

Supplemental Information – Random Forest Variable Analysis

A brief summary of how each variable was evaluated is provided below.

Note: Dynamic predictors were downloaded and analyzed from 1946 (first year Interstate Canal discharge was available) to 2013 (last year of available nitrate data (Fig. S1)).

Center Pivot Irrigated Area:

Aerial imagery was used to digitize center pivot irrigated fields (Fig. S2). When available, the analysis utilized NAIP and NAPP imagery, with LANDSAT, though at a lower resolution, providing additional imagery when the former two were unavailable (Table S1). Annual center pivot irrigated area was plotted (Fig. S3), with 1999 and 2003 identified as breaks (i.e., observed notable shifts in irrigation practice) in the data, and used to interpolate irrigated area before, between, and after the breaks. That is, estimate area for 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 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. Center pivot irrigated area was set to zero from 1946 to 1972.

Center pivot and furrow irrigated fields are believed to follow an inverse trend. In other words, as center pivot irrigated fields increase over time, fields irrigated by less efficient furrow methods are assumed to decrease. With this, it is possible less nitrate would leach through the root zone with improved water application efficiency; however, it is also possible that decreased application rates reduce the dilution of $[\text{NO}_3^-]$ in the aquifer.

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 $\text{ft}^3 \text{sec}^{-1}$ to $\text{km}^3 \text{year}^{-1}$. 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, and only a relative influence was needed for the purpose of this analysis. Each nitrate observation was assigned a lagged canal discharge value depending on the transport rate combination and total travel time.

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 impact of annually high or low stream diversions on groundwater [NO_3^-].

Area of Planted Corn:

Because the study area includes parts of two counties (i.e., Scotts Bluff and Sioux), 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 respective county within the boundary of the Dutch Flats area. Next, a ratio for every 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 annual planted corn area within Dutch Flats from 1946 to 2015 was determined by multiplying each year of total county-level planted corn by the respective ratio and summing up areas.

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 also 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). 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 retrieved 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.

Available water content influences water storage and infiltration rate, and thus is a variable influencing the amount of time that irrigated water will reside 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. This study incorporated saturated hydraulic conductivity to evaluate how the ability for which water moves through a saturated vadose zone may impact $[\text{NO}_3^-]$ in the Dutch Flats area. Higher K would suggest water moves more quickly 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 distance from the water table to the midpoint between the top and bottom of the screened interval.

Bottom Screen depth has been found an important variable in other Random Forest studies (Wheeler et al., 2015). This predictor is a factor encompassing the movement of water through both the vadose and saturated zones, and evaluates how traveling through both these zones impacts $[\text{NO}_3^-]$. Saturated-zone travel distance assesses how the distance water travels from the water table to screen midpoint influences groundwater $[\text{NO}_3^-]$.

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

To estimate saturated thickness, wells were only selected for the model dataset if they had both a $[\text{NO}_3^-]$ 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. $[\text{NO}_3^-]$ 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 shallow wells were in an unconfined aquifer, the most recent depth to groundwater record at a well nest's shallow well were 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 near, or at, the top of the confining layer. The resulting saturated thickness was consistent with estimated thicknesses in the surrounding area.

Surface elevation was used, in part, to evaluate how moving further from the North Platte River (NPR) influenced groundwater [NO₃⁻]. Because the land surface generally slopes toward the river, locations of higher elevation generally would indicate wells that are further from the NPR. This predictor may also assess the impact of canals at their respective elevations. Vadose-zone thickness has been used previously to evaluate how 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.

Distance from major and minor canals:

Canal spatial data were 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, irrespective of direction, major canal and minor canal.

Since Interstate Canal discharge is a temporally dependent variable, 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 [NO₃⁻].

Dissolved Oxygen (DO):

This dataset was obtained from NPNRD (C. Hudson, Personal Communication, 2018). The dataset had several DO values per well, however, there was not a perfect 1:1 match between the collection of a nitrate sample, and the collection of DO. Therefore, all DO measurements for each well were averaged. Thus, each well was assigned a single unique DO value that was associated with all the annual median [NO₃⁻] values for the well. Analysis of DO and isotopes of nitrate suggest there has not been a major change in nitrate reduction within Dutch Flats from 1998 to 2016 (Wells et al., 2018).

DO largely drives biological processes impacting groundwater [NO₃⁻], in which case low DO could be associated with low [NO₃⁻], and high DO with higher [NO₃⁻] because of NO₃⁻ reduction in suboxic parts of the aquifer. Alternatively, in the absence of NO₃⁻ reduction, DO and [NO₃⁻] could be correlated independently with groundwater age through DO reduction and historical change in NO₃⁻ recharge.

