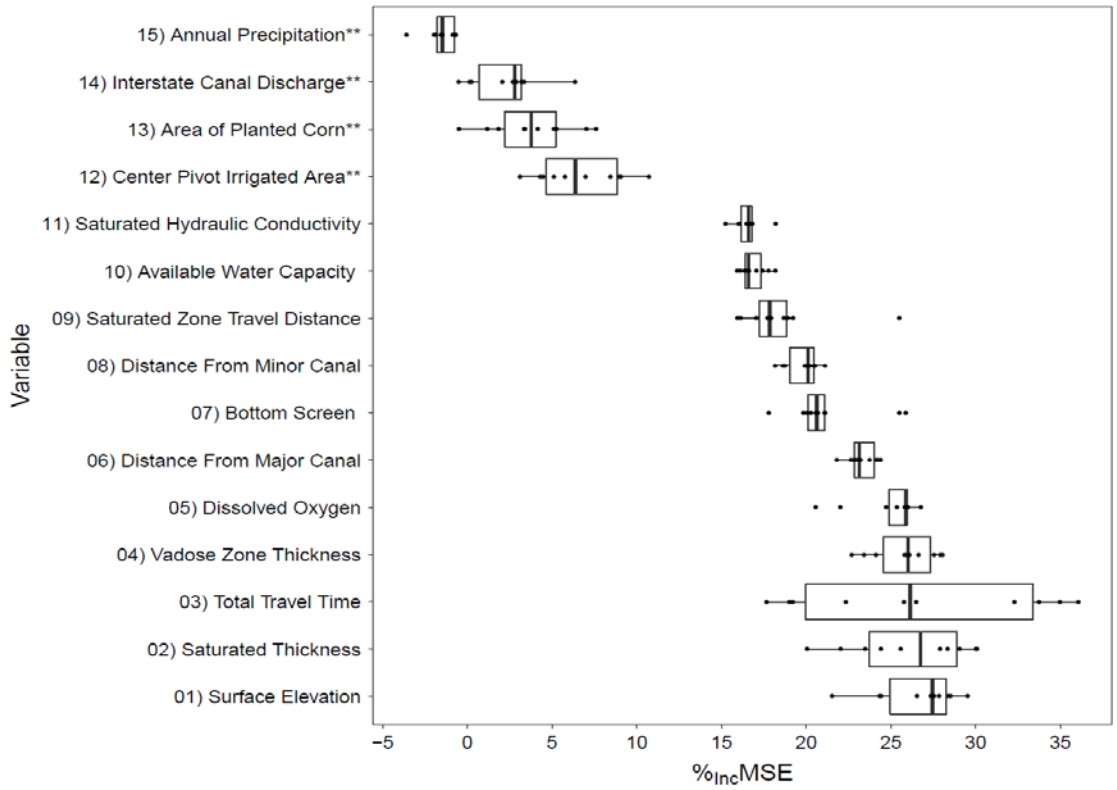


Figure 3.4. Boxplot of the $\%_{Inc}$MSE from the ten transport rate combinations further analyzed for variable importance. Each boxplot has ten points representing the median $\%_{Inc}$MSE from the 25 models (five-fold cross validation, repeated 5 times) created for each transport rate combination. **Denotes dynamic predictors

It is also recognized that groundwater [NO_3^-] may represent a mixture of groundwater ages and may be difficult to connect back to a single year event. Time series with large scatter in the data likely further inhibit Random Forest from adequately evaluating these predictors. For instance, as displayed by Figure 3.2, precipitation and Interstate Canal discharge have large annual variability, which could limit Random Forest from picking out an ideal lag time. Variable importance for precipitation further suggests this concept, where randomly assigning a precipitation value to a nitrate observation yields similar results. Interestingly, center pivot irrigated area had the largest $\%_{inc}$MSE of all dynamic predictors, where this variable also had the most distinct time series pattern. Thus, it could be suggested from the results that in order for Random Forest to properly

utilize dynamic predictors in modeling groundwater [NO_3^-], dynamic predictors require fairly low scatter, unique patterns, over an ideally large range of values. Perhaps with a stronger dynamic predictor, Random Forest could have better results evaluating these variables, where intuitively, refined predictors would improve model performance. For example, detailed fertilizer application data, if available, would be a beneficial predictor to improve conclusions related to the inability of Random Forest to model the relationship between dynamic predictors and [NO_3^-]. Further, Random Forest may not have the capabilities for evaluating dynamically changing predictors in groundwater nitrate, at least not at the spatial resolution used in this analysis. As Random Forest evaluates predictors, annual changes to a variable may impact individual wells differently. Due to a lack of greater spatial resolution available for dynamic predictors, values in the dataset were only at a regional resolution. If increased spatial resolution was available, such that individual wells could be linked to more specific dynamic values over time, it is possible they could be adequately applied to Random Forest modeling of groundwater [NO_3^-].

Thus, results from this analysis suggest that (1) the dynamic data included in the model did little to improve performance, and (2) Random Forest was not able to properly relate these four dynamic predictors to [NO_3^-] such that a realistic transport rate combination could identify a reasonable lag time. It is likely the influence of these dynamic predictors are dampened as nitrate is transported from the surface to wells such that data-driven approaches may find it difficult to sort through noise to identify relationships.

3.3.3. Optimizing Transport Rate Based On the Relative Importance of the Total Travel Time Variable

It was noted from the results of the ten transport rate combinations displayed in Figure 3.4 that the largest variation in $\%_{inc}$MSE occurred in total travel time, with a difference of 18.4% between the upper and lower values. While this variable did not fall under the dynamic data category, each well does have a unique total travel time dependent on the transport rate combination, vadose zone thickness, and distance traveled through the saturated zone at each well. It is believed total travel time and [NO_3^-] are directly related, in that shorter travel times are linked to high [NO_3^-], and long travel times to low [NO_3^-]. This trend was also observed in the analysis of apparent groundwater age and [NO_3^-] by other Dutch Flats studies (Böhlke et al., 2007, Wells et al., 2018). For further analysis, all four dynamic predictors were removed, with total travel time used as the variable to assess lag time. With 11 predictors, 288 runs were performed over the previous ranges of vadose and saturated transport rates. However, instead of using testing NSE to identify optimal transport rates, the largest $\%_{inc}$MSE in total travel time was used. Total travel time was used because of both the known relationship between this variable and [NO_3^-], as well as the variation in total travel time $\%_{inc}$MSE determined in Figure 3.4.

The second analysis without dynamic predictors had a median training NSE of 0.70 for each transport rate combination, and testing NSE values range from 0.59 to 0.60. Less variation in NSE from the second analysis likely reflects a lack in annual variability with the removal of dynamic predictors. Utilizing variable importance to evaluate transport rate combinations, $\%_{inc}$MSE of total travel time ranged from 20.6 to 31.5%. The

largest $\%_{inc}$MSE was associated with vadose and saturated zone transport rates of 3.5 m/yr and 3.75 m/yr, respectively. Figure 3.5 displays the modeled relationship between transport rates and total travel time, where larger transport rates appear to indicate total travel time with a higher variable importance. Variation between transport rate combination and largest $\%_{inc}$MSE could be attributed to inconsistencies in recharge throughout the region. For example, areas of slower transport rates could represent recharge further from canals, whereas fast transport rates are indicative of high inputs from artificial recharge.

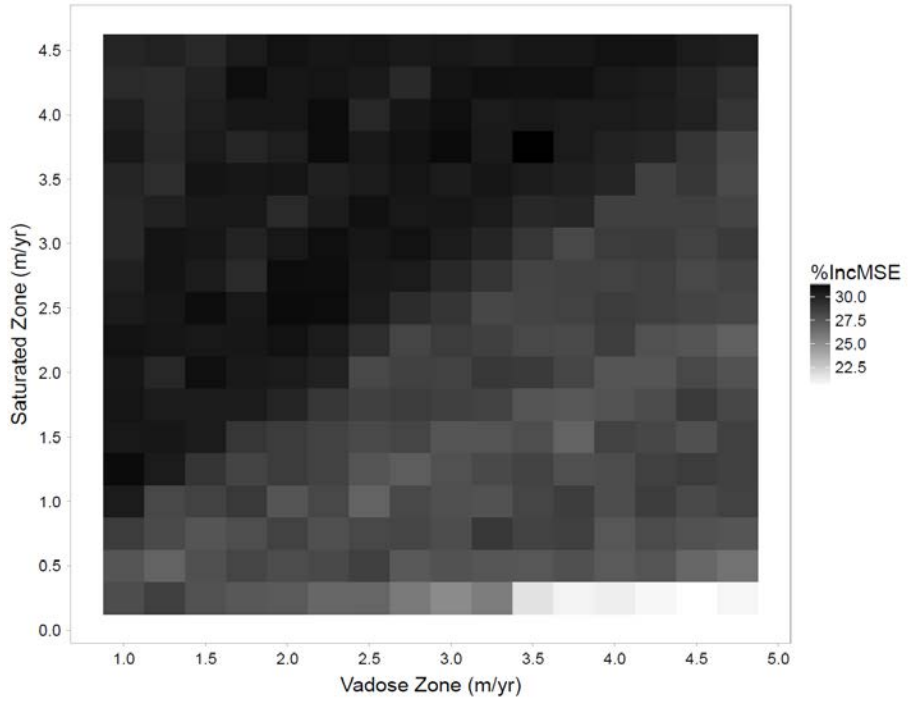


Figure 3.5. Heat map of $\%_{inc}$MSE (median from 25 models) from variable importance of total travel time for each of the 288 transport rate combinations evaluated.

Figure 3.6 shows variable importance plots from the transport rate combination of 3.5 m/yr (vadose zone) and 3.75 m/yr (saturated zone). Within this transport rate combination, total travel time was identified as the most important predictor (median

%incMSE = 31.5%). The least important predictor was available water capacity (median %incMSE = 20.6%). Converting the vadose (3.5 m/yr) and saturated zone (3.75 m/yr) transport rates to a recharge rate yielded values of 0.46 m/yr and 1.31 m/yr, respectively. Both rates compare favorably with recharge rates calculated in the area from the previous 1990s and 2016 Dutch Flats studies. A mean recharge rate of 0.38 m/yr (n = 9) for shallow wells (believed to more accurately represent vadose zone conditions) was determined from the 1998 and 2016 calculated recharge rates. However, when only looking at rates from just 1998, shallow wells had a mean recharge rate of 0.45 m/yr (n = 7).

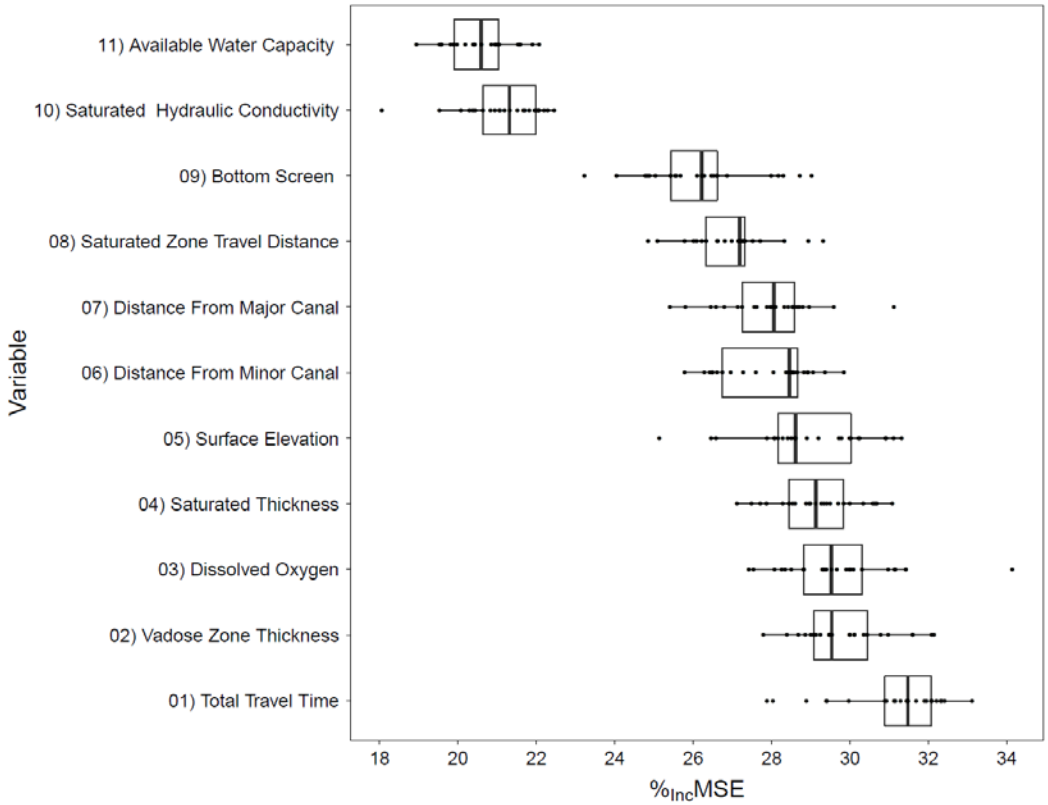


Figure 3.6. Plot from secondary analysis exploring variable importance of the transport rate combination with the largest median %incMSE in total travel time (t_{r_v}= 3.5 m/yr; t_{r_s}= 3.75 m/yr). Each point is from one of 25 Random Forest models run for this evaluation.

The overall mean recharge rates from the 1990s and Wells et al. (2018) studies, including shallow, intermediate, and deep wells, was 0.84 m/yr. While 1.31 m/yr appears high for the saturated zone, it is not unreasonable and is in the 80th percentile of all 43 Dutch Flats area calculated recharge rates. Additionally, intermediate wells were estimated using an exponential equation in the 1990s and Wells et al. (2018) studies, where a mean recharge rate of 1.27 m/yr was calculated (n = 13). Looking at this recharge rate, Random Forest, again, appears to have selected a value consistent with calculated recharge rates from the previous studies. The results from this follow up analysis suggest that even in a complex system such as the Dutch Flats, Random Forest was able to identify reasonable transport rates for both vadose and saturated zones, and it is possible this method offers an inexpensive (i.e., compared to groundwater age-dating across a large monitoring network) and accurate technique for estimating lag time from historical monitoring data.

3.3.4. Hydrogeologic Analysis of Model

From a hydrogeological perspective, Random Forest appears to function well based on what is known about the Dutch Flats area, with exception of the poor performance of dynamic predictors. Within this section, partial dependence plots, supplemented by further analysis of variable importance plots, were utilized to evaluate the model outputs. Partial dependence plots assess the impact a single predictor has on [NO_3^-] in the model, with respect to other predictors in the model. Plots display a range over which changes in predictors and [NO_3^-] are related, as shown by changes in slope, and where they are unrelated. Lines on each plot represent the median predicted [NO_3^-] from all 25 models for each predictor value. Within each model, 50 partial dependence

values are calculated for every predictor. The ribbon around the lines is the maximum and minimum predicted [NO_3^-] at that specific value. Tick marks at the base of each plot are the predictor observations used to train models. Partial dependence plots were created for each of the ten transport rate combinations further analyzed (Table 3.2), and are provided in Appendix E. Histograms for individual predictors from each of the ten transport rate combinations and the total travel time analysis, as well as predicted versus observed plots for testing and training data, may be referenced in Appendix F and Appendix G, respectively. Random Forest was able to predict [NO_3^-] fairly well, though appeared to struggle with concentrations greater than approximately 20 mg N L^{-1}. Overall, each plot from all ten analyses were similar (excluding total travel time), and differences are likely the result of having a differing number of observations in each run. After removing the four dynamic predictors for the second analysis, plots were again comparable, though slightly less variability in maximum and minimum predictions were observed. The partial dependence plot from the secondary transport rate analysis with the largest total travel time $\%_{inc}$MSE (t_{rv}= 3.5 m/yr; t_{rs}= 3.75 m/yr) is provided in Figure 3.7 below.

The inability for Random Forest to evaluate dynamic predictors was first suggested by variable importance plots (Figure 3.4), and further displayed in the partial dependence plots (Appendix E). Over the range of each dynamic predictor, only minor changes in slope were observed, though background knowledge in hydrogeologic processes suggest changes in slope should be observed in some of these plots. For instance, an increase in precipitation should result in increased [NO_3^-], yet the partial dependence plots showed a nearly flat line over the range of explanatory values.

