Papers in the Earth and Atmospheric Sciences

Earth and Atmospheric Sciences, Department of

2021

Influence of Irrigation Drivers Using Boosted Regression Trees: Kansas High Plains

Susan E. Lamb
University of Florida, susan.lamb@ufl.edu

Erin M.K. Haacker
University of Nebraska-Lincoln, ehaacker2@unl.edu

Samuel J. Smidt
University of Florida

Lamb, Susan E.; Haacker, Erin M.K.; and Smidt, Samuel J., "Influence of Irrigation Drivers Using Boosted Regression Trees: Kansas High Plains" (2021). *Papers in the Earth and Atmospheric Sciences*. 671.
https://digitalcommons.unl.edu/geosciencefacpub/671

Water Resources Research

Influence of Irrigation Drivers Using Boosted Regression Trees: Kansas High Plains

Susan E. Lamb[1] , **Erin M. K. Haacker[2]** , **and Samuel J. Smidt[1]**

[1]Soil and Water Sciences Department, University of Florida, Gainesville, FL, USA, [2]Department of Earth & Atmospheric Sciences, University of Nebraska-Lincoln, Lincoln, NE, USA

Key Points:
- The relative influence of drivers to irrigation pumping in western Kansas is modeled from 2006 to 2016 using boosted regression trees
- Site-specific factors are more influential than regional policies, and weather-related factors are more influential at longer time scales
- Irrigation decision making largely follows resource availability and operation cost considerations over localized management schemes

Supporting Information:
Supporting Information may be found in the online version of this article.

Correspondence to:
S. E. Lamb,
susan.lamb@ufl.edu

Abstract Groundwater levels across parts of western Kansas have been declining at unsustainable rates due to pumping for agricultural irrigation despite water-saving efforts. Accelerating this decline is the complex agricultural landscape, consisting of both categorical (e.g., management boundaries) and numerical (e.g., crop prices) factors that drive irrigation decisions, making integrated water budget management a challenge. Furthermore, these factors frequently change through time, rendering management strategies outdated within relatively short time scales. This study uses boosted regression trees to simultaneously analyze categorical and numerical data against annual irrigation pumping to determine the relative influence of each factor on groundwater pumping across both space and time. In all, 45 key water use variables covering approximately 19,000 groundwater wells were tested against irrigation pumping from 2006 to 2016 across five categories: (1) management/policy, (2) hydrology, (3) weather, (4) land/agriculture, and (5) economics. Study results showed that variables from all five categories were included among the top 10 drivers to irrigation, and the greatest influence came from variables such as irrigated area per well, saturated thickness, soil permeability, summer precipitation, and pumping costs (depth to water table). Variables that had little influence included regional management boundaries and irrigation technology. The results of this study are further used to target the factors that statistically lead to the greatest volumes of groundwater pumping to help develop robust management strategy suggestions and achieve water management goals of the region.

Plain Language Summary Water use for crops has lowered groundwater levels in western Kansas. Past studies have shown that this water use is driven by many factors spanning policy, economics, and the physical environment. Because of this complexity, it has been difficult to fully understand which factors most drive irrigation use relative to each other. This study uses a machine-learning model to rank the influence of 45 factors on irrigation pumping. These factors are analyzed over space (~19,000 wells across western Kansas) and time (2006–2016). Based on this study, drivers to water use include total irrigated area, summer rainfall, and depth to the water table. Factors that have little influence include management district boundaries and irrigation system type. These results are used to make water management suggestions for the region.

1. Introduction

Groundwater level decline of the High Plains Aquifer (HPA) underlying western Kansas has been extensively analyzed with most efforts focused on irrigation related to one of four domains: (1) physical (e.g., climatic and atmospheric factors and underlying geology; Edwards, 2016; MardanDoost et al., 2019; Whittemore et al., 2016), (2) agricultural (e.g., crop type, irrigation limits and management, cover crops, and irrigation efficiency; Butler et al., 2016; Kisekka et al., 2017; Pfeiffer & Lin, 2014), (3) socioeconomic (e.g., environmental policies, energy costs, irrigation efficiency, global markets, and crop prices; Hrozencik et al., 2017; Sanderson et al., 2017; Sukcharoen et al., 2020), or (4) an integrated approach among the first three categories (e.g., Haacker et al., 2019; Majumdar et al., 2020; Smidt et al., 2016). While these studies provide a comprehensive overview of the complexity of factors that drive irrigation use, little is known about the relationship between these factors across both space and time relative to irrigation pumping. For example, studies have shown that irrigation efficiency drives groundwater use (Pfeiffer & Lin, 2014), yet it cannot conclusively be determined that efficiency is more or less influential than annual weather patterns in driving total withdrawals from the aquifer. The same applies for management strategies versus crop prices or any other combination of factors compared to others. So, while there is a clear understanding of
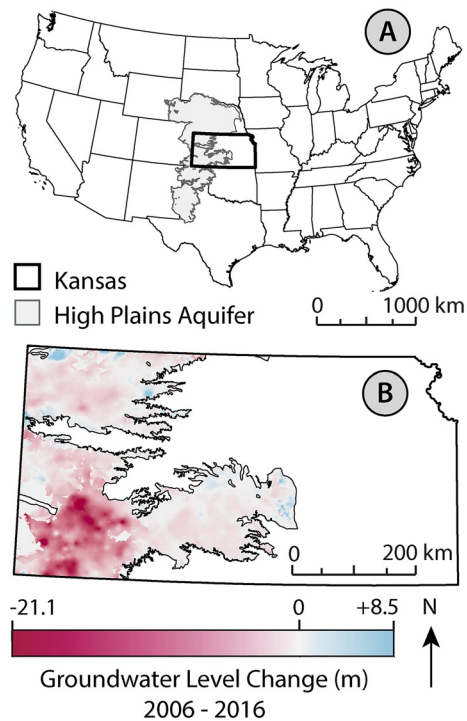
Figure 1. High Plains Aquifer site map (a) and groundwater level change for western Kansas from 2006 to 2016 (b; derived from Haacker et al., 2016).

the driving factors that lead to irrigation use, both individually and as a group (e.g., Haacker et al., 2019; Smidt et al., 2016), uncertainty still exists in understanding the spatiotemporal influence of how these factors work together to influence irrigation decision making. Measurement is a hallmark of precision agriculture and other targeted management schemes (e.g., deficit irrigation management; Rudnick et al., 2019), as farmers can often better manage the resources that they can measure. The uncertainty surrounding the relative influence (RI) of irrigation drivers leads to mismanagement and continued groundwater level decline for the region (Smidt et al., 2016).

Past studies analyzing the relationships between irrigation drivers across space and time have been limited as drivers are both numerical (e.g., average temperature and crop price) and categorical (e.g., crop type and Groundwater Management District [GMD]). One approach to overcoming these methodological data challenges has been to analyze numerical data by category and then compare results (e.g., on a GMD by GMD basis; Whittemore et al., 2016). This can be useful for resource management, since the boundaries of the GMDs correspond to some physical features of the aquifer, and these analyses provide results at the same scale at which management programs are enacted. However, this limits modeled relationships by (1) preventing categorical data from interacting directly with numerical data or (2) minimizing the scale at which drivers can impact aquifer use. Fortunately, advances in machine-learning techniques, namely boosted regression trees (BRT; Elith et al., 2008), have allowed for improved analysis when grouping disparate data. BRT is a statistical ensemble that combines regression trees and data boosting to define relationships between variables, including the simultaneous analysis of categorical and numerical variables (Elith et al., 2008). When applied to irrigation in western Kansas, BRT can accurately characterize the relative influence of sociophysical factors on annual irrigation pumping and offer strong predictive power for estimating future water use (Elith et al., 2008; Hu et al., 2017).

Here, we use BRT analysis to quantify the predominant drivers to irrigation use in western Kansas from 2006 to 2016 to (1) improve understanding of how these drivers relate, (2) develop meaningful management objectives based on these relationships, and (3) demonstrate BRT as a useful water management tool. Synchronously, we analyze 45 irrigation drivers spanning five driver groups (management/policy, hydrology, weather, land/agriculture, and economy) to quantify the predominant controls of irrigation pumping across both space and time. We further isolate dominant driver trends to target the most influential social and physical variables impacting western Kansas. The results of this analysis are then used to prioritize management efforts across the region to balance agricultural production and groundwater level declines.