Total Travel Time:

While this variable has a time component, it was treated as a static variable because it was held constant from one sample year to the next for each well. Total travel time was determined through several steps, also discussed in depth within Section 2.5 of the main paper, though a brief stepped summary is provided below.

- (1) The mean and standard deviation for recharge (R) data from Böhlke et al., 2007 and Wells et al., 2018 was determined and used to calculate a range of vertical velocities (V), or transport rates, through both the vadose and saturated zone. Equation 1a was assigned a mobile water content value of $\theta = 0.13$ and 0.35 , depending on whether rates were being calculated for the vadose and saturated zone, respectively.

a. $V = \frac{R}{\theta}$

- (2) Each $[\text{NO}_3^-]$ had a depth component (z) in the dataset, whether that be the vadose zone depth, or the vertical distance from the water table to the well mid-screen (i.e., saturated-zone vertical travel distance). From the vertical velocity determined in equation 1a and respective zone depth, equation 2a was used to independently estimate travel time through the vadose and saturated zones.

a. $\tau = \frac{z}{V}$

- (3) Lastly, based on the travel time determine for the vadose zone (τ_u) and saturated zone (τ_s), equation 3a provided a unique estimate for the total travel time at each well location.

a. $\tau_t = \tau_u + \tau_s$

The total travel time variable was used as a means to link dynamic variables to the date a NO_3^- sample was collected, or in essence, establish a lag time between surface activities when nitrate entered the system, and when it was sampled years later. Additionally, this variable, due to the known inverse relationship (e.g., as travel time increases, $[\text{NO}_3^-]$ decreases) observed in previous Dutch Flats studies, was utilized in the secondary analysis discussed in Section 3.2.