Figures 3.5 and 3.6 were further analyzed to compare Random Forest modeling to what is known about the hydrogeological relationships in the Dutch Flats (Böhlke et al., 2007, Wells et al., 2018), and assess the ability of Random Forest to be applied in complex environments. Overall, trends observed in partial dependence plots are fairly consistent with findings from the 1990s and Wells et al. (2018) studies. An octothorpe within parentheses represents the variable importance rank from Figure 3.6. Predictors such as DO (#3), depth to bottom screen (#9) and saturated zone travel distance (#8) have partial dependence plots similar to what is expected in groundwater settings. The second most important predictor was vadose zone thickness, displaying a slightly unexpected pattern, though at vadose zone depths greater than 8 m, the plot is closer to what is expected for groundwater [NO_3^-]. While it is possible that descriptive features used to evaluate soil characteristics (i.e., available water capacity (#11) and saturated hydraulic conductivity (#10)) have little influence on Dutch Flats groundwater [NO_3^-], a lack of both spatial variability and resolution could limit the ability of Random Forest to adequately evaluate these variables.

A number of predictors demonstrated the influence of leaking canals in the region, consistent with findings from Böhlke et al. (2007). Distance from both major (#7) and minor canals (#6) demonstrated the tendency for [NO_3^-] to rise with increasing distance from canals, such that leaking canals are suggested to dilute [NO_3^-]. The influence from canals is possibly reflected in other predictors, such as surface elevation (#5). Generally, the plot shows that increasing surface elevation corresponds to decreasing [NO_3^-], with a large dip occurring around 1275 m. The pattern could possibly be explained through further exploration of canal surface elevations. For example, the

mean Interstate Canal surface elevation is 1285 m (± 2.6 m) in Dutch Flats, which occurs near the large dip in groundwater [NO_3^-] of the partial dependence plot. Beyond 1285 m, only three wells are located at a greater elevation where the flat line is observed. Saturated Thickness (#4) could represent a similar relationship, where the aquifer tends to be deepest near the North Platte River, and decrease in thickness moving north and south of the river. The top ranked predictor, total travel time, reflects groundwater [NO_3^-] patterns previously discussed by Böhlke et al. (2007) and Wells et al. (2018), where [NO_3^-] decrease with increasing travel time. The plot also shows a distinct break, where groundwater [NO_3^-] are greater over a total travel time range from 0 to 7 years, and considerably lower from total travel times greater than 7 years.

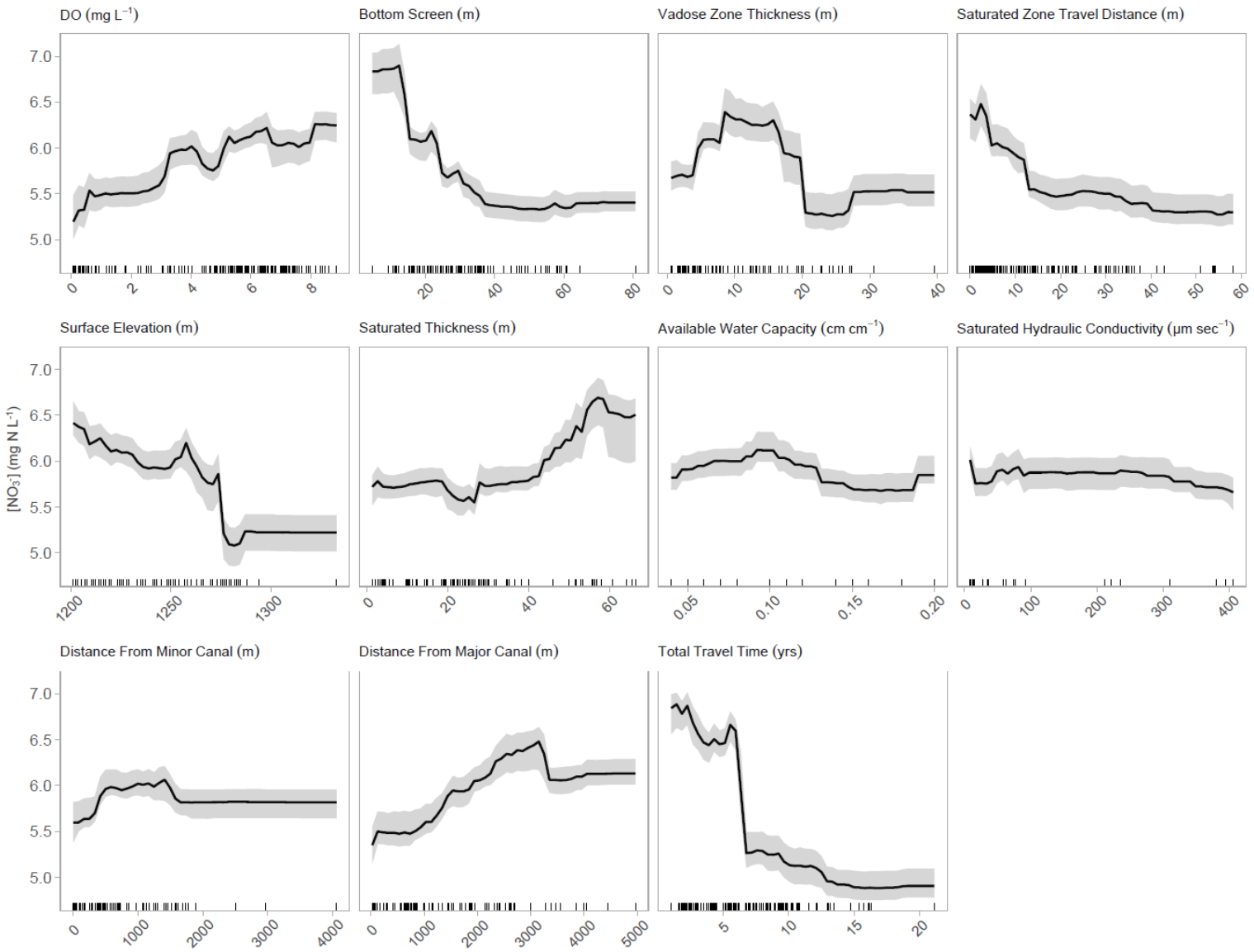


Figure 3.7. Partial dependence plot from analysis of transport rates with the largest median %$_{inc}$MSE of total travel time (t_{rv} = 3.5 m/yr; t_{rs}= 3.75 m/yr). Tick marks on each plot represent predictor observations used to train models.

3.4. Conclusions

The Dutch Flats area consists of large variations in [NO_3^-] throughout a relatively small region in western Nebraska. With numerous variables likely influencing groundwater [NO_3^-], trends were further evaluated with Random Forest, a powerful machine-learning data analysis tool. A long-term dataset of [NO_3^-] provided for an opportune setting to apply this data mining approach. More specifically, Random Forest was used to (1) estimate groundwater transport rates and lag time, (2) assess variables with the largest influence on Dutch Flats groundwater [NO_3^-], and (3) evaluate results compared to previous field surveys in the region.

The first objective involved coupling a dataset consisting of eleven static and four dynamic predictors, where Random Forest NSE values from 288 transport rate combinations failed to estimate a reasonable vadose and saturated zone transport rate. While Random Forest struggled to model dynamic predictors, applying a similar practice in alternative locations could prove useful in further investigating this technique, especially with stronger dynamic predictors whose relationship with nitrate is not dampened during vadose and saturated zone transport. With a secondary dataset of only the static predictors, an alternative approach to estimate transport rates showed potential for success by assessing variable importance of the total travel time predictor. This analysis identified a vadose and saturated zone transport rate combination consistent with rates estimated by $^3H/^3He$ age-dating in the 1990s and 2016 study.

For the second objective, Random Forest was used to rank variables from weakest to strongest predictors of Dutch Flats groundwater [NO_3^-]. Random Forest identified total

travel time as the most important predictor, followed by vadose zone thickness and dissolved oxygen, respectively.

Partial dependence plots, supplemented by variable importance, were used in the third objective to compare Random Forest results to previous field investigations. Results were largely consistent with previous findings, where canal influence, for example, displayed patterns similar to those identified in the 1990s and Wells et al. (2018) Dutch Flats studies. Findings show Random Forest has promise as an alternative approach for water managers to evaluate hydrogeologic settings. Further, the study exemplifies the ability for using Random Forest as a supplementary analysis to large-scale physical models to recognize important variables to incorporate.

The Dutch Flats area has high spatial and temporal variations in water inputs, making this an interesting area to test a relatively novel approach with respect to the application of Random Forest in a complex system. The overall success of this study suggests Random Forest has the capability to both identify reasonable transport rates (and lag time), and to a certain degree, accurately identify variables influencing groundwater [NO_3^-]. Even so, limitations were identified when using dynamic predictors to model groundwater [NO_3^-] and future studies may need to further assess the proper conditions for their application in data driven analyses.

References

Anning, D. W.; Paul, A. P.; McKinney, T. S.; Huntington, J. M.; Bexfield, L. M.; Thiros, S. A. (2012) Predicted nitrate and arsenic concentrations in basin-fill aquifers of the Southwestern United States (Report No. 5065). Scientific Investigations Report, 78; United States Geological Survey.

Babcock, H. M.; Visher, F. N.; Durum, W. H. (1951) Ground-water conditions in the Dutch Flats area, Scotts Bluff and Sioux Counties, Nebraska, with a section on chemical quality of the ground water (Report No. 126). Circular, 51.

Ball, L. B.; Kress, W. H.; Steele, G. V.; Cannia, J. C.; Andersen, M. J. (2006) Determination of canal leakage potential using continuous resistivity profiling techniques, Interstate and Tri-State Canals, western Nebraska and eastern Wyoming, 2004 (Report No. 2006–5032). Scientific Investigations Report, 59; United States Geological Survey.

Baudron, P.; Alonso-Sarría, F.; García-Aróstegui, J. L.; Cánovas-García, F.; Martínez-Vicente, D.; Moreno-Brotóns, J. (2013) Identifying the origin of groundwater samples in a multi-layer aquifer system with Random Forest classification. *Journal of Hydrology*, **499**, 303–315.

Böhlke, J. K.; Verstraeten, I. M.; Kraemer, T. F. (2007) Effects of surface-water irrigation on sources, fluxes, and residence times of water, nitrate, and uranium in an alluvial aquifer. *Applied Geochemistry*, **22** (1), 152–174.

Breiman, L. (1984) *Classification and regression trees*; New York, N.Y.: Chapman & Hall/CRC.

Cook, P. G.; Böhlke, J.-K. (2000) Determining Timescales for Groundwater Flow and Solute Transport. In: *Environmental Tracers in Subsurface Hydrology* (P. G. Cook & A. L. Herczeg, eds.), 1–30; Boston, MA: Springer US.

Efron, B. (1979) Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, **7** (1), 1–26.

Gerber, C.; Purtschert, R.; Hunkeler, D.; Hug, R.; Sültenfuss, J. (2018) Using environmental tracers to determine the relative importance of travel times in the unsaturated and saturated zones for the delay of nitrate reduction measures. *Journal of Hydrology*, **561**, 250–266.

Grimm, R.; Behrens, T.; Märker, M.; Elsenbeer, H. (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. *Geoderma*, **146** (1–2), 102–113.

Harvey, F. E.; Sibray, S. S. (2001) Delineating Ground Water Recharge from Leaking Irrigation Canals Using Water Chemistry and Isotopes. *Ground Water*, **39** (3), 408–421.

Hastie, T.; Tibshirani, R.; Friedman, J. H. (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics, 2nd ed.; New York, NY: Springer.

Hobza, C. M.; Andersen, M. J. (2010) Quantifying canal leakage rates using a mass-balance approach and heat-based hydraulic conductivity estimates in selected irrigation canals, western Nebraska, 2007 through 2009 (Report No. 2010–5226). Scientific Investigations Report, 38; United States Geological Survey.

Homer, C. G.; Dewitz, J.; Yang, L.; Jin, S.; Danielson, P.; Xian, G. Z.; Coulston, J.; et al. (2015) Completion of the 2011 National Land Cover Database for the conterminous United States – Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, **81**, 345354.

Hudson, C. (North Platte Natural Resources District, Scottsbluff, NE, USA). (2018) Personal Communication with M.J. Wells (University of Nebraska, Lincoln, NE, USA).

Irmak, S.; Odhiambo, L.; Kranz, W. L.; Eisenhauer, D. E. (2011) Irrigation Efficiency And Uniformity, And Crop Water Use Efficiency (Extension Circular No. EC732); Lincoln, NE: University of Nebraska – Lincoln. Retrieved from Available at http://extensionpubs.unl.edu/

Jones, Z. M.; Linder, F. J. (2015) Exploratory Data Analysis using Random Forests. *Proceedings of the 73rd Annual Mpsa Conference*.

Kuhn, M. (2008) Building Predictive Models in *R* Using the **caret** Package. *Journal of Statistical Software*, **28** (5).

Liao, L.; Green, C. T.; Bekins, B. A.; Böhlke, J. K. (2012) Factors controlling nitrate fluxes in groundwater in agricultural areas: Factors controlling NO_3^- fluxes in groundwater. *Water Resources Research*, **48** (6).

Luckey, R. R.; Cannia, J. C. (2006) Groundwater Flow Model of the Western Model Unit of the Nebraska Cooperative Hydrology Study (COHYST) Area, 63; Lincoln, NE: Nebraska Department of Natural Resources. Retrieved from ftp://ftp.dnr.nebraska.gov/Pub/cohystftp/cohyst/model_reports/WMU_Documentation_060519.pdf

Maupin, M. A.; Kenny, J. F.; Hutson, S. S.; Lovelace, J. K.; Barber, N. L.; Linsey, K. S. (2014) *Estimated use of water in the United States in 2010*. Circular; Reston, Virginia: United States Geological Survey.

NASS. (2018) USDA/NASS QuickStats Ad-hoc Query Tool; National Agricultural Statistics Service, United States Department of Agriculture, Retrieved from https://quickstats.nass.usda.gov/

NEDNR. (1997) 7.5 Digital Elevation Models. *Elevation Data*; Nebraska Department of Natural Resources, Retrieved from https://dnr.nebraska.gov/data/elevation-data

NEDNR. (2009) Fifty-fifth biennial report of the Department of Natural Resources, 675; Lincoln, NE: Nebraska Department of Natural Resources. Retrieved from https://dnr.nebraska.gov/sites/dnr.nebraska.gov/files/doc/surface-water/biennial-reports/BiennialReport2005-06.pdf

NEDNR. (2016) Quality-Assessed Agrichemical Contaminant Database for Nebraska Ground Water; Nebraska Department of Natural Resources. Retrieved from https://clearinghouse.nebraska.gov/Clearinghouse.aspx

Nelson, N. G.; Muñoz-Carpena, R.; Phlips, E. J.; Kaplan, D.; Sucsy, P.; Hendrickson, J. (2018) Revealing Biotic and Abiotic Controls of Harmful Algal Blooms in a Shallow Subtropical Lake through Statistical Machine Learning. *Environmental Science & Technology*, **52** (6), 3527–3535.

NOAA. (2017) Data Tools | Climate Data Online (CDO) | National Climatic Data Center (NCDC). Retrieved from https://www.ncdc.noaa.gov/cdo-web/datatools

Nolan, B. T.; Fienen, M. N.; Lorenz, D. L. (2015) A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. *Journal of Hydrology*, **531**, 902–911.

Nolan, B. T.; Gronberg, J. M.; Faunt, C. C.; Eberts, S. M.; Belitz, K. (2014) Modeling Nitrate at Domestic and Public-Supply Well Depths in the Central Valley, California. *Environmental Science & Technology*, **48** (10), 5643–5651.

NRCS. (2018) Web Soil Survey; Natural Resources Conservation Service, United States Department of Agriculture. Retrieved from https://websoilsurvey.sc.egov.usda.gov/

Ouedraogo, I.; Defourny, P.; Vanclooster, M. (2017) Validating a continental-scale groundwater diffuse pollution model using regional datasets. *Environmental Science and Pollution Research*.

Peters, J.; Baets, B. D.; Verhoest, N. E. C.; Samson, R.; Degroeve, S.; Becker, P. D.; Huybrechts, W. (2007) Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling*, **207** (2–4), 304–318.

Preston, T. (North Platte Natural Resources District, Scottsbluff, NE, USA). (2017) Personal Communication with M.J. Wells (University of Nebraska, Lincoln, NE, USA).

Rahmati, O.; Pourghasemi, H. R.; Melesse, A. M. (2016) Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. *CATENA*, **137**, 360–372.

Ransom, K. M.; Nolan, B. T.; A. Traum, J.; Faunt, C. C.; Bell, A. M.; Gronberg, J. A. M.; Wheeler, D. C.; et al. (2017) A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley aquifer, California, USA. *Science of The Total Environment*, **601**–**602**, 1160–1172.

Rodriguez-Galiano, V. F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J. P. (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, **67**, 93–104.