2. Methods

2.1. Site Description

Agriculture is the dominant land use across western Kansas. Approximately 94% of land cover was dedicated to agricultural production from 2006 to 2016, and 77% of all cropland was composed of just six commodities: winter wheat (38%), corn (19%), sorghum (12%), soybeans (3%), alfalfa hay (3%), cotton (<1%), or a double crop planting of the six (1%; USDA-CDL, 2006–2016). During the same period, HPA groundwater levels across the region declined by an average of 2.8 m (Figure 1; derived from Haacker et al., 2016). Despite this decline, water withdrawals from the underlying HPA remain essential to the agricultural production and cultural identity of western Kansas, especially due to the region's low humidity, persistent winds, and limited precipitation relative to crop water demands (Sanderson & Frey, 2014).
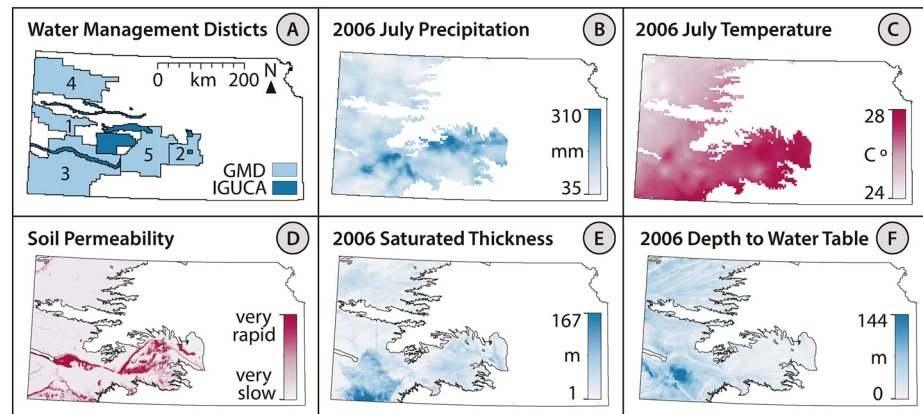
Figure 2. Example data layers for the western Kansas High Plains Aquifer region including dominant groundwater management boundaries (a; Kansas Department of Agriculture), precipitation (b; PRISM, Oregon State University), temperature (c; PRISM, Oregon State University), permeability (d; United States Geological Survey), saturated thickness (e; Haacker et al., 2019), and depth to groundwater (f; derived from Haacker et al., 2016; United States Geological Survey).

Groundwater used for irrigation in western Kansas largely sustains the expansive agricultural production of the region and can add more than $500 million in annual revenue for the state compared to dryland-only production (excluding operation costs; Smidt et al., 2019). In 2005, the groundwater was valued at $1.2 billion (as natural capital; Fenichel et al., 2015). However, continued groundwater pumping in Kansas is estimated to deplete nearly 40% of the underlying HPA in the next half century (Steward et al., 2013). This depletion will inevitably force some farmers to convert to less productive dryland operations which can produce 2–4 times lower crop yields per area (Cotterman et al., 2018; Smidt et al., 2016; Steward et al., 2013). Several strategies have been introduced to slow water loss and stabilize groundwater levels throughout the region, including efficient technologies (Pfeiffer & Lin, 2014), drought-resistant cultivars (Cotterman et al., 2018), and localized management boundaries (Deines et al., 2019). For example, five GMDs have been formed since the 1970s to develop and enforce local irrigation policies (Peck, 2006). Yet, groundwater levels have continued to decline across the region prompting the development of further management zones within and across district boundaries (K.S.A. 82a-1036, 1978; K.S.A. 82a-1041, 2012). While some areas have found success in these efforts (e.g., Deines et al., 2019), much of the region has yet to stabilize groundwater levels (e.g., Haacker et al., 2019; Whittemore et al., 2018). A clear gap remains between the intended management of groundwater, the responding use of irrigation, and the understanding of how these interact throughout the region.

Compounding this agricultural water management challenge is the socioenvironmental heterogeneities that further influence water use practices across western Kansas (Figure 2; Whittemore et al., 2018). For example, GMDs 1, 3, and 4 receive less annual precipitation (43–58 cm) than GMDs 2 and 5 (58–63 cm; Whittemore et al., 2018). In regards to management, GMDs 1, 3, and 4 operate under "planned depletion" strategies while GMDs 2 and 5 have more favorable groundwater recharge and enact a "safe yield" scheme (Peck, 2006; Whittemore et al., 2018). Other differences at the GMD level include a variability in average monthly temperature, soil permeability, depth to water table, and interstate compacts (IC). Collectively, through these heterogeneities, western Kansas is a unique case study of mixed physical and social variability with high data resolution valuable for informing progressing agricultural water use and management techniques in this region and elsewhere.

As part of the state's water management agenda, the Kansas Department of Agriculture and Kansas Geological Survey have continually managed an open-access Water Information Management and Analysis System (WIMAS) since 1996 with annual water use records for over 45,000 wells across the state. Of these wells, 19,000 are located in western Kansas with access to the HPA. Associated data include information such as well type, installation depth, depth to water, pumping amount, pumping allotment based on government regulation, irrigation system type, and crop type. No other region of the HPA has this type of data resolution, likely even in private databases. Collectively, this combination of driver complexity and data

resolution make this location an ideal selection for using BRT analysis to better understand influences on irrigation.

2.2. Data

Data used in this study can be summarized into independent variable (i.e., drivers to irrigation pumping) and dependent variable (i.e., irrigation pumping) categories. Collectively, we analyzed 45 independent variables and 1 dependent variable. The dependent variable is total annual irrigation pumping amount reported on the well level. The independent variables were identified as influential to irrigation use based on published literature and informal conversations with experts, colleagues, and stakeholders and are organized into five categories: management/policy, hydrology, weather, land/agriculture, and economics. All data used in the analysis are summarized in Table 1.

Well-specific data, such as annual pumping data and crop irrigated, for approximately 19,000 agricultural irrigation wells in western Kansas for years 2006–2016 were downloaded through the WIMAS maintained by the Kansas Department of Agriculture, Division of Water Resources and Kansas Geological Survey (KDA-DWR & KGS, version 5). These data, which included GPS coordinates for each pumping well, were then read into ArcGIS (version 10.5). Spatial driver data were downloaded as or converted to raster files in ArcGIS. If downloaded, cell sizes were kept consistent with their original format. If converted from a shapefile, we assigned a standard 11.1-m × 11.1-m cell size. Well density was produced in with a 1.11 km by 1.11 km cell size. Annual, statewide market crop value data for alfalfa, corn, sorghum, soybeans, and wheat grown in Kansas were accessed from the United States Department of Agriculture and integrated into the well data attribute table (USDA-NASS, 2006–2016). Saturated thickness and water table elevation were created using the methods described in Haacker et al. (2016). Depth to water table was created in ArcGIS using the aforementioned water table elevation data and a digital elevation model from the United States Geological Survey. The Right variable was generated by concatenating WIMAS data set variables relating to right type as well as status of the water right file, water right, and point of diversion. The shapefiles for intensive groundwater use control areas (IGUCAs) are public data and were requested from the Kansas Department of Agriculture via an Open Records Request. All data were unique annually except for boundary data (e.g., GMD, Basin), hydraulic conductivity, and soil permeability as they were effectively static over the temporal range of the study. No predetermined weights were applied to predictor variables in the analysis due to the exploratory nature of the research. All data were spatially aligned using the NAD 1927 Geographic Coordinate System and USA Contiguous Albers Equal Area Conic projected coordinate system. All driver data values were then spatially attributed to each well annually from 2006 to 2016 and exported in spreadsheet format (i.e., columns are variables and rows are individual wells) for use in the boosted regression tree modeling.

2.3. Description of Boosted Regression Trees

Boosted regression trees draw on both statistical and machine-learning techniques to determine the relative influence of each predictor variable (i.e., independent variable) on a response variable (i.e., dependent variable; Elith et al., 2008). Specifically, BRT relies on two algorithms: (1) regression trees and (2) boosting. Regression trees are decision trees that have repeating binary splits to identify how a subset of predictor variables relate to the response variable based on a defined predictor value as a split point (e.g., May precipitation >35 mm; Elith et al., 2008). Decision tree complexity (i.e., total number of trees and number of nodes in a tree) is defined by the user. Each regression tree culminates with the calculation of a residual value. Boosting then uses the results of a regression tree (i.e., residual) to improve the fit of the next tree (i.e., improve the residual). This sequence progresses stage-wise through the model rather than stepwise, thus existing trees are not changed but instead the model estimates are updated as new trees are added (Elith et al., 2008). This process continues until the defined number of trees is reached or the residual has reached its optimum value, at which point improvements in model estimates are negligible. A learning rate further defines the contribution, or weight, of each tree to the model. Based on the results of the boosted regression trees, the model can then quantify the relative influence of each predictor variable on the response variable. Refer to Hastie et al. (2009) for more information and extended derivation of regression trees, Ridgeway (2007) for boosting, and Elith et al. (2008) and Friedman and Meulman (2003) for an integration of the two.