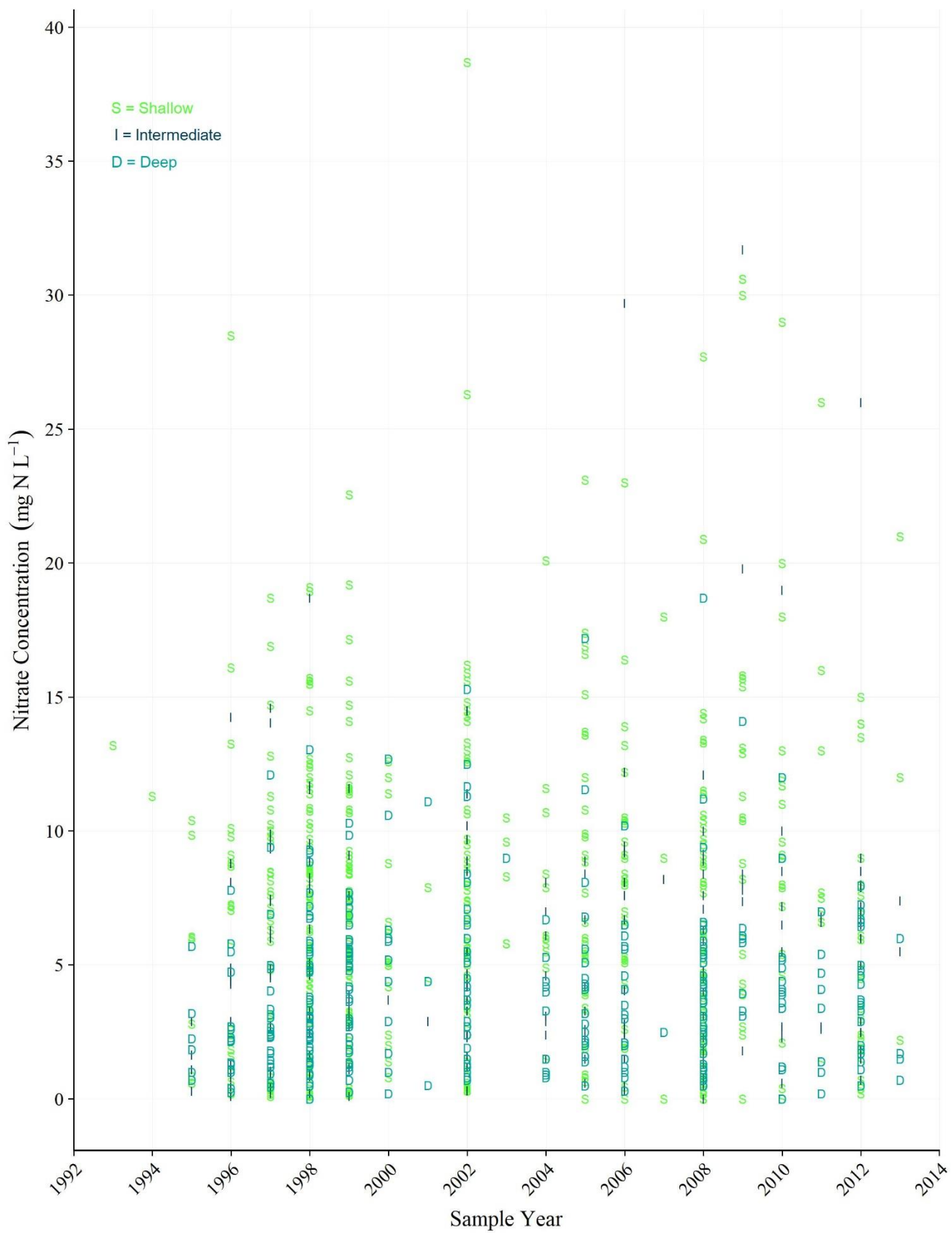


Figure S1: Plot of nitrate data used in the analysis from 1993 to 2013 (n = 1,049). Each [NO₃⁻] is classified within the plot by whether the sample was collected from a shallow (S), intermediate (I), or deep well (D).