Rodriguez-Galiano, V. F.; Mendes, M. P.; Garcia-Soldado, M. J.; Chica-Olmo, M.; Ribeiro, L. (2014) Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). *Science of The Total Environment*, **476**–**477**, 189–206.

USBR. (2018) Hydromet: Archive Data Access; United States Bureau of Reclamation. Retrieved from https://www.usbr.gov/gp/hydromet/hydromet_arcread.html

USGS. (2012) NHDPlus High Resolution; United States Geological Survey. Retrieved from https://nhd.usgs.gov/NHDPlus_HR.html

Verstraeten, I.M.; Sibray, S. S.; Cannia, J. C.; Tanner, D. Q. (1995) Reconnaissance of ground-water quality in the North Platte Natural Resources District, western Nebraska, June-July 1991 (Report No. 94–4057). Water-Resources Investigations Report, 114; United States Geological Survey.

Verstraeten, I.M.; Steele, G. V.; Cannia, J. C.; Hitch, D. E.; Scripter, K. G.; Böhlke, J. K.; Kraemer, T. F.; et al. (2001a) Interaction of surface water and ground water in the Dutch Flats area, western Nebraska, 1995-99 (Report No. 2001–4070). Water-Resources Investigations Report, 56; United States Geological Survey.

Verstraeten, I.M.; Steele, G. V.; Cannia, J. C.; Böhlke, J. K.; Kraemer, T. E.; Hitch, D. E.; Wilson, K. E.; et al. (2001b) Selected field and analytical methods and analytical results in the Dutch Flats area, western Nebraska, 1995-99 (Report No. 2000–413). Open-File Report, 53; Reston, VA: United States Geological Survey.

Wells, M.; Gilmore, T.; Mittelstet, A.; Snow, D.; Sibray, S. (2018) Assessing Decadal Trends of a Nitrate-Contaminated Shallow Aquifer in Western Nebraska Using Groundwater Isotopes, Age-Dating, and Monitoring. *Water*, **10** (8), 1047.

Wheeler, D. C.; Nolan, B. T.; Flory, A. R.; DellaValle, C. T.; Ward, M. H. (2015) Modeling groundwater nitrate concentrations in private wells in Iowa. *Science of The Total Environment*, **536**, 481–488.

Wiesmeier, M.; Barthold, F.; Blank, B.; Kögel-Knabner, I. (2011) Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant and Soil*, **340** (1–2), 7–24.

Young, A. (University of Nebraska, Lincoln, NE, USA). (2016) Personal Communication with M.J. Wells (University of Nebraska, Lincoln, NE, USA).

APPENDICES

Appendix A – Results from iNoble Version 2.2 workbook developed by the International Atomic Energy Agency (IAEA) to model groundwater age from 2016 samples.

Model	Constraints	**Name**	He (cm3STP/g)	Erorr	Ne (cm3STP/g)	Erorr	Ar (cm3STP/g)	Erorr
CE (Aeschbach)	T, Ex-A, F or R varied	**1C-D**	6.1E-08	1.2E-09	2.0E-07	4.0E-09	0.00035	7.1E-06
CE (Aeschbach)	T, Ex-A, F or R varied	**1C-M**	7.5E-08	1.5E-09	3.1E-07	6.1E-09	0.00041	8.3E-06
CE (Aeschbach)	T, Ex-A, F or R varied	**1E-D**	7.3E-08	1.5E-09	2.1E-07	4.1E-09	0.00037	7.3E-06
CE (Aeschbach)	T, Ex-A, F or R varied	**1E-M**	5.1E-08	1.0E-09	2.1E-07	4.2E-09	0.00036	7.2E-06
CE (Aeschbach)	T, Ex-A, F or R varied	**1E-S**	5.3E-08	1.1E-09	2.3E-07	4.5E-09	0.00038	7.6E-06
CE (Aeschbach)	T, Ex-A, F or R varied	**1G-I**	4.8E-08	9.6E-10	1.8E-07	3.7E-09	0.00034	6.8E-06
CE (Aeschbach)	T, Ex-A, F or R varied	**1G-I**	4.6E-08	9.2E-10	1.8E-07	3.6E-09	0.00034	6.9E-06
CE (Aeschbach)	T, Ex-A, F or R varied	**1L-D**	1.2E-07	2.5E-09	1.8E-07	3.5E-09	0.00036	7.3E-06
CE (Aeschbach)	T, Ex-A, F or R varied	**2D-D**	6.2E-08	1.2E-09	2.0E-07	4.1E-09	0.00036	7.1E-06
CE (Aeschbach)	T, Ex-A, F or R varied	**2D-M**	4.6E-08	9.1E-10	1.8E-07	3.7E-09	0.00032	6.4E-06
CE (Aeschbach)	T, Ex-A, F or R varied	**2D-S**	4.8E-08	9.5E-10	2.0E-07	4.1E-09	0.00036	7.2E-06

Name	Kr (cm3STP/g)	Erorr	Xe (cm3STP/g)	Erorr	3He/4He (uncorrected)	Erorr	Altitude (m)	Temp
1C-D	8.3E-08	1.7E-09	1.1E-08	2.2E-10	1.3E-06	2.0E-08	1248.2	18.2
1C-M	9.0E-08	1.8E-09	1.2E-08	2.3E-10	1.4E-06	2.0E-08	1248.2	19.6
1E-D	8.3E-08	1.7E-09	1.1E-08	2.2E-10	4.1E-06	2.0E-08	1240.9	15.2
1E-M	8.5E-08	1.7E-09	1.1E-08	2.2E-10	2.0E-06	2.0E-08	1240.9	15.1
1E-S	8.7E-08	1.7E-09	1.2E-08	2.3E-10	1.5E-06	2.0E-08	1240.9	14.3
1G-I	8.0E-08	1.6E-09	1.1E-08	2.1E-10	1.4E-06	2.0E-08	1219.6	12.7
1G-I	8.1E-08	1.6E-09	1.1E-08	2.2E-10	1.4E-06	2.0E-08	1222.3	14.9
1L-D	9.0E-08	1.8E-09	1.3E-08	2.5E-10	7.7E-07	2.0E-08	1211.9	12.5
2D-D	8.2E-08	1.6E-09	1.0E-08	2.1E-10	1.8E-06	2.0E-08	1247.7	17.3
2D-M	7.4E-08	1.5E-09	9.6E-09	1.9E-10	1.9E-06	2.0E-08	1247.7	17.1
2D-S	8.3E-08	1.7E-09	1.2E-08	2.3E-10	1.5E-06	2.0E-08	1247.7	15.8

Name	Salinity (permil)	Sampling Date	Noble Gas Extraction Date	Tritium (TU)	Tritium Error (TU)	NGT	Error	Excess Air (cc/kg)	Error
1C-D	0	8/17/2016	4/5/2017	5.0	0.3	14.0	0.4	232.2	5.2
1C-M	0	8/17/2016	3/1/2017	5.6	0.3	11.3	0.2	18.5	0.6
1E-D	0	8/16/2016	4/6/2017	7.1	0.5	14.3	0.3	185.4	3.6
1E-M	0	8/16/2016	4/5/2017	6.1	0.3	15.6	0.3	278.1	7.8
1E-S	0	8/16/2016	4/5/2017	7.1	0.4	10.8	0.3	39.1	1.0
1G-I	0	4/18/2016	4/5/2017	7.1	0.2	10.9	0.1	0.8	0.0
1G-I	0	10/12/2016	3/1/2017	6.7	0.4	9.6	0.1	0.4	0.0
1L-D	0	8/17/2016	4/6/2017	6.9	0.2	6.5	0.2	0.0	0.0
2D-D	0	8/16/2016	4/5/2017	6.3	0.2	16.1	0.4	175.2	6.1
2D-M	0	8/16/2016	4/5/2017	5.1	0.3	13.9	0.3	1.4	0.0
2D-S	0	8/16/2016	4/6/2017	6.2	0.2	10.6	0.2	33.5	0.4

Name	Sum(Chi^2)	F values	Terrigenic 4He	Error	[3He]trit (TU)	Error	T-3He age (Years)	Error	3He/4He corrected	Error
1C-D	2.1	0.8	1.4E-08	2.0E-09	4.9	2.3	12.0	4.4	1.3E-06	2E-08
1C-M	2.3	0.3	4.1E-10	1.9E-09	0.5	1.2	1.5	2.9	1.4E-06	2E-08
1E-D	0.6	0.8	2.4E-08	2.0E-09	92.2	2.9	47.0	1.2	4.0E-06	2E-08
1E-M	3.0	0.8	1.6E-09	2.3E-09	13.7	1.9	20.9	1.7	2.0E-06	2E-08
1E-S	1.8	0.7	1.3E-10	1.6E-09	2.2	1.1	4.9	1.9	1.5E-06	2E-08
1G-I	8.0	0.2	4.7E-09	9.6E-10	2.8	1.7	5.9	3.1	1.4E-06	2E-08
1G-I	3.8	0.3	4.1E-09	9.2E-10	2.3	1.5	5.3	3.0	1.4E-06	2E-08
1L-D	3.6	0.0	8.3E-08	2.5E-09	14.5	2.3	20.2	2.0	7.6E-07	2E-08
2D-D	4.4	0.8	1.3E-08	2.2E-09	17.8	2.0	23.9	1.6	1.8E-06	2E-08
2D-M	9.7	0.3	1.5E-09	9.1E-10	11.0	1.3	20.5	1.6	1.9E-06	2E-08
2D-S	0.2	0.8	2.8E-10	1.3E-09	2.2	0.9	5.4	1.7	1.5E-06	2E-08

Appendix B – Chapter 2 Additional Information

Table B1. Nitrate data from current study, North Platte Natural Resources District, and Nebraska Agricultural Contaminant Database. This dataset was used to create comparisons in Table 2.3 and Figure 2.6.

Well ID	1990s $[NO_3^-]$ * (mg N L^{-1})	2008 $[NO_3^-]$ (mg N L^{-1})	2016 $[NO_3^-]$ (mg N L^{-1})
10A-S	1.4	0.0	0.0
10E-S	16.5	5.4	0.0
10K-S	7.6	3.9	4.4
10M-S	9.5	4.5	6.4
10N-S	3.0	2.1	0.0
1C-S	9.9	8.7	9.5
1E-S	13.7	13.4	33.5
1G-S	8.6	9.7	34.5
1H-S	11.3	8.1	13.7
1J-S	16.3	11.4	13.8
1M-S	10.9	10.4	16.1
1N-S	3.3	2.1	4.3
2E-S	4.9	3.2	4.7
2F-S	16.1	0.9	30.5
2J-S	12.4	8.0	8.8
3B-S	2.6	0.3	0.0
3C-S	1.8	0.0	3.8
3F-S	4.8	27.7	0.5
4A-S	0.5	0.5	0.7
5A-S	2.9	2.3	1.4
5B-S	0.9	1.9	2.2
5D-S	11.1	8.8	6.5
5F-S	4.7	4.6	3.5
5G-S	11.4	13.3	19.8
6C-S	4.8	5.6	5.9
6D-S	19.2	20.9	26.4
6E-S	8.9	5.1	6.3
6F-S	3.6	0.7	0.9
6G-S	2.1	1.8	1.7
6H-S	3.6	4.3	3.5
6M-S	4.5	8.1	6.4
6N-S	6.9	9.1	5.7
7A-S	5.3	3.8	3.3
7B-S	6.5	8.7	2.6
7C-S	14.3	14.2	11.2
7D-S	3.3	1.3	5.0
7H-S	14.2	10.6	17.8
8B-S	5.3	4.6	11.3
8C-S	5.6	4.7	5.7
8D-S	5.4	4.7	6.5
8E-S	0.1	0.0	0.0
8G-S	1.4	1.8	5.0
9D-S	5.4	1.2	0.0

9E-S	9.6	8.7	6.4
11D-M	2.7	2.3	2.2
1C-M	3.0	4.8	5.7
1E-M	9.0	7.1	4.6
1G-M	9.6	4.0	7.1
1H-M	4.0	3.1	3.5
1J-M	7.8	9.2	4.0
1M-M	11.9	5.9	12.9
2C-M	2.3	1.6	2.3
2D-M	5.0	1.9	1.3
2F-M	19.8	12.1	21.4
2J-M	9.4	8.9	4.7
2L-M	4.8	4.8	1.4
3C-M	1.1	1.3	0.9
3E-M	7.9	10.0	9.2
3F-M	6.2	12.1	0.0
5B-M	0.9	0.6	0.6
10A-D	3.6	6.5	9.4
10E-D	5.2	5.1	4.1
10K-D	5.0	5.3	6.2
10M-D	9.4	5.9	7.6
11A-D	4.8	3.9	4.2
1C-D	2.4	3.0	4.3
1E-D	2.3	3.6	3.1
1G-D	8.9	4.0	6.2
1L-D	4.9	6.3	1.1
2C-D	2.8	2.4	1.9
2D-D	1.4	1.2	1.4
2F-D	3.1	3.1	3.8
2L-D	1.6	1.3	1.0
3B-D	1.2	0.5	0.3
3C-D	1.2	1.1	1.6
3E-D	3.7	3.7	3.9
3F-D	11.5	5.7	4.7
5B-D	0.9	0.8	0.6
6G-D	6.2	1.3	2.0
6H-D	0.1	0.9	0.6
6M-D	2.4	1.1	1.4
7A-D	1.9	1.7	1.7
7C-D	5.8	9.4	5.7
7D-D	3.7	3.3	4.1
8D-D	5.7	5.4	8.1
9D-D	2.8	3.1	3.1
9E-D	5.5	4.6	3.4

Note: * Value shown is from 1998 or 1999, or the average from the two years.

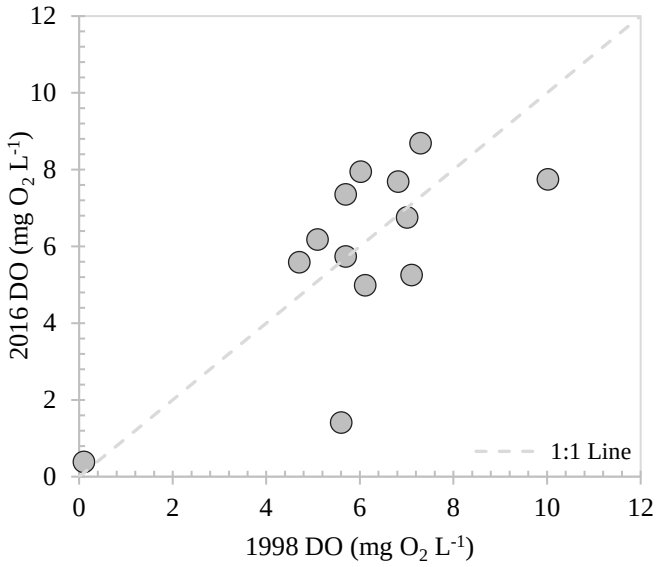


Figure B1. Comparison of groundwater dissolved oxygen (DO) in 2016 to 1998, where DO was mostly similar between both studies.

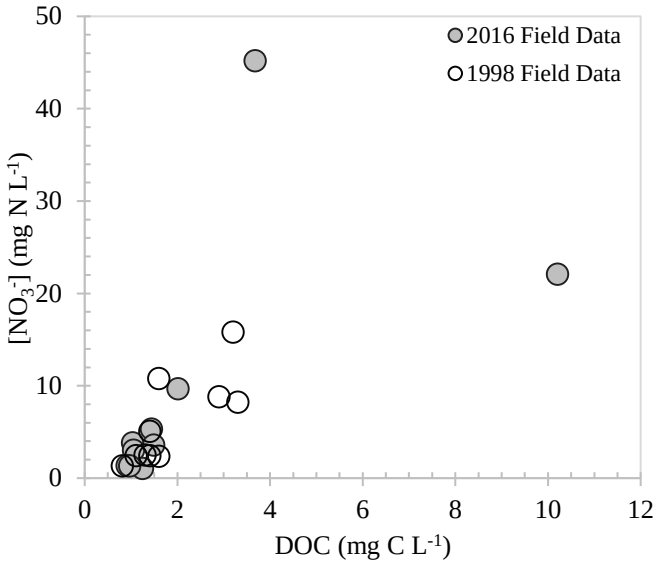


Figure B2. Comparison of [NO_3^-] in 2016 and 1998 to dissolved organic carbon (DOC). Large [NO_3^-] and DOC at Well 1G-S are consistent with isotopes indicating high organic nitrogen source.