Table 1
Data, Descriptions, and Sources

Variable	Description	Units/scale	Mean; standard deviation	Source
Dependent variable				
Amt_Irr	Total annual irrigation pumping	m^3/year	204,479; 152,316	1
Independent variables				
Management/policy				
GMD	Groundwater Management District	Boundary	Categorical	1
IGUCA	Intensive Groundwater Use Control Area	Boundary	Categorical	2
RC	Rattlesnake Creek Management Plan	Boundary	Categorical	3
IC	Interstate compacts	Boundary	Categorical	3
Right	Water rights	Concatenated	Categorical	1
Auth_Amt	Authorized amount of irrigated water	m^3/year	326,391; 273,905	1
Auth_Rate	Authorized pumping rate	m^3/min	3; 12	1
Auth_Area	Authorized irrigated area	ha/year	169; 262	1
Hydrology				
Basin	River basin	Boundary	Categorical	1
ST	Saturated thickness	Meters, 250-m × 250-m	36; 27	6
WT_Elev	Interpolated water table elevation	Meters, 250-m × 250-m	771; 171	6
K	Aquifer hydraulic conductivity	Contours, 0–90 m/day	Categorical	5
Weather				
T01, T02, ...	Mean monthly temperature	Celsius, 4-km × 4-km	13; 2	4
T_Annual	Mean annual temperature	Celsius, 4-km × 4-km	13; 1	4[a]
P01, P02, ...	Total monthly precipitation	mm, 4-km × 4-km	48; 36	4
P_Annual	Mean annual precipitation	mm, 4-km × 4-km	48; 15	4[a]
Land/agriculture				
SP	Mean soil permeability	Very slow to very rapid	Categorical	3
System	Irrigation technology	Type	Categorical	1
Area_Irr	Reported annual irrigated area	ha/year	58; 57	1
WD	Well density	Wells/km^2, 0.01° × 0.01°	0.4; 0.2	1
Crop	Reported commodity	Type	Categorical	1
Economics				
CV_Area	Annual KS crop price per unit area	$/ha	Crop specific	5
Depth	Depth to water table	Meters, 250-m × 250-m	41; 28	3, 6[a]

Note. Mean and standard deviation of collected data from 2006 to 2016. Source key: (1) Water Information Management and Analysis, maintained by the Kansas Department of Agriculture, Division of Water Resources and Kansas Geological Survey; (2) Kansas Department of Agriculture, Division of Water Resources – Open Records Request; (3) United States Geological Survey; (4) PRISM Climate Group, Oregon State University; (5) United States Department of Agriculture; (6) Haacker et al. (2016).
[a]Indicates that data were derived from these sources rather than utilized directly.

In addition to its ability to analyze both numerical and categorical data, a BRT model was selected for this study as it is not sensitive to outliers and missing data in predictor variables (Elith et al., 2008). BRT was chosen over a random forest model, which tends to perform poorly when there are many low-influencing variables (Hastie et al., 2009). Furthermore, statistical approaches such as generalized additive models were not utilized as, unlike BRT, interactions between predictor variables are not automatically modeled (Elith et al., 2008; Hastie et al., 2009).

2.4. Construction of BRT Model

We used foundational R code (R Development Core Team, 2018) from Elith et al. (2008) to implement the BRT. While this code utilized the *gbm* R package, we used its expanded version for this study, the *dismo* R package (v. 3.5.2 for Mac and v. 3.5.1 and v.3.6.2 for Windows; Hijmans et al., 2017). This code was then modified to fit the conditions of our data set.

In BRT analyses, five conditions must be predetermined: (1) the total number of trees to be used in the analysis, (2) distribution of the loss function, (3) bag fraction (i.e., the amount of data randomly selected at each step and not replaced), (4) the tree complexity (i.e., number of nodes in each tree), and (5) the learning rate of the model (i.e., the influence of each tree in the model; a value of 0–1) (Elith et al., 2008). Here, we did not set a limit to the number of trees and allowed the model run until it optimized the residual loss on the response variable. We used a Laplace distribution because it is optimal for data sets with a continuous response variable (Ridgeway, 2007) and provides a more robust fit to the data compared to other distribution options (Lampa et al., 2014; Murphy, 2012), which we believe is more suitable for agricultural data that often have large variability. The bag fraction was set at 0.5. The learning rate was set to 0.05, and the tree complexity to 24, as determined through a calibration and validation process.

We ran a BRT model using the 2016 data for learning rate (*lr*) values of 0.1, 0.05, and 0.01 and tree complexity (*tc*) values of 2–18 in increments of 2. We utilized two values from models run on each combination of parameters (1) a tenfold cross validation (CV) statistic and (2) r-squared statistic. For each combination of parameters. We maintain that CV is an appropriate metric for evaluating the best model, as BRT differs from other statistical methods in that there are no *p* values and degrees of freedom are difficult to identify (Elith et al., 2008). Also, CV can be a more robust sensitivity analysis for machine-learning models rather than Akaike information criterion (Hauenstein et al., 2017). For the r-squared analysis, we conducted a BRT model on 50% of the data and used the function predict.gbm to predict the remaining 50% of data. This prediction was plotted against the observed values to determine the r-squared statistic.

BRT is a stochastic technique, thus the CV and r-squared statistics change marginally between iterations even if conducted on the same parameter combinations. Because of this, we utilized the mean CV statistics for each learning rate, and the mean r-squared statistics for each tree complexity. Identifying the highest mean CV and r-squared statistics, we determined that a learning rate of 0.05 and a tree complexity of 24 would produce the best performing model. Please reference the Supplemental Information for more detail on the calibration and validation steps.

2.5. Application of BRT Models

Once parameters were identified, we conducted four groups of BRT models: (1) annually, (2) annually with pumping normalized by irrigated area, (3) all years grouped together, and (4) all years grouped together with pumping normalized by irrigated area. Annual models (Groups 1 and 2) utilized unique data for each year, whereas aggregated models (Groups 3 and 4) examined all years of data simultaneously and time was not distinguished as a variable. We normalized pumping by irrigated area in two of the groups to eliminate an anticipated strong correlation between total pumping and total irrigated area. We also initially ran each year as an individual BRT model to systematically flag the noninfluential variables using the *gbm.simplify* function of the R code. This allowed us to manually reduce noise by eliminating noninformative variables prior to our analysis. Note that the collective influence of variables will always add up to 100, and even noninformative variables will be assigned a nonzero value in a BRT analysis, albeit quite small. We used the simplified sets for Groups 1 and 2. We could not use the simplified sets for Groups 3 and 4 as noninformative variables were flagged in some years but not others (Table 2), leaving incomplete driver data sets for the groups with years combined together. We also used only a randomly selected 25% of data for each variable in Groups 3 and 4, as this still included over 2.7 million data points and challenged computing capabilities (Dell Precision 5820 desktop computer, 2018, Windows 10, Intel(R) Xeon(R) W-2123 CPU @ 3.60 GHz, 32 GB RAM). A summary of completed models is outlined in Table 2.

Table 2
BRT Models (Size, Cross Validation, and Deviance)

Year	Variables removed	Number of trees	CV (mean, standard deviation)	Deviance (mean, standard deviation)
Group 1: 11 models, annual, simplified (gbm.simplify)				
2006	4 (GMD, IC, IGUCA, RC)	2,150	0.834, 0.003	43.257, 0.379
2007	4 (GMD, IC, IGUCA, RC)	1,600	0.836, 0.004	40.124, 0.438
2008	2 (IC, RC)	1,650	0.862, 0.003	41.055, 0.333
2009	4 (GMD, IC, IGUCA, RC)	1,900	0.851, 0.002	38.424, 0.349
2010	4 (GMD, IC, IGUCA, RC)	2,000	0.857, 0.002	38.644, 0.273
2011	2 (GMD, RC)	2,000	0.874, 0.004	44.883, 0.387
2012	2 (IGUCA, RC)	1,800	0.859, 0.004	42.913, 0.420
2013	5 (GMD, IC, IGUCA, RC, Right)	1,450	0.860, 0.003	41.058, 0.404
2014	5 (GMD, IC, IGUCA, RC, Right)	1,800	0.860, 0.002	40.486, 0.371
2015	6 (GMD, IC, IGUCA, K, RC, Right)	1,600	0.833, 0.003	37.077, 0.188
2016	4 (GMD, IC, IGUCA, RC)	1,900	0.845, 0.003	36.923, 0.302
Group 2: 11 models, annual, simplified (gbm.simplify), normalized by area irrigated				
2006	4 (GMD, IC, IGUCA, RC)	2,150	0.550, 0.030	0.315, 0.002
2007	4 (GMD, IC, IGUCA, RC)	2,150	0.631, 0.007	0.292, 0.002
2008	2 (IC, RC)	1,900	0.641, 0.014	0.296, 0.002
2009	4 (GMD, IC, IGUCA, RC)	2,350	0.617, 0.016	0.278, 0.003
2010	4 (GMD, IC, IGUCA, RC)	2,150	0.634, 0.005	0.282, 0.001
2011	2 (GMD, RC)	2,700	0.560, 0.007	0.336, 0.003
2012	2 (IGUCA, RC)	2,550	0.566, 0.007	0.327, 0.003
2013	5 (GMD, IC, IGUCA, RC, Right)	1,950	0.653, 0.009	0.301, 0.003
2014	5 (GMD, IC, IGUCA, RC, Right)	2,500	0.619, 0.006	0.294, 0.002
2015	6 (GMD, IC, IGUCA, K, RC, Right)	2,200	0.553, 0.040	0.278, 0.002
2016	4 (GMD, IC, IGUCA, RC)	2,450	0.632, 0.006	0.270, 0.002
Group 3: 1 model, combined years (25% from each year)				
Combined years	No removed variables	2,650	0.857, 0.002	41.199, 0.230
Group 4: 1 model, combined years (25% from each year), normalized by area irrigated				
Combined years	No removed variables	2,750	0.597, 0.027	0.310, 0.002