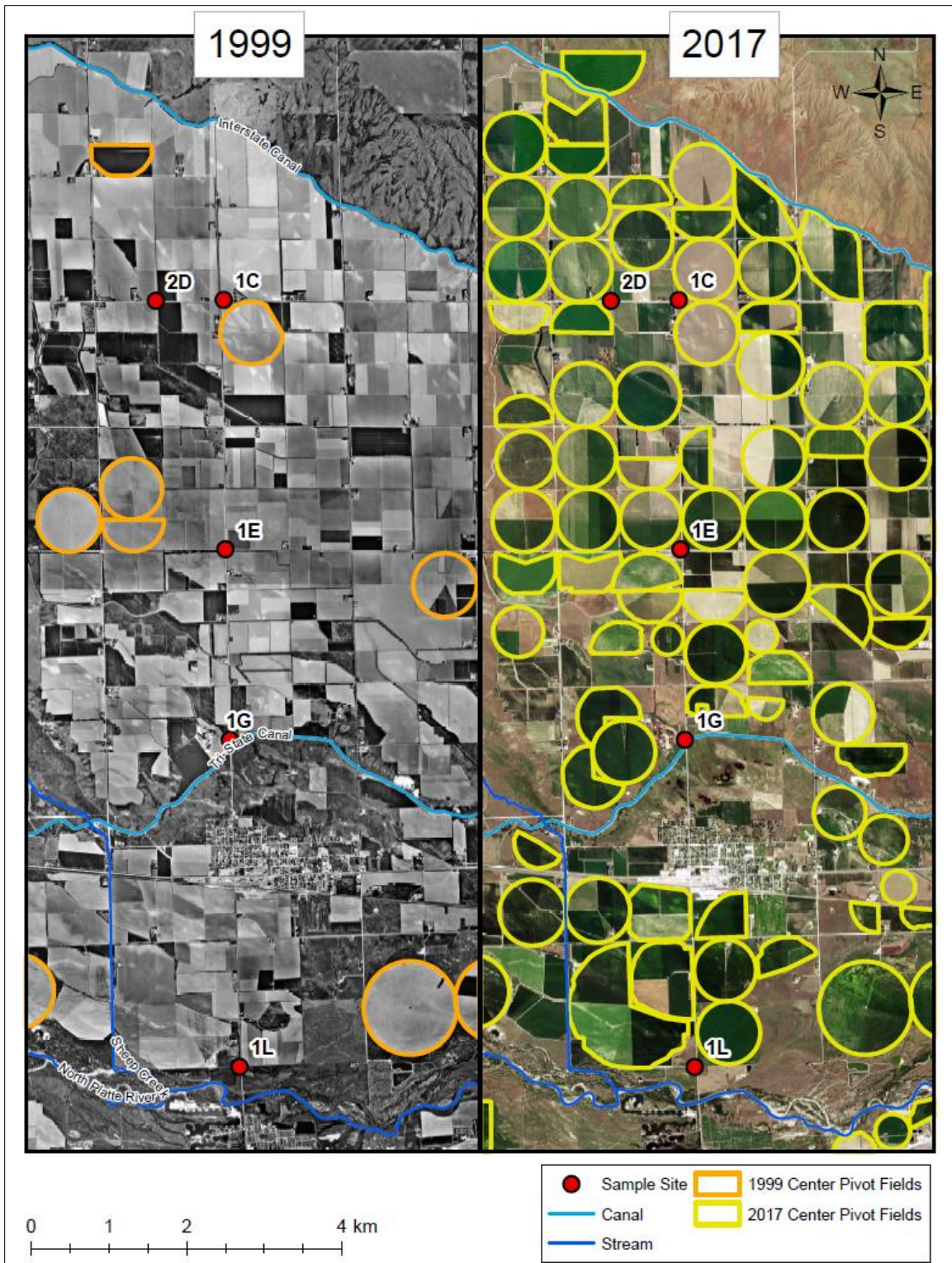


Figure S2: Visual comparison of 1999 Dutch Flats center pivot irrigated fields to 2017 center pivot irrigated fields using NAIP, NAPP, (USDA 2017) and LANDSAT imagery (USGS 2017) (Table S1). Sample sites shown in the figure are from a representative subset of wells selected for comparison in the Wells et al., 2018 study.

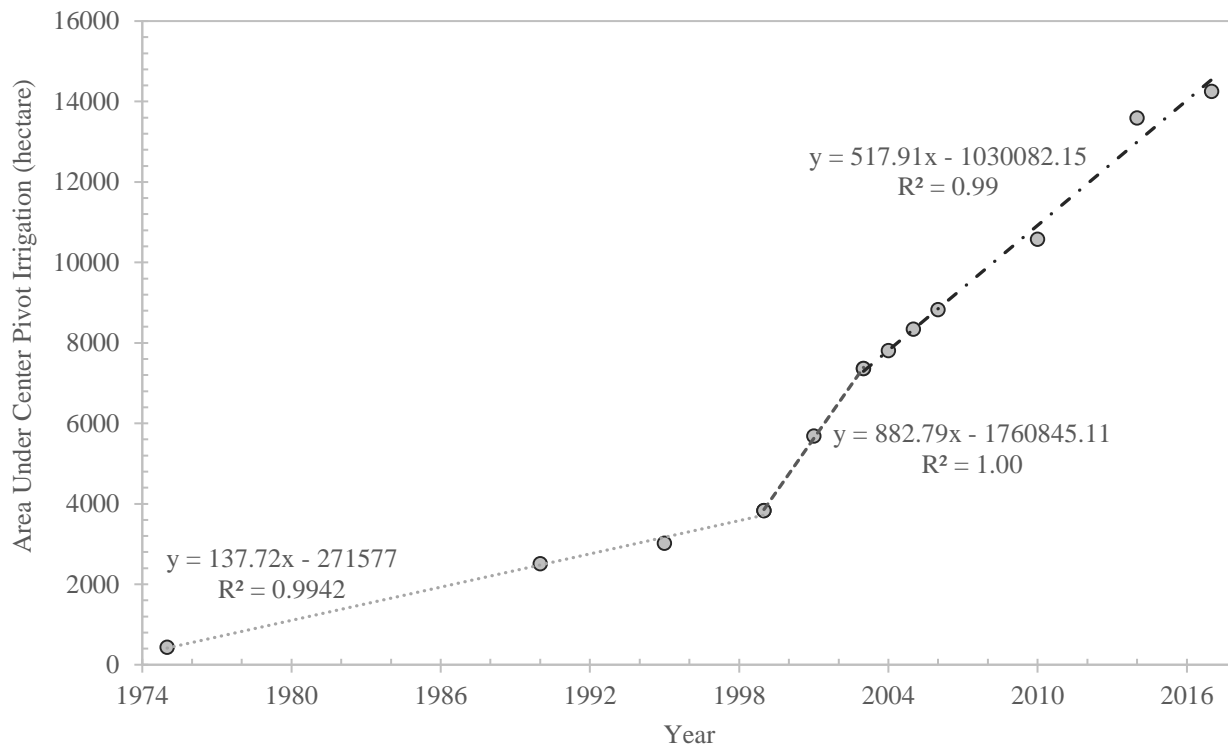


Figure S3: Center Pivot Irrigated Area based on observations from aerial imagery. The years 1999 and 2003 were used as breaks for determining linear regression equation.

Table S1: Years when aerial imagery was utilized to digitize center pivot irrigated fields.

Year	Estimated Center Pivot Irrigated Area (hectare)
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 analysed with NAIP or NAPP imagery