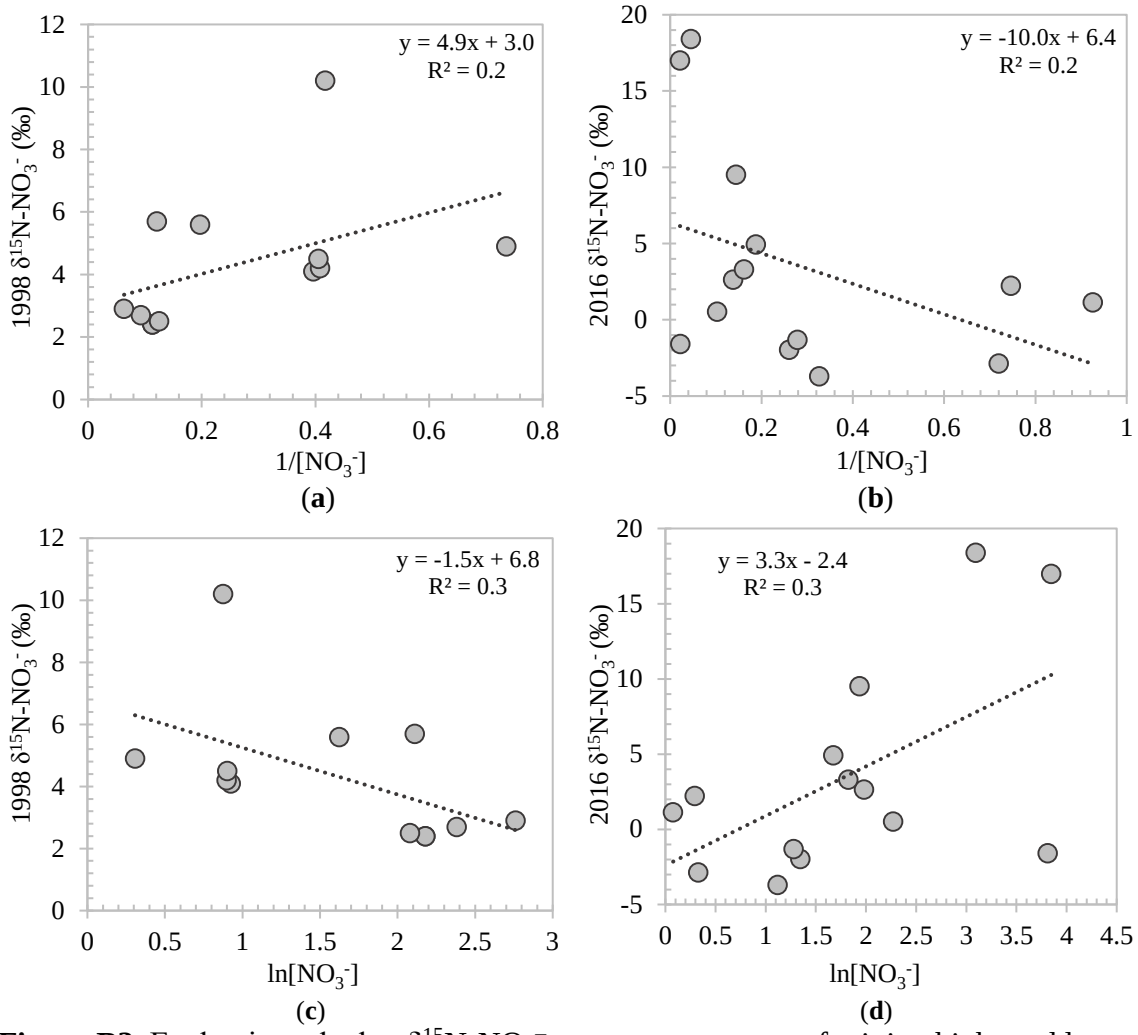


Figure B3. Evaluating whether $\delta^{15}N\text{-}NO_3^-$ suggests processes of mixing high- and low-[NO_3] groundwater in: (**a**) 1998; and (**b**) 2016; or groundwater denitrification in: (**c**) 1998; and (**d**) 2016 [64]. Simple groundwater mixing nor denitrification were indicated from the composition of nitrogen isotopes throughout the Dutch Flats.

Appendix C – Random Forest Variable Analysis

A brief summary of how each variable was evaluated is provided below. Dynamic predictors were downloaded and analyzed from 1946 (first year Interstate Canal discharge was available) to 2013 (last year of nitrate data available).

<u>Center Pivot Irrigated Area:</u>

Aerial imagery was used to digitize center pivot irrigated fields (Figure C1). High resolution imagery was utilized when available (NAIP, NAPP), with LANDSAT providing additional imagery for the analysis. Two years were identified as breaks (1999 and 2003) in the data, and used to interpolate irrigated area before, between, and after the breaks. That is, years before the shift (<1999), during the major shift (1999 – 2003), and years following the major shift (>2003). Linear regression was used to estimate center pivot irrigated area for years that

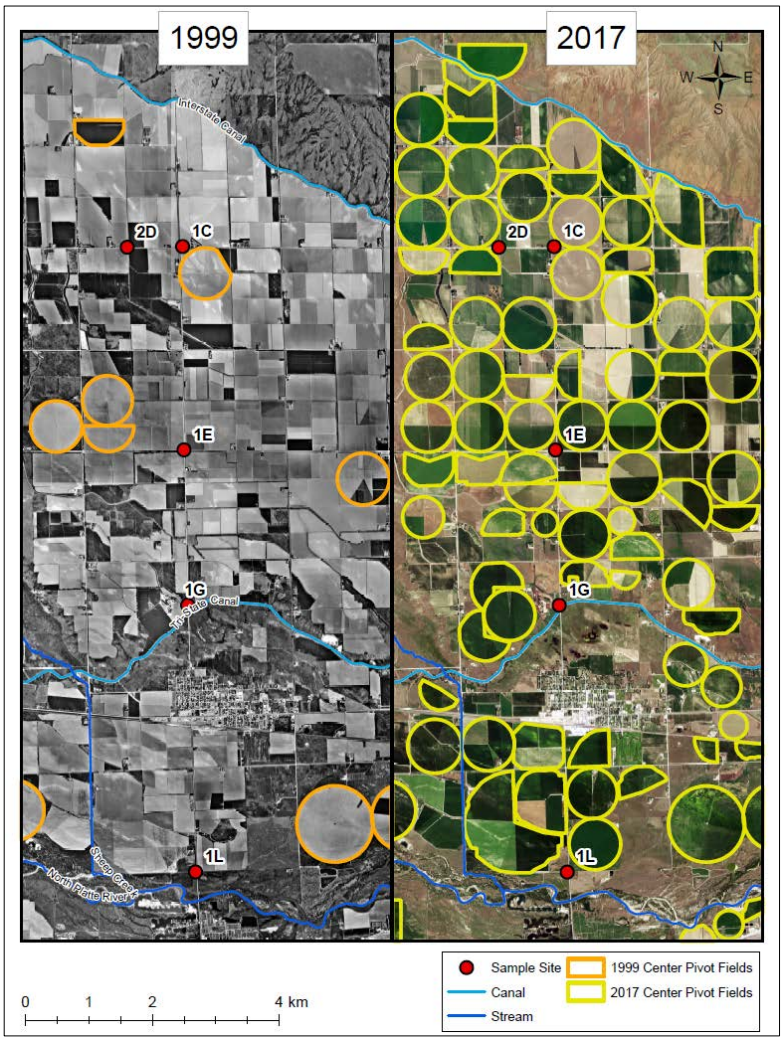
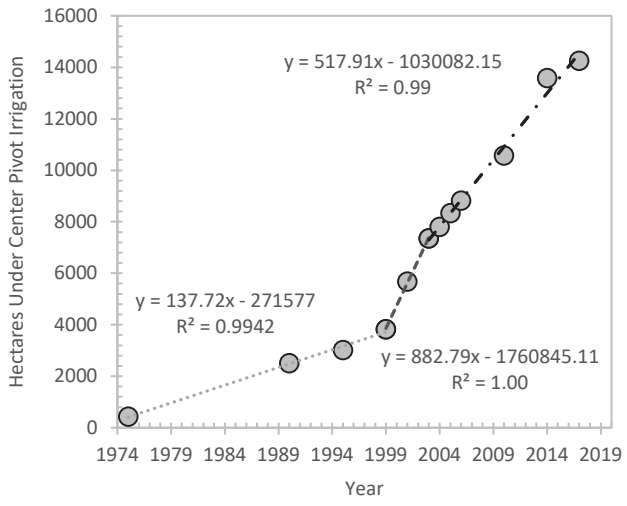


Figure C1: Visual comparison of 1999 center pivot irrigated fields to 2017.

were not digitized, providing a dynamic dataset for the study. Before the shift, center pivot fields were installed at approximately 138 hectares per year. During the largest conversion from furrow to center pivot, fields increased from 1999 – 2003 at a rate of 883 hectares per year, followed by 518 hectares per year thereafter. In 1999 area under center pivot irrigation was approximately 3,830 hectares. As of 2017, irrigated hectares increased by nearly 270%, to 14,253 hectares.

Rationale for Inclusion: Center pivot and furrow irrigated fields are believed to follow an inverse trend. In other words, as center pivot irrigated fields increase over time, the less efficient furrow irrigated fields will decrease. With this, it is believed that less nitrate will leach through the root zone with water application improvements. Thus, this variable was included to see if the model identified it as an important variable, such that lower $[NO_3^-]$ would be linked to larger areas of center pivot, or higher $[NO_3^-]$ with less center pivot irrigated area.



Year	Center Pivot Irrigated Hectares
1975	429
1990	2507
1995	3015
1999*	3830
2001	5685
2003*	7361
2004*	7804
2005*	8341
2006*	8822
2010*	10577
2014*	13591
2017	14253

*Years with available High-Resolution imagery

Interstate Canal Discharge:

Annual Interstate Canal discharge was retrieved from the Bureau of Reclamation's Hydromet data archive (USBR, 2018). Data were downloaded, and records were converted from cm^3/second to km^3/year. Since unique annual values were available for each year over the period of record, this was a dynamic dataset. Datasets were also downloaded for both the Tri-State and Mitchell-Gering Canal, however, only an Interstate Canal discharge value was assigned to each nitrate observation. This approach was justified because it is unknown which canal influences a well, regardless of the distance for a canal. Further, annual canal discharge from the Tri-State and Mitchell-Gering Canals were also compared to the Interstate Canal, in which it was determined canals follow similar annual trends. Each nitrate observation was assigned a lagged canal discharge value dependent on the optimally identified total travel time.

Rationale for Inclusion: Due to the known influence of canal leakage on groundwater [NO_3^-], annual volume of water diverted into the Interstate Canal was used as a dynamic predictor to investigate if the model could identify the influence of annually high or low stream diversions on groundwater [NO_3^-].

Planted Corn Area:

Due to the study area incorporating two counties, planted corn area was analyzed in three steps. First, the total annual planted area was downloaded from the USDA National Statistics Service (NASS) for both Scotts Bluff and Sioux Counties from 1946 to 2015 (NASS, 2018). Second, the USDA-NASS georeferenced Cropland Data Layer was downloaded through Geospatial Data Gateway (https://gdg.sc.egov.usda.gov), and projected into ArcMap 10.4. From 2002 to 2015, estimated corn area was determined for

each year in the two respective counties. The area of planted corn was also determined for each county within the boundary of the Dutch Flats area. Next, a ratio of each year was estimated by comparing the area of planted corn for each county within Dutch Flats, to the planted corn area for the entirety of each county. Ratios were averaged from 2002 to 2015, with the Dutch Flats county-level planted corn area having ratios of 0.18 (± 0.01) and 0.77 (± 0.03) compared to total planted corn in Scotts Bluff and Sioux Counties, respectively. Third, estimated planted corn area within Dutch Flats from 1946 to 2015 was predicted by multiplying each year of total county-level planted corn by the respective ratio and summing up areas.

Rationale for Inclusion: This variable was used as a dynamic predictor, and was included as a proxy for the limited amount of long-term fertilizer data available since corn requires high fertilizer inputs. Statistical analysis of the long-term dataset suggested there has been a significant decrease in fertilizer application in Scotts Bluff County, while planted corn area has significantly increased. However, it is possible there are more uncertainties related to application surveys associated with fertilizer than planted corn area.

Precipitation:

Annual sums for precipitation were downloaded from the National Climatic Data Center of National Oceanic and Atmospheric Administration (NOAA) at the Scottsbluff W.B. Heilig Field Airport, NE US (NOAA, 2017).

Rationale for Inclusion: With several years of consistent data, precipitation was used as a dynamic predictor, and lagged with respect to the sample collection date. Higher years of precipitation could potentially lead to increased leaching or fertilizer runoff.

Available Water Capacity and Saturated Hydraulic Conductivity:

Available water capacity (AWC) and saturated hydraulic conductivity (K) were from Web Soil Survey, which is maintained by the U.S. Department of Agriculture (USDA) Natural Resources Conservation Service. Spatial data was downloaded from Geospatial Data Gateway (https://gdg.sc.egov.usda.gov). AWC is the amount of water that is retained by soil and available for plant uptake. Saturated hydraulic conductivity describes the movement of water through saturated soil. Data was retrieved for both Sioux and Scotts Bluff Counties, and evaluated with the *Soil Data Viewer* tool developed by NRCS. Values of each respective predictor were extracted to each well.

Rationale for Inclusion (AWC): AWC influences water storage, and thus the amount of time that irrigated water will stay within the root zone. Lower values of AWC could influence leaching, as more nitrate may leach below the root zone before plants can assimilate the nutrients.

Rationale for Inclusion (Saturated Hydraulic Conductivity): This study incorporated K to evaluate how the movement of water through saturated soils might impact [NO_3^-] in the Dutch Flats area. Higher K would suggest water moves quicker through the soil, potentially transporting nitrate to groundwater at higher rates.

Bottom Screen & Saturated Zone Travel Distance:

Bottom screen depth was already within the nitrate dataset retrieved from the Quality-Assessed Agrichemical Contaminant Database (NEDNR, 2016). Saturated zone travel distance is the distance below the water table to the screen midpoint. If the screen crossed the water table, common for shallow wells, the saturated zone travel distance was the screen midpoint between the bottom screen and water table. In deeper wells, the

saturated zone travel distance value was found as the midpoint between the upper and lower screens.

Rationale for Inclusion (Lower Screen): Well depth has been found an important variable in other Random Forest studies (Wheeler et al., 2015). This predictor is a factor of both vadose and saturated zone, and evaluates how traveling through both these zones impacts [NO_3^-].

Rationale for Inclusion (Saturated Zone Travel Distance): Evaluates how the distance water travels from the water table to screen midpoint influences groundwater [NO_3^-]. It is believed longer travel distances will have lower [NO_3^-], as there is more time for biological processes to reduce concentrations.

Surface Elevation (DEM), Vadose Zone Thickness, and Saturated Thickness:

To estimate saturated thickness, wells were only selected for the Random Forest model if they had both a [NO_3^-] and depth to groundwater value. Depth to groundwater records for the NPNRD (n = 49,765; 1929 – 2016) were retrieved from the University of Nebraska Conservation and Survey Division (CSD) (A. Young, Personal Communication, 2016). Depth to groundwater records from 2017 (n = 806) were also sent from the NPNRD GIS Coordinator (T. Preston, Personal Communication, 2017). Wells from the nitrate database were joined with the 2017 NPNRD monitoring well data. [NO_3^-] that did not have a 2017 depth to groundwater value were checked with the CSD database for additional depth to groundwater records. Of the 2,829 nitrate samples in the Dutch Flats area, 2,651 samples had a well matching a depth to groundwater value from the 2017 monitoring well and/or CSD depth to groundwater dataset. Under the

assumption wells were in an unconfined aquifer, the most recent depth to groundwater record at a well nest's shallow well was assigned to the entire well nest.

Base of aquifer contours were acquired from NPNRD (T. Preston, Personal Communication, 2017). A 30 meter base of aquifer surface was interpolated using the *Topo to Raster* tool in ArcMap 10.4. A base of aquifer and 30 meter surface Digital Elevation Model (DEM) raster value were extracted to each well point, representing the surface elevation and base of aquifer elevation at each well. Depth to groundwater at each well was used to determine the water table elevation. This in turn was used to estimate saturated thickness of the aquifer by subtracting the interpolated base of aquifer elevation from the water table elevation. One well (8C-S) returned a negative aquifer thickness value. For this specific well, the saturated thickness was estimated by taking bottom of the well and subtracting it from the water table elevation, and assuming the bottom of the well was located at the top of the confining layer. The resulting saturated thickness was consistent with estimated thicknesses in the surrounding area.

Rationale for Inclusion (Surface Elevation (DEM)): Due to varying landscape, surface elevation DEM was used to explore, in part, how moving further from the North Platte River (NPR) influenced groundwater [NO_3^-]. Since the landscape generally slopes toward the river, locations of higher elevation would suggest a well is further from the NPR. This predictor may also assess the impact of canals at their respective elevations.