3. Results

The relative influences of each irrigation driver on groundwater pumping amount for the Group 1 models (annual, simplified) and Group 3 model (years combined, unsimplified) are displayed in Figure 3; each variable is further color coded by driver category.

Drivers are arranged by decreasing mean RI for Group 1 over the temporal range of the study, where the boxes on the Group 1 boxplots are the interquartile range (IQR) and the whiskers extend to the minimum and maximum values. Hollow points indicate RI outliers from Group 1 results. In Group 1, irrigated area was the most dominant driver with a mean RI of 18.0%, followed by authorized amount, saturated thickness, and authorized pumping rate (mean RI of 5.3%, 5.3%, and 4.7%, respectively). Localized management drivers had the least influence on irrigation pumping, with GMDs accounting for 0.1% of influence, IGUCAs for 0.1%, and IC for 0.2%. The RI of weather-related drivers ranged from an 0.6% to 5.3%, with annual precipitation being the greatest influencing weather variable by mean. Weather variables also had the greatest occurrence of RI outliers across time, likely corresponding to extreme weather events. Regardless of the month, precipitation had a higher median RI on irrigation pumping than did temperature. For management/policy variables, well-scale policies were strong drivers to irrigation pumping whereas regional-scale,
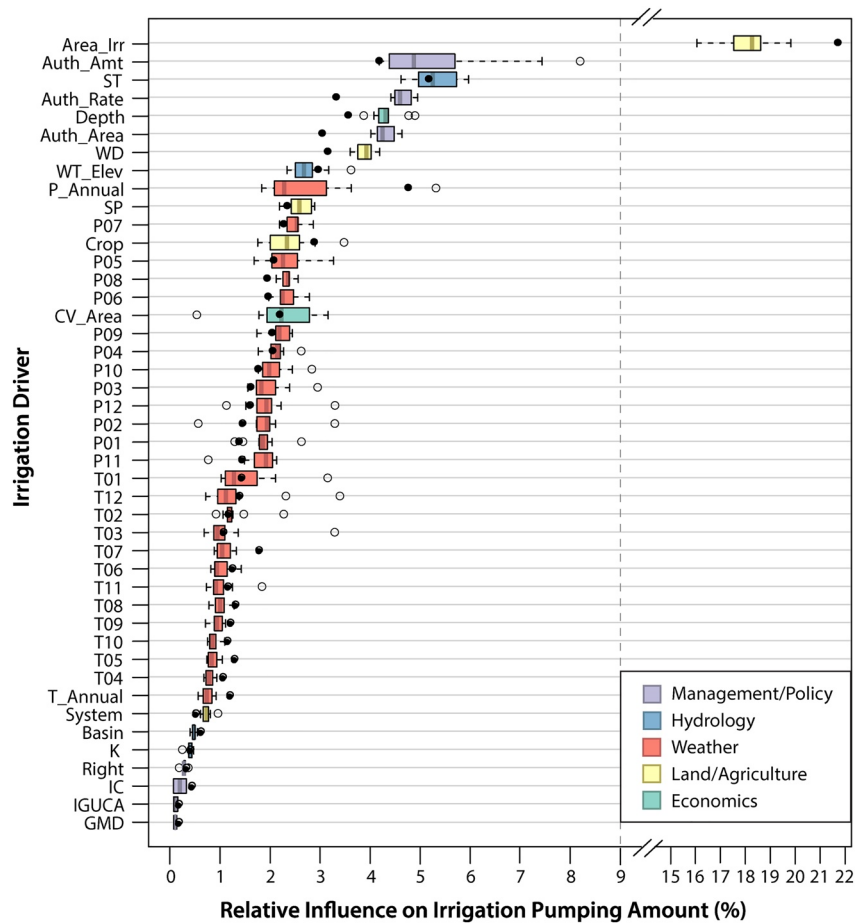
Figure 3. Boxplots showing RI outputs from BRT runs conducted annually from 2006 to 2016 (Group 1). Black points show RI from a single BRT run conducted on collective data from 2006 to 2016 (Group 3). Note scale change following *x*-axis break. RI, relative influence; BRT, boosted regression trees.

boundary level policies were not strong drivers. The Rattlesnake Creek Management Plan was also a district boundary examined in this study but was classified as noninfluential by *gbm.simplify* across all years and was not considered in the Group 1 model and thus excluded from Figure 3. Generally, variables with the greatest variance in RI across years were among top influencing variables, and those with the smallest variance in RI across years were among low-influencing variables. With the exception of irrigated area, all driver variables reported RI values of less than 10% in any given year, with most being less than 5% in all years. All five variable categories are represented in the top nine influencing drivers, with the top three drivers accounting for 28.7% of RI by mean across years.

As for Group 3 results (all years combined), the top three drivers included irrigated area with a RI of 21.7%, saturated thickness with 5.2%, and annual precipitation with 4.8%. Outside the large RI increase for annual precipitation, trends remained consistent between Groups 1 and 3. In addition to annual precipitation, irrigated area also had a notable increase in RI; combined, these increases resulted in the slight RI decrease for most other variables as all contributing variables sum to 100%. The overall complexity of variables influencing irrigation pumping is further highlighted through all five categories contributing to the top nine driving factors and nearly all drivers contributing less than 5% to the total influence on pumping.

The relative influences of each irrigation driver on groundwater pumping amount for the Group 2 models (annual, simplified, and normalized by irrigated area) and Group 4 model (years combined, unsimplified, and normalized by irrigated area) are displayed in Figure 4; each variable is further color coded by driver category.
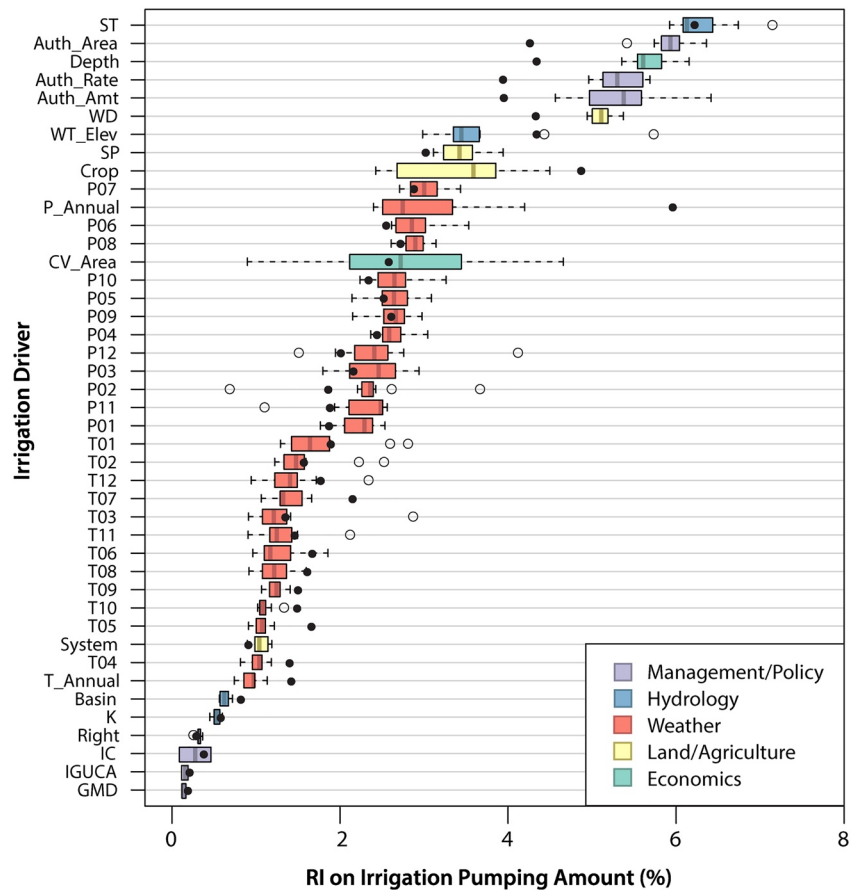
Figure 4. Boxplots showing RI outputs from BRT runs conducted annually from 2006 to 2016 on data normalized by irrigated area (Group 2). Black points show RI from a single BRT run conducted on collective data from 2006 to 2016 on data also normalized by irrigated area (Group 4). RI, relative influence; BRT, boosted regression trees.

Drivers are arranged by decreasing mean RI for Group 2 over the temporal range of the study, where whiskers on the Group 2 boxplots are likewise set at minimum and maximum values and the box is the IQR, and the hollow dots are outlier RI values. While the collective RI values still summed to 100%, normalizing pumping by irrigated area allowed for more detailed characterization of other less influential variables. Here, trends remained similar to those observed in Group 1, with the exception of relative shuffling of the top five variables. In Group 2, saturated thickness was the top influencing driver (mean RI of 6.3%) and authorized area was second (mean RI of 5.9%). Depth to water (cost to pump; mean RI of 5.7%) moved ahead of authorized rate (mean RI of 5.4%) when compared to Group 1. Crop type experienced increased variability across years, as did crop value per hectare.

Variables with the largest mean RI no longer observed the greatest variance in RI across years, although generally larger RI values resulted in greater variance. All driver variables still reported RI values of less than 10% in any given year, with most being less than 5% in all years. All five variable categories are represented in the top 10 influencing drivers, with the top three drivers accounting for 17.7% of RI by mean across years.

In Group 4 (all years combined, normalized by irrigated area), the top three influencing drivers were saturated thickness (6.2%), annual precipitation (6.0%), and crop type (4.9%). Similar to Group 3, annual precipitation had the greatest increase in RI, resulting in a decrease in RI for other top drivers. Outside the top 10 drivers, Group 4 results matched closely with the results of Group 2. All five categories were present in the top 10 variables, and all variables had RI values of less than 8% with most contributing to less than 5% of the total influence on groundwater pumping.
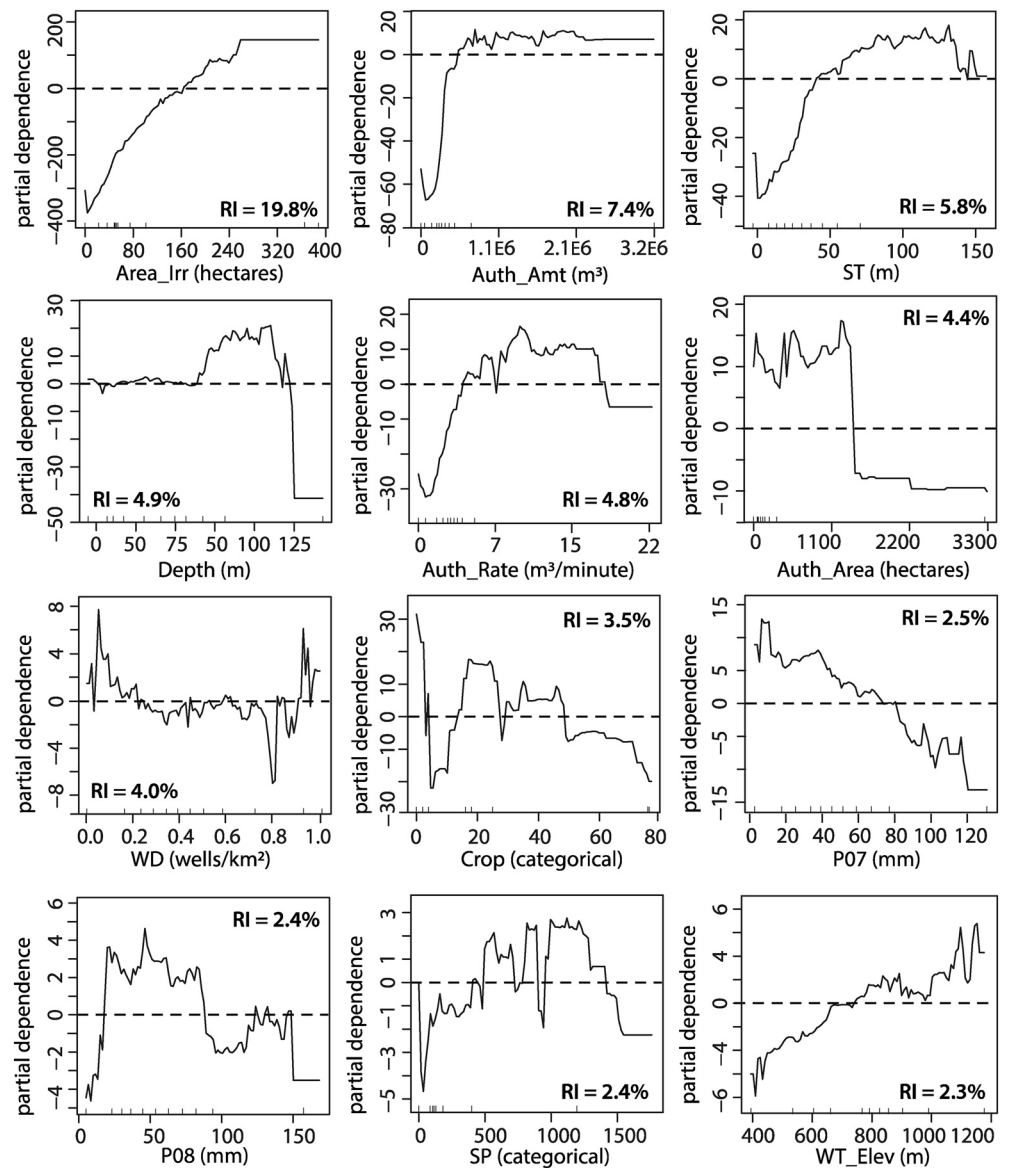
Figure 5. Partial dependence plots showing 12 influential variables in 2012 (Group 1), ordered by decreasing relative influence. Tick marks along the inside of the *x*-axes represent the distribution of data in deciles.

While RI is a meaningful metric, it fails to characterize whether a given variable contributes to inhibits pumping. This type of influence is displayed using partial dependence (PD) plots. Figure 5 displays PD plots for 12 example variables from Group 1 in 2012, where positive *y*-axis values indicate that corresponding *x*-axis values are more likely to predict irrigation pumping (stronger prediction). The reverse is also true, negative *y*-axis values indicate that corresponding *x*-axis values are less likely to predict irrigation pumping (weaker prediction). The magnitude of these values describes the strength of the correlation. These functions reveal the individual impact of the driver on irrigation pumping after the mean impact of the other drivers in the model is considered (Elith et al., 2008). Because of this, individual variables with strong interactions may not be well represented in these plots, but it remains a useful tool for understanding the general relationship between a predictor variable and the response variable (Friedman, 2001; Friedman & Meulman, 2003). Reference the Supplemental Information for examples regarding variable interactions.

Drivers are arranged in order of decreasing RI and can be summarized into three general patterns: (1) increasing PD slopes, (2) decreasing PD slopes, and (3) nonlinear PD slopes. Variables with increasing PD
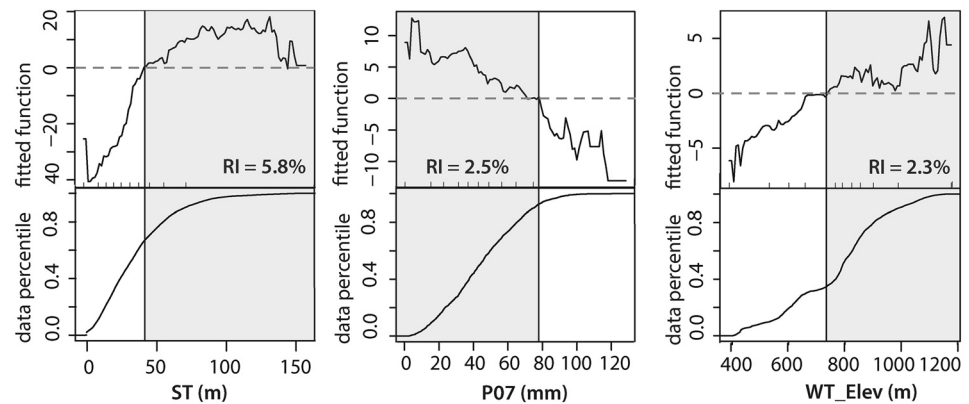
Figure 6. Partial dependence plots (top) with corresponding empirical cumulative distribution functions (bottom) to show influence of driver on irrigation pumping along driver data ranges.

slopes include irrigated area, authorized amount, saturated thickness, depth to water, authorized rate, and water table elevation. For example, as the number of irrigated areas increased, so did the occurrence of pumping, creating a positive PD slope. Variables with decreasing PD slopes included July, August, and October precipitation. Here, low precipitation values correlated with pumping activity and high precipitation values did not correlate with pumping activity, leading to a negative PD slope. Variables with nonlinear PD slopes included authorized area, well density, and soil permeability. These variables did not have a predictable PD pattern, showing both correlation with pumping and no pumping at variable points within the range of their data values. This relationship is also common in categorical data which do not have linear characteristics or behaviors to generate meaningful slope depictions.