Rationale for Inclusion (Vadose Zone Thickness): Vadose zone thickness has been previously used to evaluate how a quickly a contaminant may reach the water table, and was assessed by Rodriguez-Galiano et al. (2014). Vadose zone thickness may also play a role in nitrate storage in the unsaturated zone.

Rationale for Inclusion (Saturated Thickness): This variable is related to groundwater age distribution, and is included in the exponential equation (Vogel, 1967).

Distance from major and minor canals:

Canal spatial data was from the USGS National Hydrography Dataset (USGS, 2012). The three largest canals were identified, based on canal discharge, in the Dutch Flats area as the Interstate, Tri-State, and Mitchell-Gering Canals. These canals were used to determine the "Distance from Major Canal" variable. Further analysis was conducted via aerial imagery to digitize lower order irrigation canals that were not in the NHD database. These canals were used to determine the "Distance from minor canal" variable. The *Near* tool in ArcMap 10.4 was used to calculate the distance each well was from the closest major canal and minor canal.

Rationale for Inclusion: Interstate Canal discharge is a temporally dependent variable; however, the distance from major and minor canals was used to evaluate the spatial component of canals in the region. Due to the high leakage potential of these canals, the variable was included to evaluate the influence of a well's proximity to canals in Dutch Flats on groundwater [NO_3^-].

Dissolved Oxygen (DO):

This dataset was sent directly from NPNRD (C. Hudson, Personal Communication, 2018). The dataset had several DO values, however, there was not a perfect 1:1 match between the collection of a nitrate sample, and the collection of DO. Therefore, each well's record of DO was averaged. Thus, each well had a unique value with respect to other wells, but the same value applied to all the nitrate samples within

that well for each annual median [NO_3^-]. Analysis of DO and isotopes of nitrate suggest there has not been a major change in denitrification within Dutch Flats from 1998 to 2016 (Wells et al., 2018).

Rationale for Inclusion: DO largely drives biological processes impacting groundwater [NO_3^-], where it would be expected low DO would be associated with low [NO_3^-], and high DO with higher [NO_3^-].

Total Travel Time:

This variable was largely addressed within the thesis. Please reference sections *3.2.5* and *3.3.3* for further information.

Appendix D – Random Forest script and Holland Computing Center cluster interface

```
########################################
########## Random Forest #################
#### Dutch Flats Nitrate Concentration####
########################################

# Clear workspace & load packages ------------------------------------------
rm(list=ls(all=TRUE))
library(randomForest)
library(caret)
library(usdm)

#Assign file and directory names ------------------------------------------
  filename<-"OptimizeDataset"
  setwd("/lustre/work/gilmoregwlab/mwells5/FinalRF/")

#Read command line arguments (indicates file line index, previously `g`
args = commandArgs(trailingOnly=TRUE)
g <- as.numeric(args[1])

  # Upload data ------------------------------------------
  filein<-paste("/lustre/work/gilmoregwlab/mwells5/FinalRF/",filename,".csv",sep="")
  d<-read.csv(skip=(1049*g), filein ,header= FALSE, nrows=1049)
  dorig<-d
  RemNA<-na.omit(d)

  # Random forests: k-fold cross-validation ----------------------------------------
  ev<-RemNA[,c(8:18)]
  nit<-RemNA[,c(7)]

  dirout<-"/lustre/work/gilmoregwlab/mwells5/FinalRF/"
  d.sub<-data.frame(nit,ev)
  colnames(d.sub)[1]<-"rv"
  rv<-"rv"

  # pre-allocate dataframes
  vi.df<-data.frame(run=NULL,variables=NULL,percMSE=NULL)
  nse.df<-data.frame(fold=NULL,run=NULL,train=NULL,test=NULL)
  r2.df<-data.frame(fold=NULL,run=NULL,r2.train=NULL,r2.test=NULL)

  for (s in 1:5)
  {
  set.seed(777+s)

  # shuffle data
  d.sub = d.sub[sample(1:nrow(d.sub)),]

  # where are maxes and mins of predictors?
  min = apply(d.sub, 2, which.min)
  max = apply(d.sub, 2, which.max)
  intrain = unique(c(min, max))
  intest = (1:nrow(d.sub))[!(1:nrow(d.sub) %in% intrain)]
  d.sub.intrain<-d.sub[intrain,]
  d.sub.intest<-d.sub[intest,]

  for (k in 1:5) { # 5 for 5-fold cross-validation
  # obtain folds
  folds = cut(1:nrow(d.sub.intest), breaks = 5, labels = FALSE) # breaks should equal number of folds
  d.train = rbind(d.sub.intrain, d.sub.intest[folds != k, ])
  d.test = d.sub.intest[folds == k, ]

  # prepare training scheme - 5-fold cross-validation is applied to tune the mtry parameter
  control <- trainControl(method="repeatedcv", number=5, repeats=5, savePredictions = "final")
  # train the model
  model <- train(form = rv~.,
                 data=d.train,
                 method="rf",
                 trControl=control,
                 importance = TRUE)

    # training NSE & R2
  pred = predict(model$finalModel, newdata = d.train)
  NSE.train = 1 - mean((pred - d.train$rv)^2)/var(d.train$rv)#; print(NSE)
  r2.train = 1 - sum((d.train$rv - pred)^2)/sum((d.train$rv - mean(d.train$rv))^2)

  # testing NSE & R2
  pred = predict(model$finalModel, newdata = d.test)
  NSE.test = 1 - mean((pred - d.test$rv)^2)/var(d.test$rv)#; print(NSE)
  r2.test = 1 - sum((d.test$rv - pred)^2)/sum((d.test$rv - mean(d.test$rv))^2)

  #update nse.df
  nse.df<-rbind(nse.df,data.frame(fold=k,run=g,train=NSE.train,test=NSE.test))
  r2.df<-rbind(r2.df,data.frame(fold=k,run=g,train=r2.train,test=r2.test))

  # estimate variable importance
  imp<-importance(model$finalModel)
  imp.df<-data.frame(variables = names(imp[,1]),percMSE = imp[,1],run=g)
  vi.df<-rbind(vi.df,imp.df)

  file=paste(dirout,"OutputFolder/",rv,"_rep",s,"_fold",k,".Rdata",sep="")
  save(model,file=file)
  }

  }
  write.csv(nse.df,file=paste(dirout,"NSE/",g,"_NSE.csv",sep=""),row.names=F)
  write.csv(r2.df,file=paste(dirout,"r2/",g,"_r2.csv",sep=""),row.names=F)
  write.csv(vi.df,file=paste(dirout,"vi/",g,"_vi.csv",sep=""),row.names=F)
```

```
#!/bin/sh
#SBATCH --time=24:00:00
#SBATCH --mem-per-cpu=64gb
#SBATCH --job-name=RFModel
#SBATCH --error=FinalRF.%J.stderr
#SBATCH --output=FinalRF.%J.stdout
#SBATCH --array=0-287

module load R/3.4

Rscript FullCode_modified3.R $SLURM_ARRAY_TASK_ID
```

```
login as: mwells5
Using keyboard-interactive authentication.
Password:
Access denied
Using keyboard-interactive authentication.
Password:
Using keyboard-interactive authentication.
Duo two-factor login for mwells5

Enter a passcode or select one of the following options:

 1. Duo Push to XXX-XXX-5341
 2. Phone call to XXX-XXX-5341
 3. SMS passcodes to XXX-XXX-5341 (next code starts with: 1)

Passcode or option (1-3): 1
Success. Logging you in...
Last login: Wed Jul 25 12:05:59 2018 from 129.93.167.84
      /\                 _____          \/                   /\
          _             /  __ \                                  .>- /\
         /(.._      \/   | /   \/_ __ __ _ _ __   ___           _..)
           \`\\          | |    | '__/ _` | '_ \ / _ \    \/  //"/
            |>    \/     | \__/\ | | (_| | | | |  __/          |
    ________/|_____________\____/_|  \__,_|_| |_|\___|________/|_________

Have questions? See the documentation!
https://hcc-docs.unl.edu/

For SLURM docs, see https://hcc-docs.unl.edu/display/HCCDOC/Submitting+Jobs

HCC currently has no storage that is suitable for HIPAA, PHI, PID, classified
or other data sets for which access is legally restricted.  Users are not
permitted to store such data on HCC machines.

You have no HCC acknowledgment credits currently
For details, please see https://hcc-docs.unl.edu/x/CYRBAQ

                [       /home       ][       /work        ][      /common      ]
U:mwells5------>[0.5% (0.1GB/20GB)  ][0.0% (0.1GB/50TB)   ][0.0% (0GB/1TB)     ]
G:gilmoregwlab->[0.0% (0.1GB/500GB) ][0.0% (0.1GB/50TB)   ][0.0% (0GB/1.9PB)   ]
E:------------->[18.6% (1.9TB/10TB) ][65.7% (954.1TB/1.4PB)][0.2% (4TB/1.9PB)   ]
key: (U)ser, (G)roup, (E)ntire system
above output generated by 'hcc-du' command, type 'hcc-du -h' for more options

Purge policy exists on /work, see https://hcc-docs.unl.edu/x/4BEF
[mwells5@login.crane ~]$ cd /lustre/work/gilmoregwlab/mwells5/FinalRF
[mwells5@login.crane FinalRF]$ ls
caret_6.0-80.tar.gz   FullCode_modified3.R  ggplot2_2.2.1.tar.gz    index.html  OptimizeDataset.csv  pd  randomForest_4.6-14.tar.gz    RFSubmit_modifi
FullCode_modified2.R  FullCode_modified.R   ggplot2_2.2.1.tar.gz.1  NSE         OutputFolder         r2  randomForest_4.6-14.tar.gz.1  vi
[mwells5@login.crane FinalRF]$ module load R/3.4
[mwells5@login.crane FinalRF]$ module list

Currently Loaded Modules:
  1) R/3.4



[mwells5@login.crane FinalRF]$ sbatch RFSubmit_modified
Submitted batch job 15606347
[mwells5@login.crane FinalRF]$ squeue -u mwells5
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
  15606347_[0-287]     batch  RFModel  mwells5 PD       0:00      1 (Priority)
[mwells5@login.crane FinalRF]$ scancel 15606347
[mwells5@login.crane FinalRF]$
```

Appendix E – Partial Dependence Plots

Figure E1: Partial Dependence plot including 607 observations at t_{r_v} = 1.0 m/year and t_{r_s} = 0.25 m/year.

Figure E2: Partial Dependence plot including 1,049 observations at t_{r_v} = 1.0 m/year and t_{r_s} = 4.5 m/year.

Figure E3: Partial Dependence plot including 1,049 observations at t_{r_v}= 2.75 m/year and t_{r_s}= 2.25 m/year.

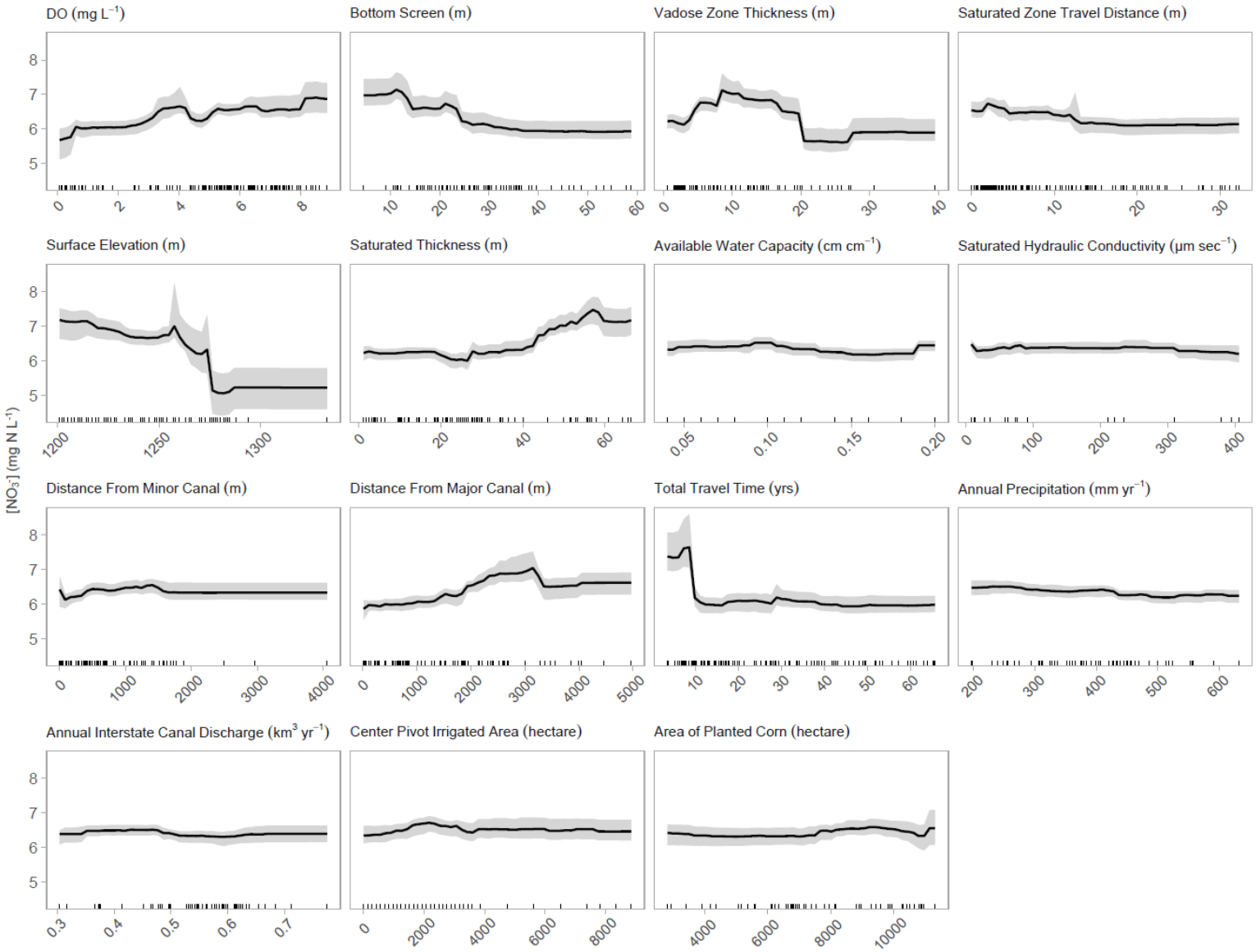


Figure E4: Partial Dependence plot including 861 observations at t_{r_v} = 2 m/year and t_{r_s} = 0.5 m/year.

Figure E5: Partial Dependence plot including 1,049 observations at t_{r_v}= 3.75 m/year and t_{r_s}=4m/year.

Figure E6: Partial Dependence plot including 1,049 observations at t_{r_v}= 4.5 m/year and t_{r_s}= 3 m/year.