Because outliers can seemingly skew the results of PD plots, decile marks are also included along the *x*-axis of each variable to communicate the magnitude of data represented in the plot. To further demonstrate how these data are interpreted, we plotted empirical cumulative distribution plots of driver data beneath three example PD plots for three variables in 2012 (Group 1), where shaded values have positive *y*-axis values, thus are correlated with pumping (Figure 6). These plots also identify the threshold at which pumping is likely to occur (or no longer occur). For example, pumping was likely to occur when saturated thickness was more than 41-m, July precipitation less than 77-mm, or water table elevation more than 737-m.

4. Discussion

4.1. Correlative Relationships

In this study, we seek to confirm correlations between predictor variables and irrigation pumping rather than infer causality. In discussion of results, we offer potential explanations for these correlations as we aim characterize what drives irrigation use in this region. Beyond identifying strong correlations, the BRT model also determines noisy variables. These variables have the lowest relative influence, indicating that they not only are uncorrelated with pumping but also do not *drive* pumping. Therefore, future water management efforts can be guided away from these low-influencing variables and redirected toward stronger drivers. This is a valuable deduction as a wide range of variables impact irrigation decision making and any reduction of noise helps target effective management.

Furthermore, some predictor variables in this study are endogenous and have potential to cause pumping, result from pumping, or likely both. For example, water table depth is highly correlated to pumping in this study which could be a cause of pumping and/or result from pumping. As this work focuses on irrigation drivers, we propose possible reasons why water table depth could drive pumping without claiming certain causality. Parsing causality from consequence becomes even more difficult with a variable like depth as it is temporally nonautonomous. In this way, pumping in a given year can be impacted by the change in water table depth from the previous year, which can then impact water table depth in the upcoming year. Depth is unlike autonomous variables such as precipitation, which are not linked to their impact in prior growing

seasons (i.e., the amount of rain in a given year is not largely impacted by the amount of rain in the previous year). Rather, depth can both be a driver and consequence of irrigation pumping. Even if this BRT model was conducted using predevelopment water table depths rather than current measurements, this unique relationship could not be fully captured.

4.2. Strong Versus Weak Drivers

Strong drivers were primarily those specific to the conditions of an *individual* well (e.g., total irrigated area, authorized rate, authorized amount, authorized area, and well density). Weak drivers were primarily those more representative of *regional* characteristics (e.g., groundwater doctrine, GMD, other localized management boundaries, river basin, or hydraulic conductivity of the aquifer). This difference between well-scale and regional influence can likely be attributed to individual farmers making decisions to irrigate on a case-by-case basis within the regulations of their larger-scale governing frameworks. For example, a water user more inclined to irrigate compared to a neighbor under the same regional governance will amplify well-specific drivers and buffer the influence of regional governance drivers, as regional drivers are the same between two disparate users. This relationship likely explains why well-specific governance (e.g., authorized area) is a strong driver, whereas regional governance (e.g., state groundwater doctrine; Right), which defines the well-specific governance, is a weak driver.

More directly, this relationship suggests that users are not limited or forced to change behavior by their regional governance and may be self-electing to change behavior despite their legal rights to business-as-usual. For example, GMDs 1, 3, and 4 have less groundwater supply but operate under "planned depletion" doctrine, and GMDs 2 and 5 have greater groundwater supply under the more conservative "safe yield" doctrine (Peck, 2006; Whittemore et al., 2018). But this does not mean all farmers in planned depletion zones are seeking to deplete the aquifer. This is especially true as GMD (and thus management scheme) was a weak driver. Several grassroots movements have been observed throughout the study region to prioritize water conservation given the limitation that regional governance does not adequately align conservation goals, and the impacts of these movements on reducing water usage may also be captured here; water management decisions have been found to correlate with close network members such as families, friends, and neighbors (Nian et al., 2020). Another contributing explanation as to why well-specific drivers are stronger than regional drivers may be that Local Enhanced Management Areas (LEMAs) introduced in some portions of the study area have enacted policies shown to reduce irrigation pumping within the frameworks of larger-scale governing regulations (Deines et al., 2019; Whittemore et al., 2018). However, LEMAs were not evaluated in this BRT model as the driver did not cover the full temporal range of the study.

Additionally, precipitation drivers were always stronger drivers to pumping than temperature drivers, and growing season precipitation was more influential than off-season precipitation. This is not only logically supported as adequate soil moisture is a critical metric to plant production (Basso & Ritchie, 2014), but this is also supported by the greater variability in seasonal values for precipitation compared to temperature. Irrigation is often used as a tool for overcoming or reducing the negative impacts of seasonal variability (Whittemore et al., 2016), where water applications can be used to buffer the impacts of drought conditions or extreme temperatures (Basso & Ritchie, 2014; Whittemore et al., 2016). As precipitation was more variable than temperature during this study, it is reasonable that it would be the stronger driver.

Irrigation technology, such as flood or center pivot, was not an influential variable despite the frequent discussion that higher-efficient systems lead to water savings (e.g., Schaible & Aillery, 2012). This may be due to nearly 76% of all wells being high-efficiency LEPA systems (KDA-DWR & KGS, 2021), in which case reduced variability within the driver has less influence on the pumping results. Also possible, this low influence is because efficient irrigation has been documented not to result in water savings (Pfeiffer & Lin, 2014), as farmers can irrigate more area for a reduced price compared to inefficient technologies. This behavioral response further supports total irrigated area being the predominant driver to pumping and the growing understanding that efficient technology does not reduce water use as long as there are incentives for farmers to irrigate more area (Pfeiffer & Lin, 2014; Smidt et al., 2016). Interestingly, few drivers in this study can be controlled directly by farmers, and those that can were among high-influencing variables (i.e., number of acres irrigated, crop type). Irrigation system type is the only farmer-controlled driver among low-influencing variables.

4.3. Partial Dependence Behaviors

Most of the top influencing drivers followed a positive slope (i.e., as driver values increase, so does the connection to pumping), with the exception of well density, crop type, and precipitation; well density and precipitation followed a negative slope, and crop type was categorical. While the partial dependence of these drivers was mostly as expected (e.g., dry conditions led to pumping), we did observe unexpected results in moderately influential drivers. For example, the results of depth to water were opposite what seemed intuitive. Instead of greater depths leading to a negative pumping slope (i.e., increased pumping costs leading to decreased pumping), we found greater depths to water led to more pumping. This is perhaps because pumping is so heavily established in these well locations that increased costs are a necessary, or unavoidable, operational factor or because decreased well yields may demand greater time spent for a center pivot to make its way around a field (Rad et al., 2020), potentially affecting the ratio between evaporation and infiltration. It may also be feasible that smaller farm behaviors, acting more in line with expected operational costs, are buffered in this model by larger farms with larger operational budgets. In addition, as water table depth changes, aquifer transmissivity can change nonlinearly, causing a ranging impact on wells in a region (Korus & Hensen, 2020). Another possible explanation is that variable is dominated by planned depletion management schemes (Peck, 2006), but this seems unlikely as most of the region is not under planned depletion strategies. Likewise, increased well density did not result in a positive pumping slope. Instead, fewer wells per area resulted in greater pumping. This may be in part because more wells per area can share the water demand of a larger area compared to an individual well, so total pumping per well can be reduced. However, this may be because fewer wells per area may be correlated to more individual landowners and subsequently different operational decisions. In this case, the connection between pumping and fewer wells per area may be attributed to smallholder farmers maximizing short-term profits or mitigating seasonal variability risk through increased pumping (Whittemore et al., 2016); these ownership data were not available specifically for each well and could not be evaluated in this study.

4.4. Irrigation and Climate

For Groups 1 and 2 (annual), the weather category was the last category listed among top variables, while it ranked as the first category for Groups 3 and 4 (all years combined). Considering that irrigation can be used as means of climate control (Whittemore et al., 2016), it was intriguing that weather variables were not more influential in Groups 1 and 2. This could result for two reasons: (1) the total influence is shared across many weather drivers, so a single driver is ultimately buffered: many weather drivers sum to have a large influence, but no single driver is largely influential, and (2) weather-related variables in the study region are sufficient for dryland agricultural production, as irrigation is used to capture incentives other than baseline production. However, it seems unlikely the weather conditions are sufficient for dryland production in western Kansas given regional water demands of the produced commodities and precipitation patterns (Cotterman et al., 2018). Even with the increase of drought-resistant cultivars (Hu & Xiong, 2014), the shared influence of weather drivers is a likely explanation for the lack of highly influential weather drivers in Groups 1 and 2. This makes logical sense as seasonal weather extremes are often short lived (U.S. Drought Monitor, 2020), not typically observed in repeated years with the same intensity (U.S. Drought Monitor, 2020), and are not always limiting to crop production as crops can partially rebound within a season. We also found that the RIs of weather-related variables were not noticeably higher during drought years within the study range (2011–2014; U.S. Drought Monitor, 2020). This may point to the practice of taking irrigated area out of production during abnormally dry conditions in order to meet higher irrigation demands of the remaining fields (Deines et al., 2017; Nie et al., 2018).

Furthermore, weather variables in Groups 1 and 2 may be relatively weaker drivers due to the spread of collective influence across many variables because seasonal extremes are combined into one variable in Groups 3 and 4. In these groups, annual precipitation became the third most influential driver on irrigation pumping. So, while annual precipitation may have less influence at the annual scale, its combined influence at longer time scales (multiyear) on regional pumping is significant. This is further supported in Groups 1 and 2 where annual weather-related variables were more influential than monthly, just as combined years were more influential than annual. Collectively, these relationships suggest that climate, acting at longer

Table 3
Relative Influence of Drivers by Category Across BRT Model Runs

	Relative influence of drivers on irrigation pumping (%)				
	Management/policy	Hydrology	Weather	Land/agriculture	Economics
Group 1	15.0	9.0	42.2	27.7	6.6
Group 2	17.5	11.2	50.5	12.9	8.5
Group 3	11.7	9.2	42.8	30.6	5.7
Group 4	13.3	12.0	54.7	13.1	6.9

Note. Groups 1 and 2 do not sum to 100 as the reported values represent means across 11 models where Groups 3 and 4 are single values reported across one model.

time scales than weather, is likely to play a significant role on the pumping patterns across the region. Consequently, climate change may reconfigure irrigation drivers in this region.

4.5. Management

This study highlights that irrigation decision making typically follows two questions: (1) how much is available to irrigate, both in water volume and land area (e.g., irrigated area, saturated thickness, authorized area, authorized rate, and authorized amount), and (2) how much does it cost (e.g., depth)? Other drivers or considerations appear to be marginal compared to the answers to these two questions. Observed trends further indicate irrigation is a default behavior and is *intensified* by weather conditions and not necessarily *a result of* weather conditions. So, while water use is an annual decision, compounding weather-related variables appear to shape behaviors at longer time scales. Collectively, this means water conservation strategies (even in planned depletion zones) would be better suited to focus on well-specific policies designed within the framework of these two questions while stabilizing water use incentives over longer time scales.

As each driver category was represented in the top 10 most influential drivers in each model group, policies must also be well rounded and account for the variations across categories through time, rather than emphasizing a suite of specific drivers. Water management is complex in this region and must be approached as such to avoid unintended water use outcomes. The summed totals of each category for each model group are reported in Table 3. The Weather category not only contributed the highest collective influence of all variable categories but also contributed the highest number of variables in the model (26 out of 45). The Economics category not only contributed the lowest collective influence of all categories but also had the lowest number of variables in the model (2 out of 45). The contribution of the Land/Agriculture category was about 2 times higher in Groups 1 and 3 than 2 and 4 due to the inclusion of total irrigated area in Groups 1 and 3.

5. Conclusion

Although many political, economic, and physical factors impact irrigation decision making in western Kansas and elsewhere, characterization of their relative influence on pumping has largely remained unknown. To quantify the influence of irrigation drivers, we utilized a BRT machine-learning technique on data across space and time to characterize the impact of 45 drivers relating to five categories (management/policy, hydrology, weather, land/agriculture, and economy) on irrigation pumping from approximately 19,000 wells across western Kansas from 2006 to 2016. BRT is a useful and informative tool for analyzing water use decision making and can effectively capture both numerical and categorical variable relationships across both space and time. In addition to total driver influence, BRT can also be used to understand the magnitude of influence as well the conditions in which a user typically decides to stop irrigating. In the future, this technique can also be used with other models to improve their irrigation prediction (e.g., agent-based irrigation models; Mewes & Schumann, 2019). From this study, we have identified four main conclusions:

(1) Influences on irrigation use in this region are complex, as all five variable categories were represented in the top 10 most influential variables under all modeling scenarios. In addition, the influence of many

drivers, like precipitation and crop value, varied from year to year. As a result of this complexity and variability, effective policy should focus on comprehensive, multifaceted measures rather than targeting individual, undesired behaviors.

(2) Well-specific drivers were considerably more influential to irrigation use than regional-specific drivers. This relationship suggests irrigation applications are a user-by-user decision not largely influenced by preexisting regulatory frameworks. Instead, water use decisions in this region are more a function of maximizing crop production across disparate and self-motivated water conservation strategies.

(3) Decisions to irrigate can largely be summarized in response to two questions: (1) how much is available to irrigate, both in water volume and land area (e.g., irrigated area, saturated thickness, authorized area, authorized rate, and authorized amount), and (2) how much does it cost (e.g., depth to water, well yield as a function of saturated thickness)? Other considerations contribute notably less to overall use.

(4) While influential in the short-term, weather-related factors have a greater influence at longer time scales due to varying impact at shorter time scales (e.g., seasonal compared to annual time scales, annual compared to multiannual time scales). This increased influence at longer time scales suggests irrigation use in this region may be susceptible to changes in irrigation patterns and behaviors under changing climate scenarios.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Model data are available at http://www.hydroshare.org/resource/c0b6ebc880f54c92b1c4a633b0b85353 (Smidt, 2020).

References

Basso, B., & Ritchie, J. T. (2014). Temperature and drought effects on maize yield. *Nature Climate Change*, *4*(233). 233. https://doi.org/10.1038/nclimate2139

Butler, J. J., Whittemore, D. O., Wilson, B. B., & Bohling, G. C. (2016). A new approach for assessing the future of aquifers supporting irrigated agriculture. *Geophysical Research Letters*, *43*, 2004–2010. https://doi.org/10.1002/2016GL067879

Cotterman, K. A., Kendall, A. D., Basso, B., & Hyndman, D. W. (2018). Groundwater depletion and climate change: Future prospects of crop production in the Central High Plains Aquifer. *Climatic Change*, *146*, 187–200. https://doi.org/10.1007/s10584-017-1947-7

Deines, J. M., Kendall, A. D., Butler, J. J., & Hyndman, D. W. (2019). Quantifying irrigation adaptation strategies in response to stakeholder-driven groundwater management in the US High Plains Aquifer. *Environmental Research Letters*, *14*, 044014. https://doi.org/10.1088/1748-9326/aafe39

Deines, J. M., Kendall, A. D., & Hyndman, D. W. (2017). Annual irrigation dynamics in the U.S. Northern High Plains derived from Landsat satellite data. *Geophysical Research Letters*, *44*, 9350–9360. https://doi.org/10.1002/2017GL074071

Edwards, E. C. (2016). What lies beneath? Aquifer heterogeneity and the economics of groundwater management. *Journal of the Association of Environmental and Resource Economists*, *3*, 453–491. https://doi.org/10.1086/685389

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Fenichel, E. P., Abbott, J. K., Bayham, J., Boone, W., Haacker, E. M. K., & Pfeiffer, L. (2015). Measuring the value of groundwater and other forms of natural capital. *Proceedings of the National Academy of Sciences*, *113*(9), 2382–2387. https://doi.org/10.1073/pnas.1513779113

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232.

Friedman, J. H., & Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, *22*(9), 1365–1381. https://doi.org/10.1002/sim.1501

Haacker, E. M. K., Cotterman, K. A., Smidt, S. J., Kendall, A. D., & Hyndman, D. W. (2019). Effects of management areas, drought, and commodity prices on groundwater decline patterns across the High Plains Aquifer. *Agricultural Water Management*, *218*, 259–273. https://doi.org/10.1016/j.agwat.2019.04.002

Haacker, E. M. K., Kendall, A. D., & Hyndman, D. W. (2016). Water level declines in the High Plains Aquifer: Predevelopment to resource senescence. *Groundwater*, *54*, 231–242. https://doi.org/10.1111/gwat.12350

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York, NY: Springer.