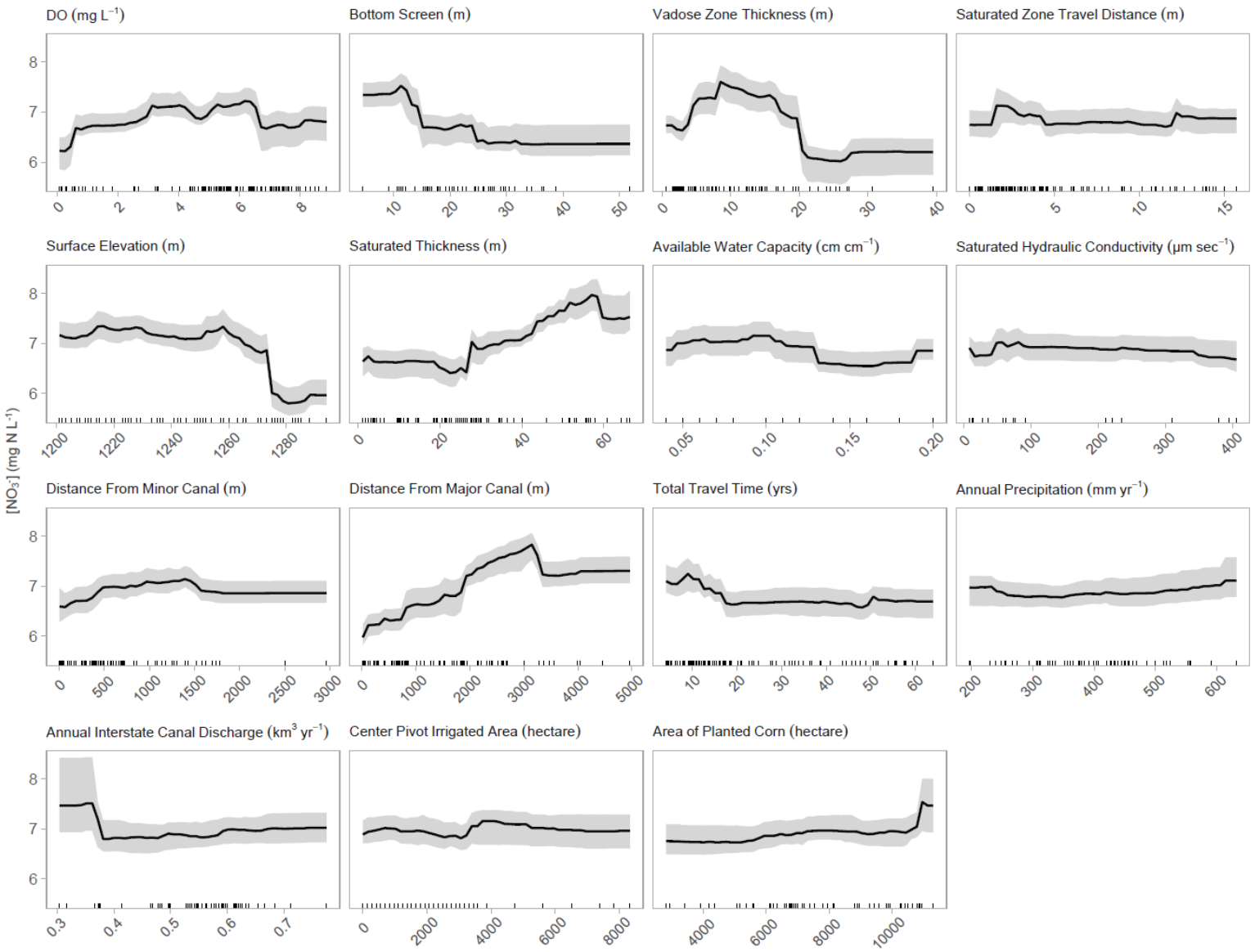


Figure E7: Partial Dependence plot including 664 observations at t_{r_v}= 4.75 m/year and t_{r_s}= 0.25 m/year.

Figure E8: Partial Dependence plot including 1,049 observations at t_{r_v}= 4.75 m/year and t_{r_s}= 4.5 m/year.

Figure E9: Partial Dependence plot including 878 observations at t_{r_v} = 4 m/year and t_{r_s} = 0.5 m/year.

Figure E10: Partial Dependence plot including 1,049 observations at t_{r_v}= 4 m/year and t_{r_s}= 3.5 m/year.

Appendix F – Histograms for individual explanatory variables from each of the ten combinations further evaluated and total travel time analysis

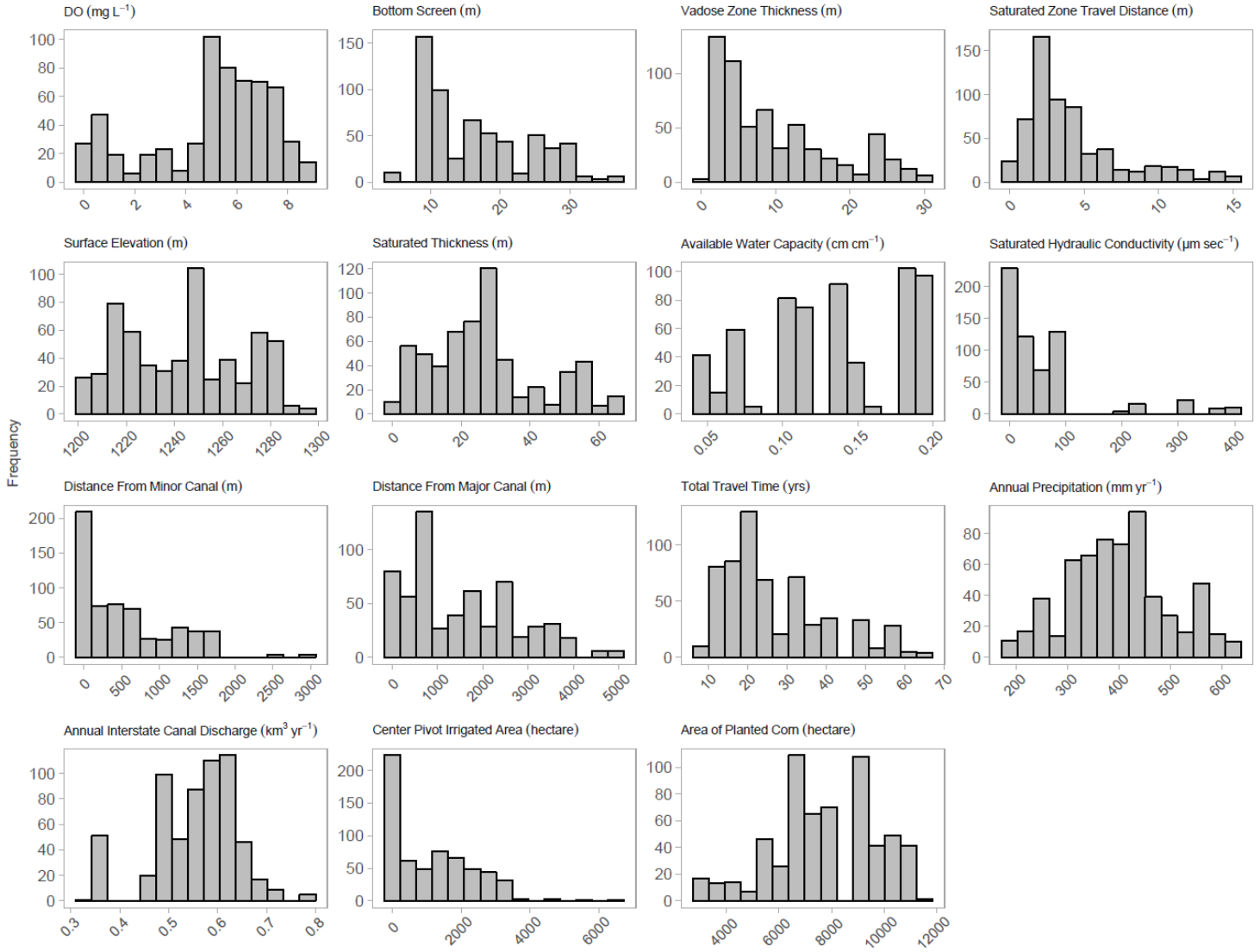


Figure F1: Explanatory variable distribution for 607 observations at t_{r_v} = 1.0 m/year and t_{r_s} = 0.25 m/year.

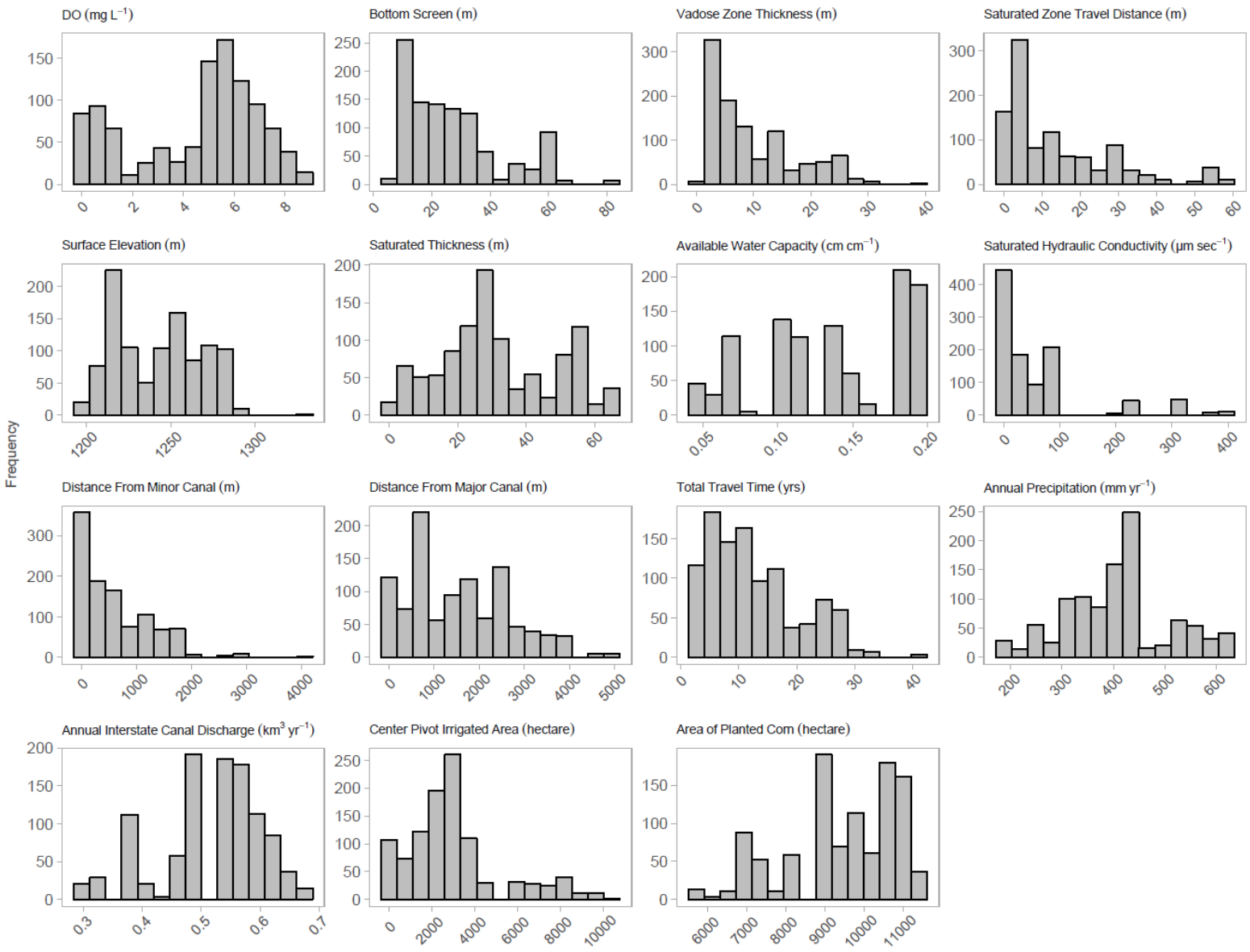


Figure F2: Explanatory variable distribution for 1,049 observations at t_{r_v}= 1.0 m/year and t_{r_s}= 4.5 m/year.

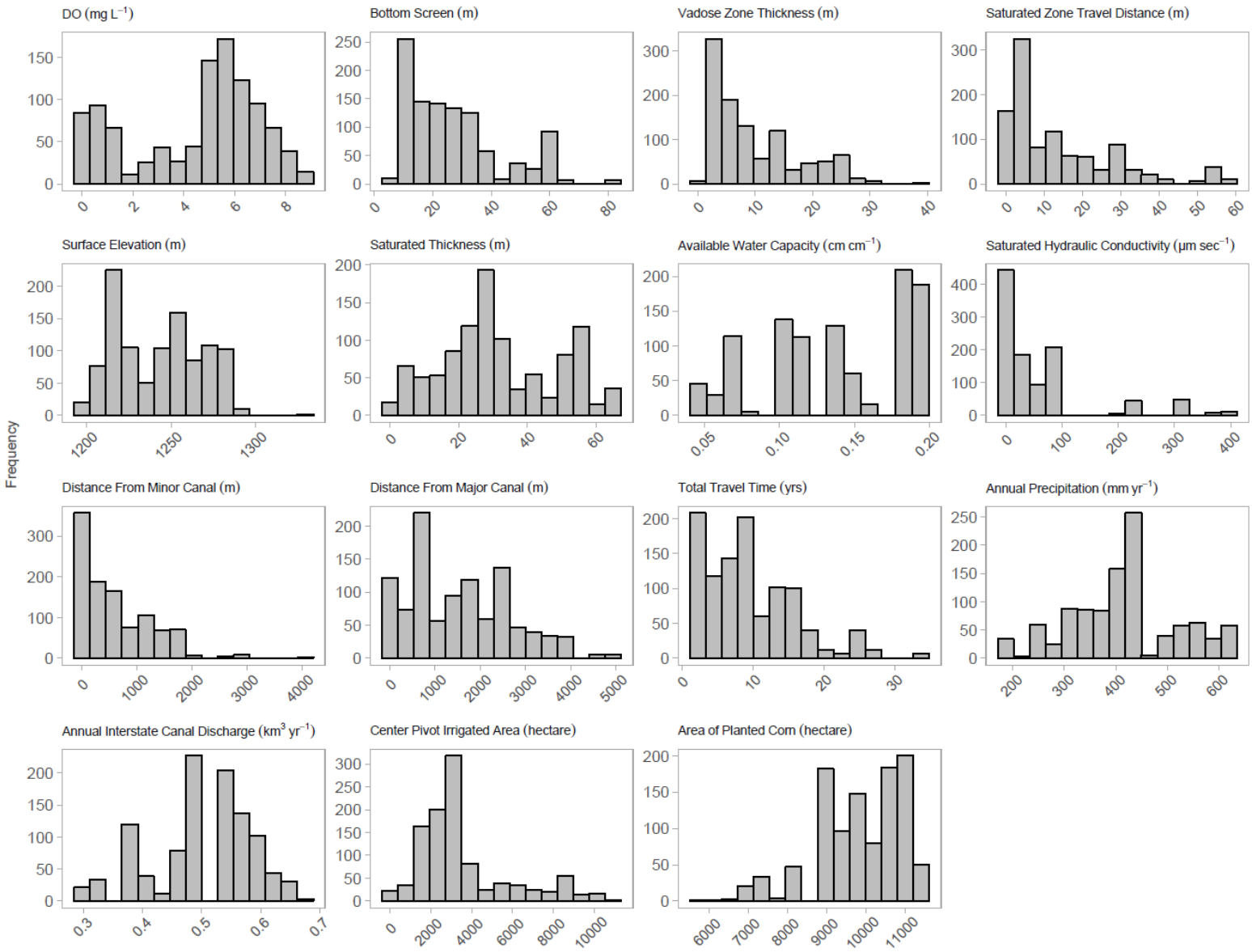


Figure F3: Explanatory variable distribution for 1,049 observations at t_{r_v}= 2.75 m/year and t_{r_s}= 2.25 m/year.

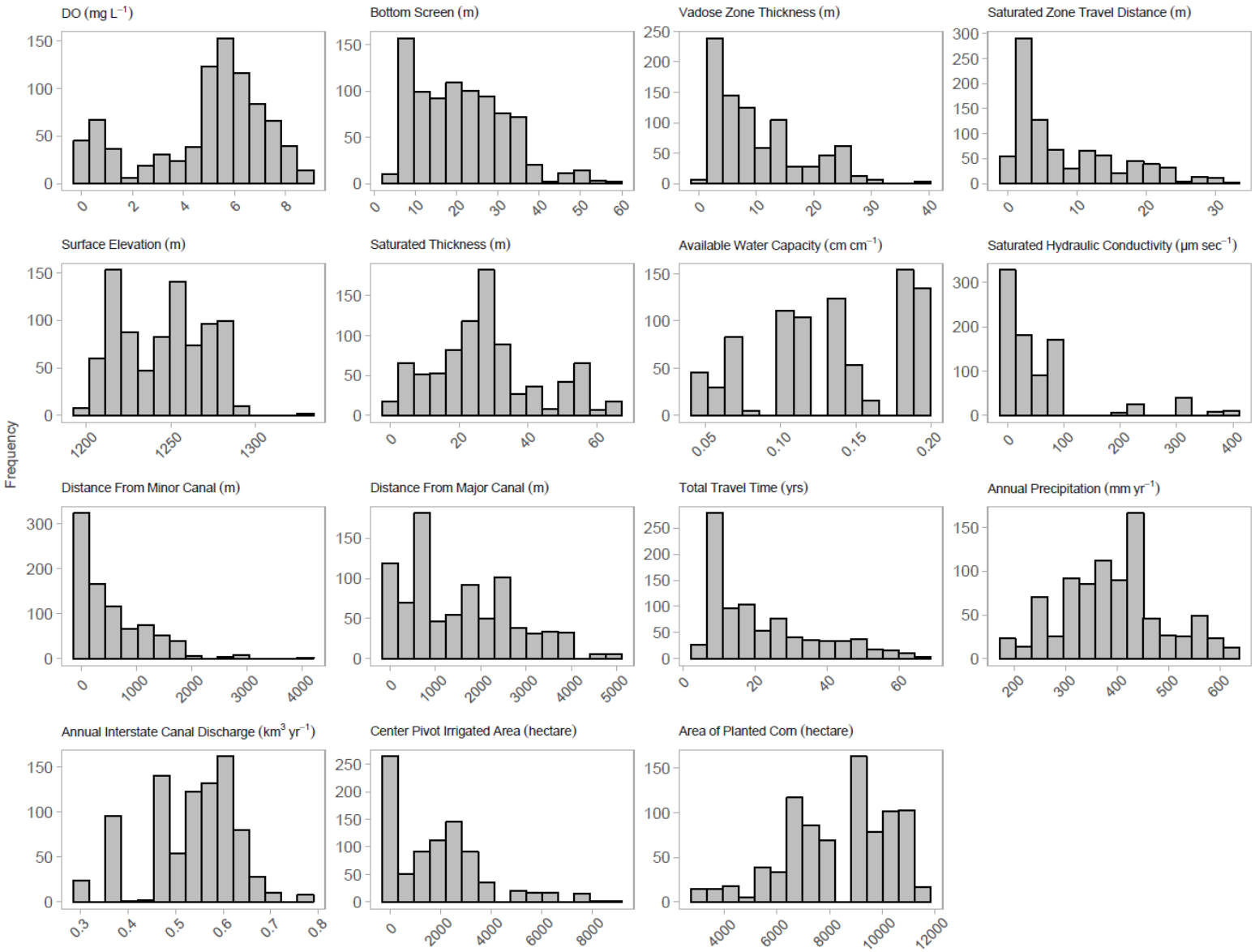


Figure F4: Explanatory variable distribution for 861 observations at t_{r_v}= 2 m/year and t_{r_s}= 0.5 m/year.