Hauenstein, S., Wood, S. N., & Dormann, C. F. (2017). Computing AIC for black-box models using generalized degrees of freedom: A comparison with cross-validation. *Communications in Statistics - Simulation and Computation*, *47*(5), 1382–1396. https://doi.org/10.1080/03610918.2017.1315728

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). *Package "dismo"*. CRAN. http://cran.nexr.com/web/packages/dismo/dismo.pdf

Hrozencik, R. A., Manning, D. T., Suter, J. F., Goemans, C., & Bailey, R. T. (2017). The heterogeneous impacts of groundwater management policies in the Republican River Basin of Colorado. *Water Resources Research*, *53*, 10757–10778. https://doi.org/10.1002/2017WR020927

Hu, H., & Xiong, L. (2014). Genetic engineering and breeding of drought-resistant crops. *Annual Review of Plant Biology*, *65*, 715–741. https://doi.org/10.1146/annurev-arplant-050213-040000

Hu, Y., Quinn, C. J., Cai, X., & Garfinkle, N. W. (2017). Combining human and machine intelligence to derive agents' behavioral rules for groundwater irrigation. *Advances in Water Resources*, *109*, 29–40. https://doi.org/10.1016/j.advwatres.2017.08.009

KDA-DWR & KGS, version 5. (2021). WIMAS, Water Information Management and Analysis System, 2006–2016. Kansas Geological Survey and the Kansas Department of Agriculture, Division of Water Resources. Retrieved from http://hercules.kgs.ku.edu/geohydro/wimas/index.cfm

Kisekka, I., Oker, T., Nguyen, G., Aguilar, J., & Rogers, D. (2017). Revisiting precision mobile drip irrigation under limited water. *Irrigation Science*, *35*, 483–500. https://doi.org/10.1007/s00271-017-0555-7

Korus, J. T., & Hensen, H. J. (2020). Depletion percentage and nonlinear transmissivity as design criteria for groundwater-level observation networks. *Environmental Earth Sciences*, *79*, 382. https://doi.org/10.1007/s12665-020-09123-y

K.S.A. 82a-1036. (1978). Initiation of proceedings for designation of intensive groundwater use control areas; duties of chief engineer; findings. Retrieved from https://www.ksrevisor.org/statutes/chapters/ch82a/082a_010_0036.html

K.S.A. 82a-1041. (2012). Local enhanced management areas; establishment procedures; duties of chief engineer; hearing; notice; orders; review. Retrieved from https://www.ksrevisor.org/statutes/chapters/ch82a/082a_010_0041.html

Lampa, E., Lind, L., Lind, M., & Bornefalk-Hermansson, A. (2014). The identification of complex interactions in epidemiology and toxicology: A simulation study of boosted regression trees. *Environmental Health*, *13*(57). https://doi.org/10.1186/1476-069X-13-57

Majumdar, S., Smith, R., Butler, J., Jr., & Lakshmi, V. (2020). Groundwater withdrawal prediction using integrated multitemporal remote sensing data sets and machine learning. *Water Resources Research*, *56*, e2020WR028059. https://doi.org/10.1029/2020WR028059

MardanDoost, B., Brookfield, A. E., Feddema, J., Sturm, B., Kastens, J., Peterson, D., & Bishop, C. (2019). Estimating irrigation demand with geospatial and in-situ data: Application to the High Plains Aquifer, Kansas, USA. *Agricultural Water Management*, *223*, 105675. https://doi.org/10.1016/j.agwat.2019.06.010

Mewes, B., & Schumann, A. (2019). The potential of combined machine learning and agent-based models in water resources management. *Hydrologie und Wasserbewirtschaftung*, *63*, 332–338. https://doi.org/10.5675/hywa_2019.6_2

Murphy, K. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA; London, UK: MIT Press.

Nian, Y., Huang, Q., Kovacs, K. F., Henry, C., & Krutz, J. (2020). Water management practices: Use patterns, related factors, and correlations with irrigated acres. *Water Resources Research*, *56*, e2019WR025360. https://doi.org/10.1029/2019WR025360

Nie, W., Zaitchik, B. F., Rodell, M., Kumar, S. V., Anderson, M. C., & Hain, C. (2018). Groundwater withdrawals under drought: Reconciling GRACE and land surface models in the United States High Plains Aquifer. *Water Resources Research*, *54*, 5282–5299. https://doi.org/10.1029/2017WR022178

Peck, J. C. (2006). Groundwater management in Kansas: Brief history and assessment. *The Kansas Journal of Law & Public Policy*, *15*(3), 441–466

Pfeiffer, L., & Lin, C.-Y. C. (2014). Does efficient irrigation technology lead to reduced groundwater extraction? Empirical evidence. *Journal of Environmental Economics and Management*, *67*, 189–208. https://doi.org/10.1016/j.jeem.2013.12.002

Rad, M. R., Brozović, N., Foster, T., & Mieno, T. (2020). Effects of instantaneous groundwater availability on irrigated agriculture and implications for aquifer management. *Resource and Energy Economics*, *59*, 101129. https://doi.org/10.1016/j.reseneeco.2019.101129

R Development Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org

Ridgeway, G. (2007). *Generalized boosted models: A guide to the gbm package*. CRAN. https://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf

Rudnick, D. R., Irmak, S., West, C., Chávez, J. L., Kisekka, I., Marek, T. H., et al. (2019). Deficit irrigation management of maize in the High Plains Aquifer region: A review. *Journal of the American Water Resources Association*, *55*(1), 38–55. https://doi.org/10.1111/1752-1688.12723

Sanderson, M. R., Bergtold, J. S., Heier Stamm, J. L., Caldas, M. M., & Ramsey, S. M. (2017). Bringing the "social" into sociohydrology: Conservation policy support in the Central Great Plains of Kansas, USA. *Water Resources Research*, *53*, 6725–6743. https://doi.org/10.1002/2017WR020659

Sanderson, M. R., & Frey, R. S. (2014). From desert to breadbasket... to desert again? A metabolic rift in the High Plains Aquifer. *Journal of Political Ecology*, *21*, 516. https://doi.org/10.2458/v21i1.21149

Schaible, G., & Aillery, M. (2012). Water conservation in irrigated agriculture: Trends and challenges in the face of emerging demands. *USDA-ERS Economic Information Bulletin*, *99*. https://doi.org/10.2139/ssrn.2186555

Smidt, S. J. (2020). Irrigation drivers boosted regression; Kansas HPA. *HydroShare*. http://www.hydroshare.org/resource/c0b6ebc880f54c92b1c4a633b0b85353

Smidt, S. J., Haacker, E. M. K., Kendall, A. D., Deines, J. M., Pei, L., Cotterman, K. A., et al. (2016). Complex water management in modern agriculture: Trends in the water–energy–food nexus over the High Plains Aquifer. *The Science of the Total Environment*, *566–567*(567), 988–1001. https://doi.org/10.1016/j.scitotenv.2016.05.127

Smidt, S., Kendall, A., & Hyndman, D. (2019). Increased dependence on irrigated crop production across the CONUS (1945–2015). *Water*, *11*, 1458. https://doi.org/10.3390/w11071458

Steward, D. R., Bruss, P. J., Yang, X., Staggenborg, S. A., Welch, S. M., & Apley, M. D. (2013). Tapping unsustainable groundwater stores for agricultural production in the High Plains Aquifer of Kansas, projections to 2110. *Proceedings of the National Academy of Sciences*, *110*, E3477–E3486. https://doi.org/10.1073/pnas.1220351110

Sukcharoen, K., Golden, B., Vestal, M., & Guerrero, B. (2020). Do crop price expectations matter? An analysis of groundwater pumping decisions in Western Kansas. *Agricultural Water Management*, *231*, 106021. https://doi.org/10.1016/j.agwat.2020.106021

USDA-CDL. (2006–2016). *National Agricultural Statistics Service Cropland Data Layer*. Retrieved from https://nassgeodata.gmu.edu/CropScape

USDA-NASS. (2006–2016). *Crop values annual summary*. Retrieved from https://usda.library.cornell.edu/concern/publications/k35694332

U.S. Drought Monitor. (2020). National Drought Mitigation Center University of Nebraska-Lincoln, United States Department of Agriculture, National Oceanic and Atmospheric Administration. Retrieved from https://droughtmonitor.unl.edu/About.aspx

Whittemore, D. O., Butler, J. J., & Wilson, B. B. (2016). Assessing the major drivers of water-level declines: New insights into the future of heavily stressed aquifers. *Hydrological Sciences Journal*, *61*, 134–145. https://doi.org/10.1080/02626667.2014.959958

Whittemore, D. O., Butler, J. J., & Wilson, B. B. (2018). Status of the High Plains Aquifer in Kansas. *Kansas Geological Survey Technical Series*, *22*, 1–28.