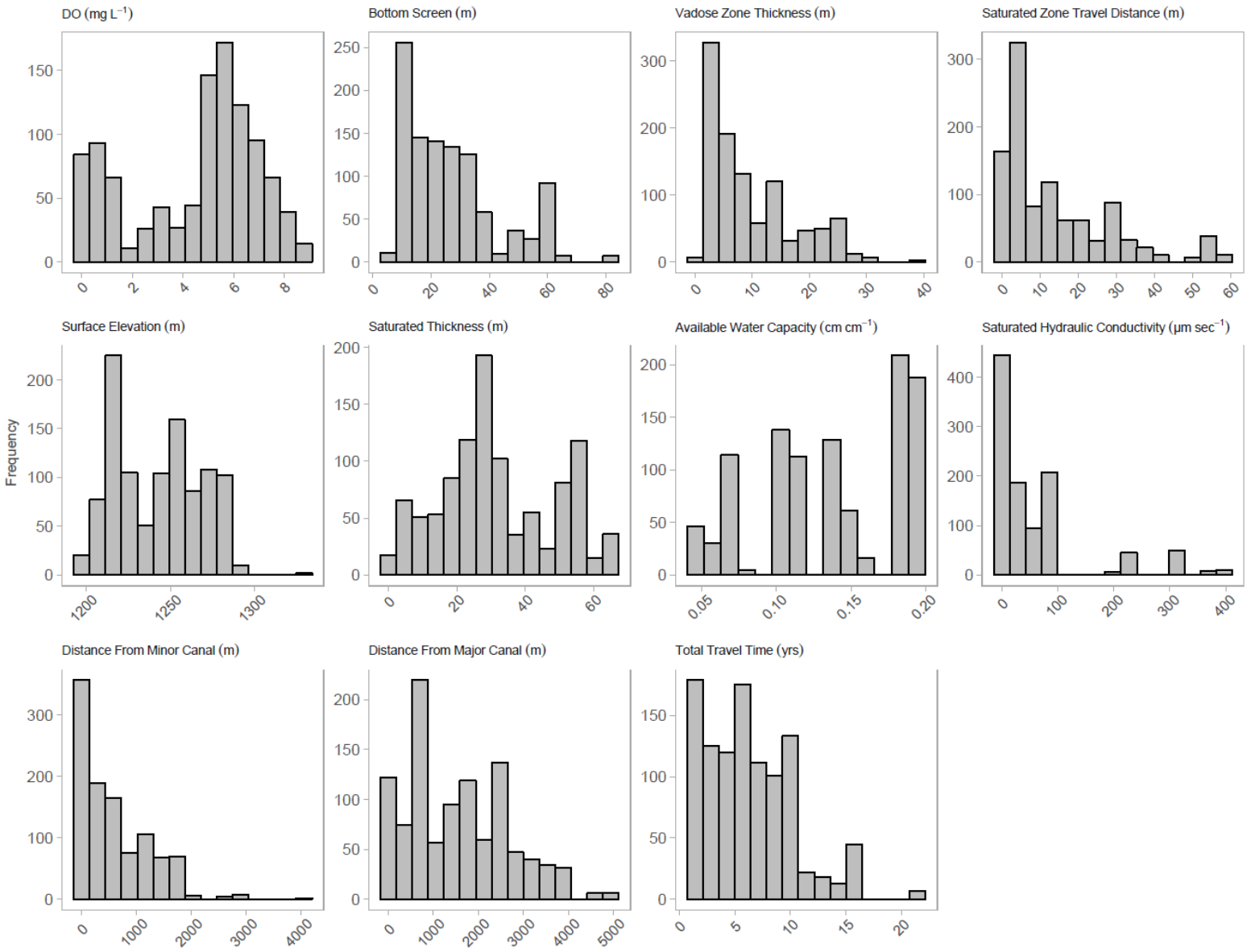


Figure F5: Explanatory variable distribution for 1,049 observations at t_{r_v} = 3.5 m/year and t_{r_s} = 3.75 m/year.

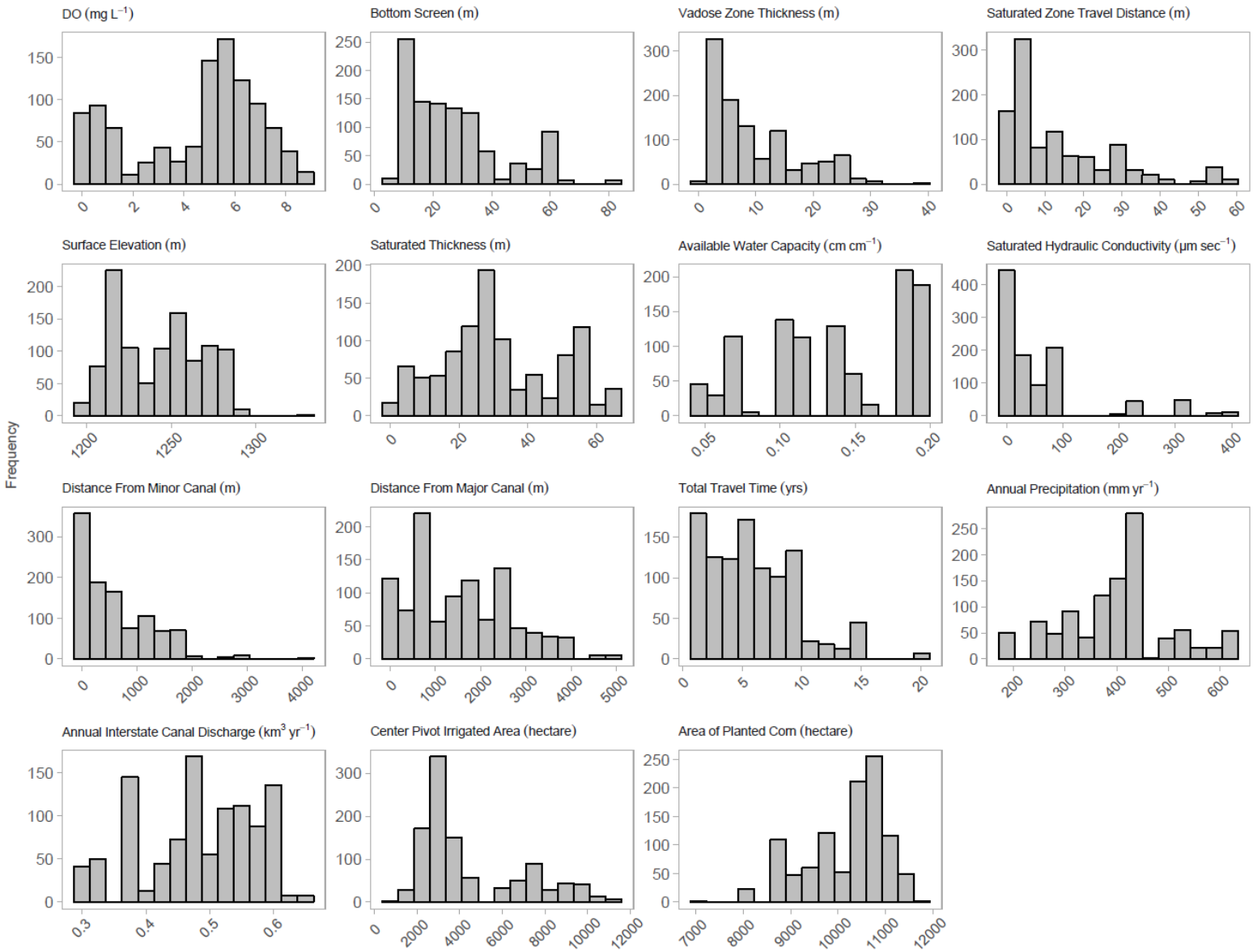


Figure F6: Explanatory variable distribution for 1,049 observations at t_{r_v}= 3.75 m/year and t_{r_s}= 4m/year.

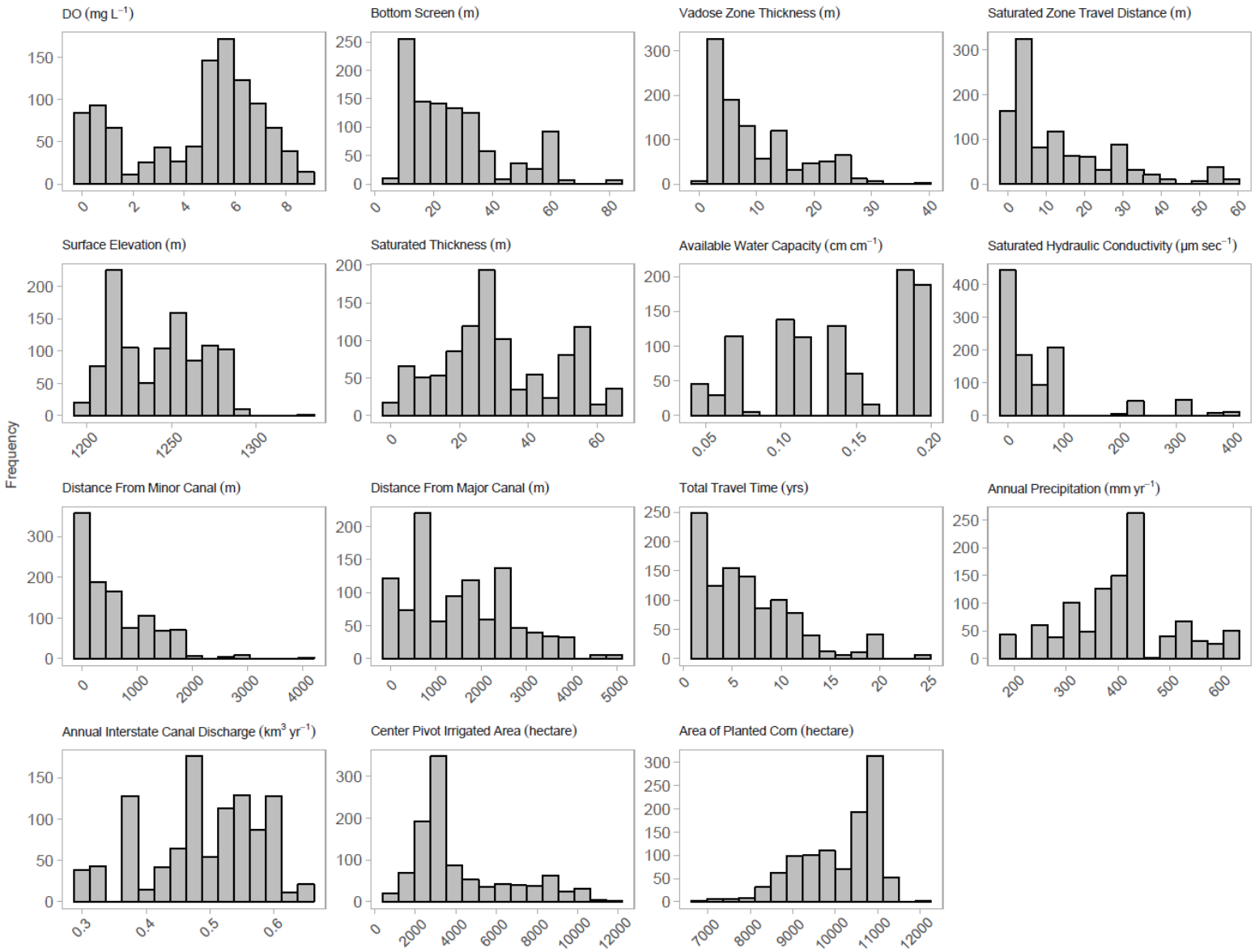


Figure F7: Explanatory variable distribution for 1,049 observations at t_{r_v} = 4.5 m/year and t_{r_s} = 3 m/year.

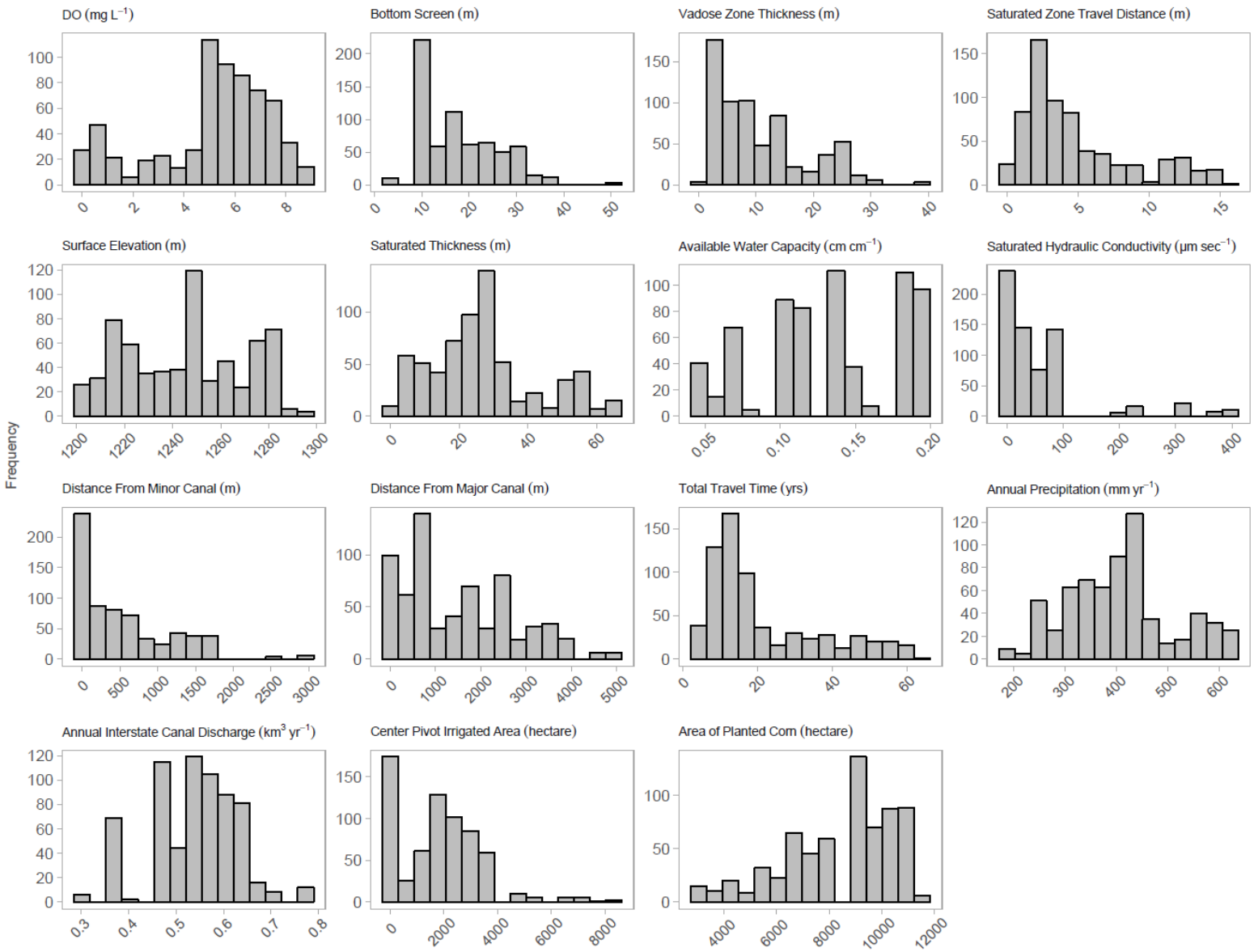


Figure F8: Explanatory variable distribution for 664 observations at t_{r_v}= 4.75 m/year and t_{r_s}= 0.25 m/year.

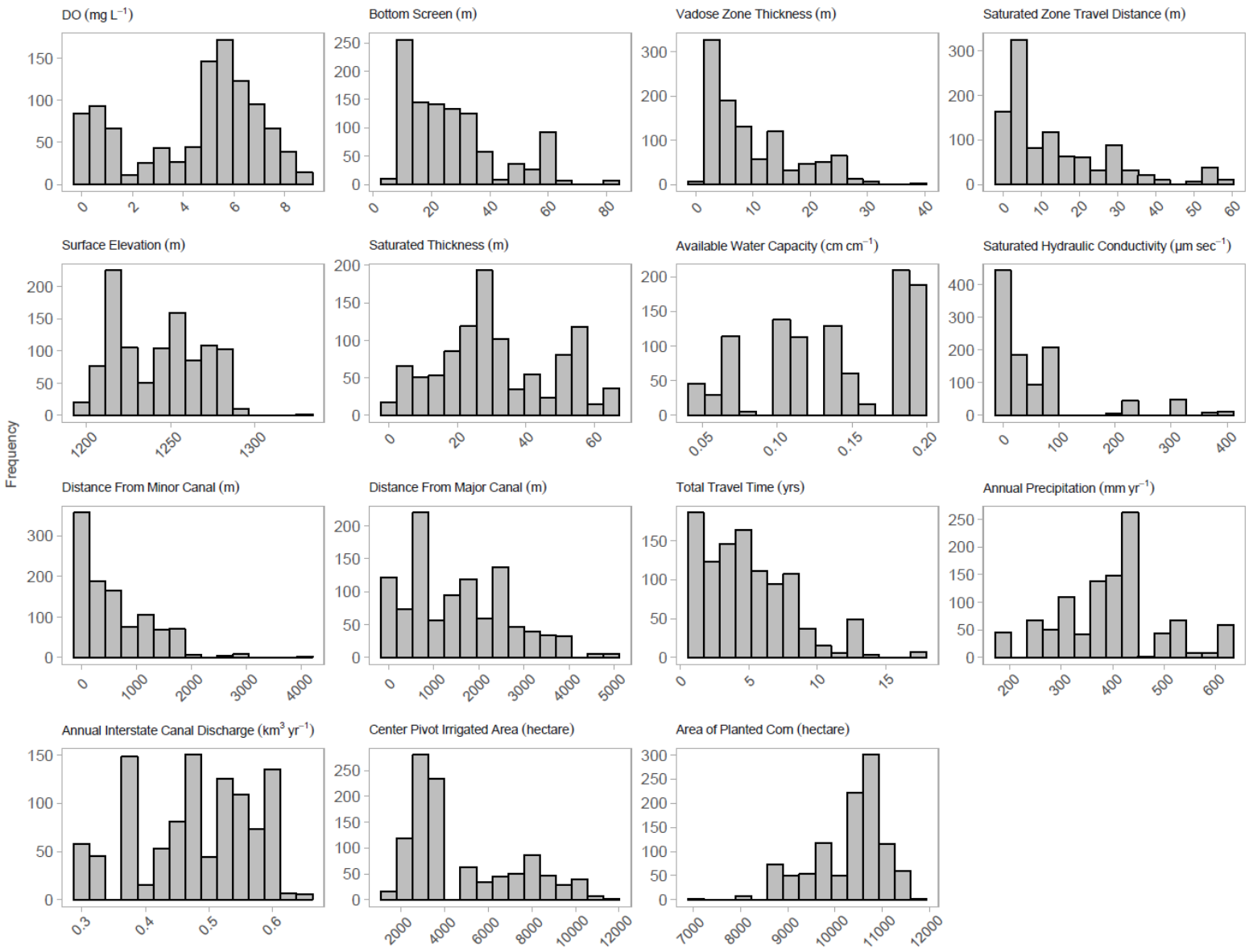


Figure F9: Explanatory variable distribution for 1,049 observations at t_{r_v}= 4.75 m/year and t_{r_s}= 4.5 m/year.

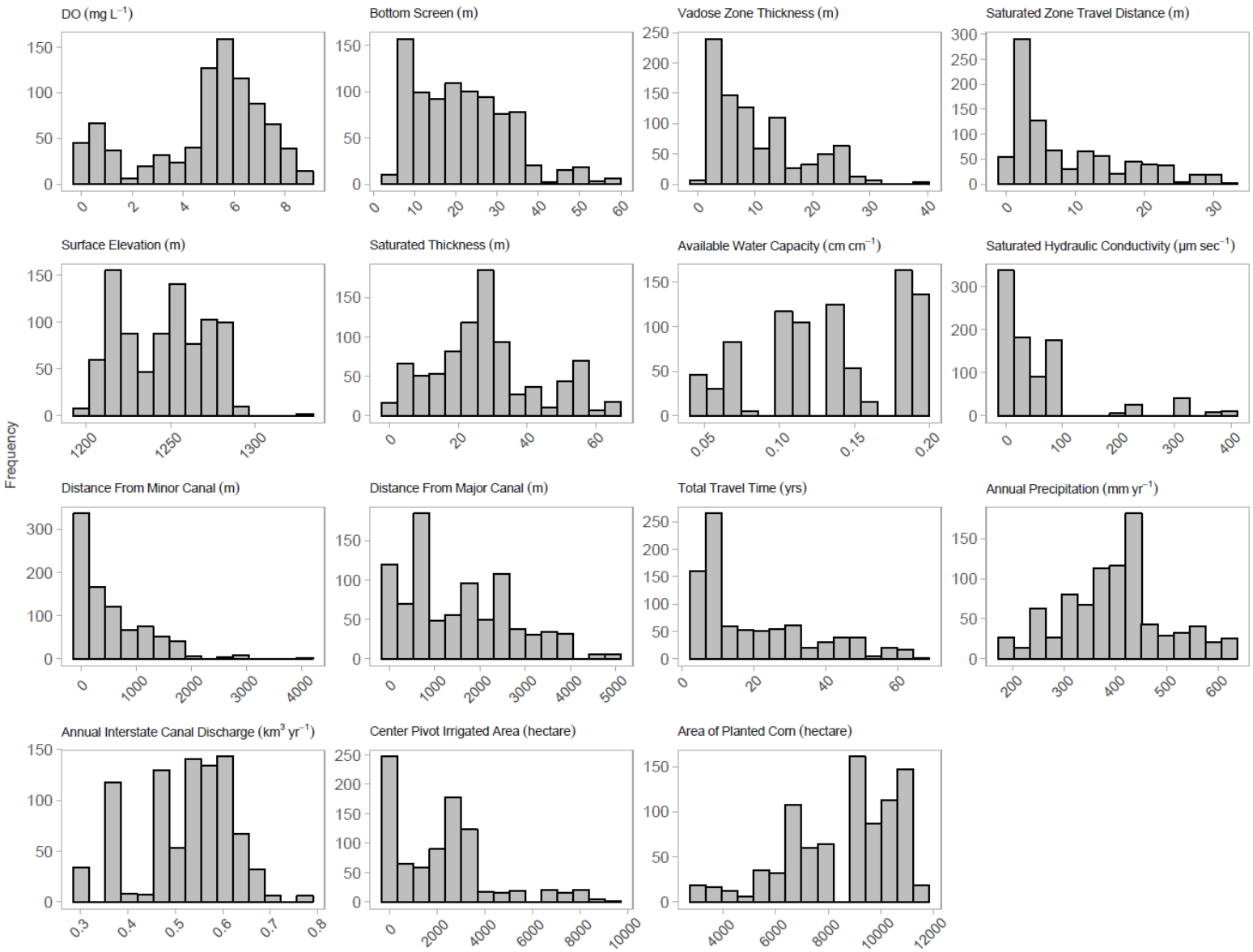


Figure F10: Explanatory variable distribution for 878 observations at t_{r_v} = 4 m/year and t_{r_s} = 0.5 m/year.

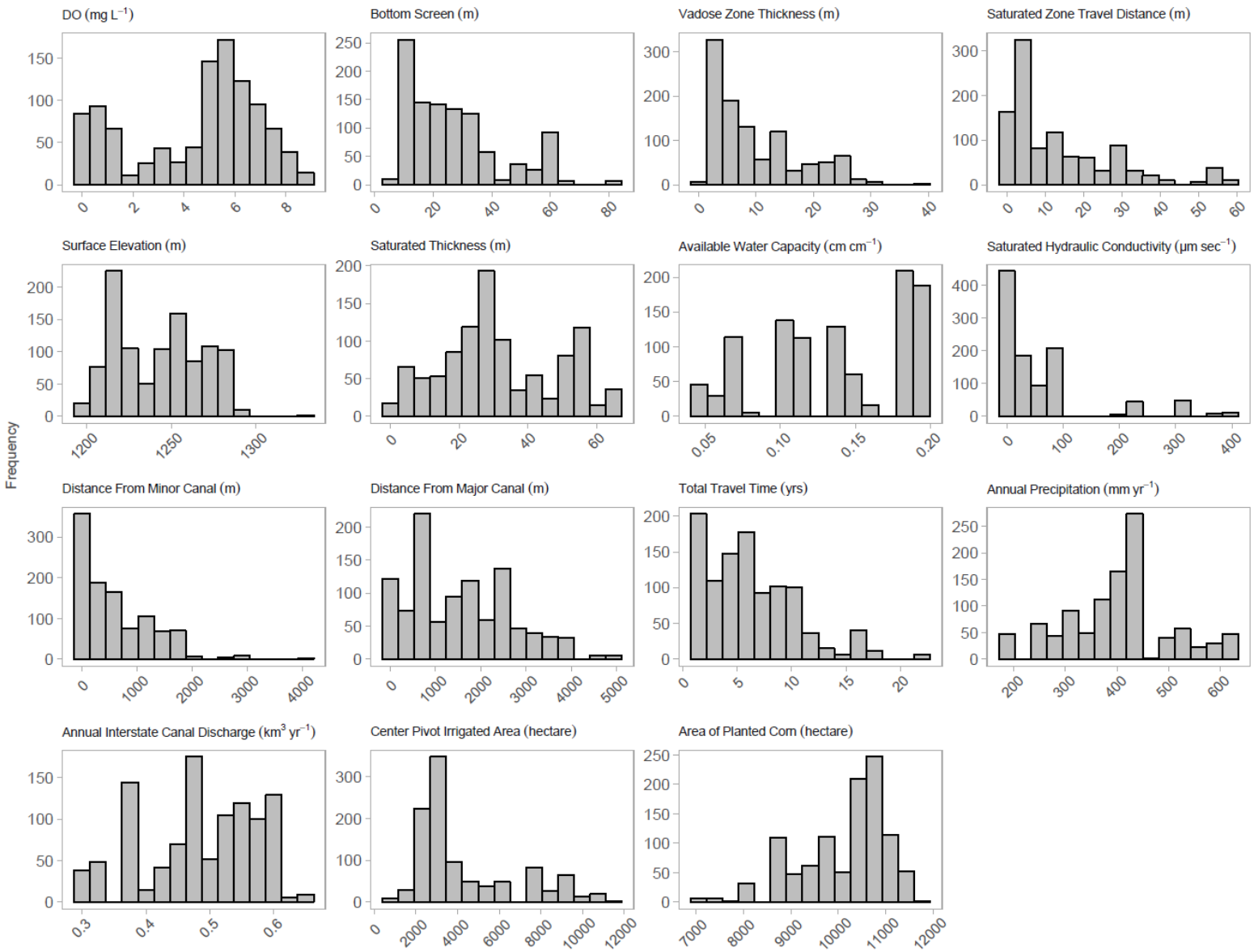


Figure F11: Explanatory variable distribution for 1,049 observations at t_{r_v}= 4 m/year and t_{r_s}= 3.5 m/year

Appendix G – Predicted vs. Observed plot for training and testing datasets.

Figure G1: Predicted [NO_3^-] vs. observed [NO_3^-] for training and testing data at t_{r_v} = 1.0 m/year and t_{r_s} = 0.25 m/year.

Figure G2: Predicted [NO_3^-] vs. observed [NO_3^-] for training and testing data at t_{r_v} = 1.0 m/year and t_{r_s} = 4.5 m/year.

Figure G3: Predicted [NO_3^-] vs. observed [NO_3^-] for training and testing data at t_{r_v}= 2.75 m/year and t_{r_s}= 2.25 m/year.

Figure G4: Predicted [NO_3^-] vs. observed [NO_3^-] for training and testing data at t_{r_v}= 2 m/year and t_{r_s}= 0.5 m/year.

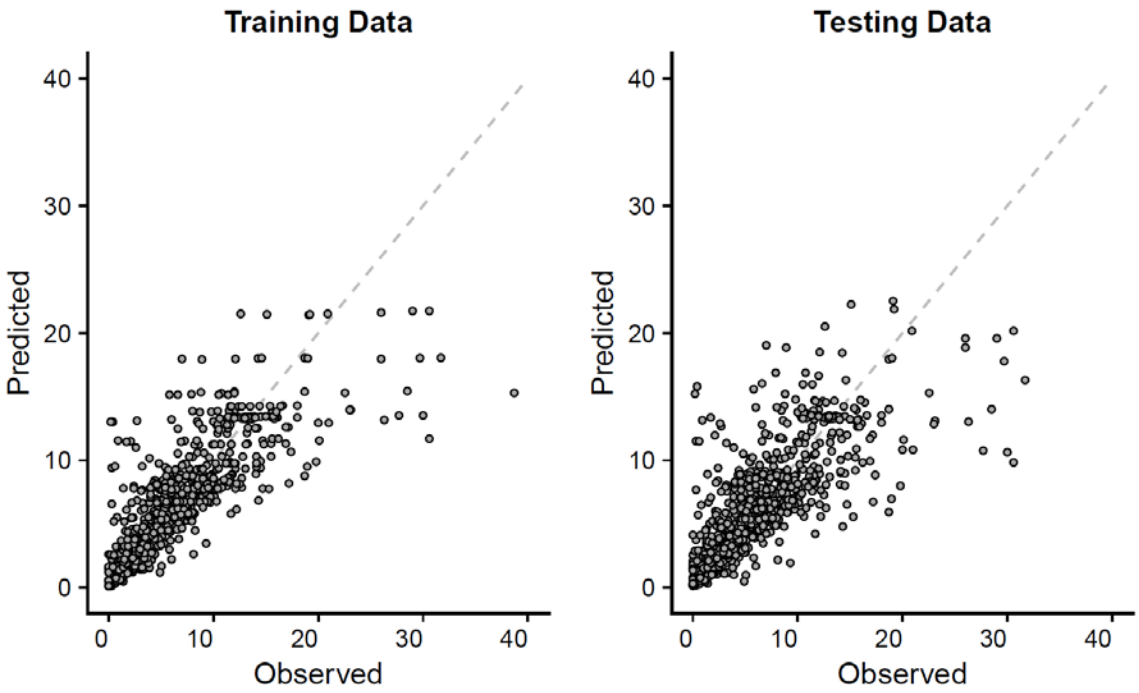


Figure G5: Predicted [NO_3^-] vs. observed [NO_3^-] for training and testing data at t_{r_v}= 3.5 m/year and t_{r_s}= 3.75 m/year.

Figure G6: Predicted [NO_3^-] vs. observed [NO_3^-] for training and testing data at t_{r_v}= 3.75 m/year and t_{r_s}= 4m/year.

Figure G7: Predicted [NO_3^-] vs. observed [NO_3^-] for training and testing data at t_{r_v}= 4.5 m/year and t_{r_s}= 3 m/year.

Figure G8: Predicted [NO_3^-] vs. observed [NO_3^-] for training and testing data at t_{r_v}= 4.75 m/year and t_{r_s}= 0.25 m/year.

Figure G9: Predicted [NO_3^-] vs. observed [NO_3^-] for training and testing data at t_{r_v}= 4.75 m/year and t_{r_s}= 4.5 m/year.

Figure G10: Predicted [NO_3^-] vs. observed [NO_3^-] for training and testing data at t_{r_v}= 4 m/year and t_{r_s}= 0.5 m/year.

Figure G11: Predicted [NO_3^-] vs. observed [NO_3^-] for training and testing data at t_{r_v}= 4 m/year and t_{r_s}= 3.5 m/year.