ASSESSING RELATIONSHIPS BETWEEN PATTERNS OF ONSITE WASTEWATER TREATMENT SYSTEM MAINTENANCE AND ENVIRONMENTAL VARIABLES IN ATHENS-CLARKE COUNTY, GEORGIA

by

KYLE NEIL CONNELLY

(Under the Direction of Krista Capps and Seth Wenger)

ABSTRACT

Millions of households rely on onsite wastewater treatment systems (OWTSs) to safely treat domestic wastewater; however, OWTSs can become environmental hazards when they are not sited properly or maintained. To better understand the drivers of OWTS maintenance and failure, we examined relationships between site-level OWTS characteristics and system repair, pumping, and anomalous pumping records in Athens-Clarke County, Georgia, USA. We found that the oldest OWTSs (> 50 years) had the highest probabilities of being repaired and exhibiting signs of hydraulic failure. Notably, newer OWTSs (2-10 years) were approximately equally as likely as older systems to exhibit signs of hydraulic failure. These findings suggest that repair and replacement efforts should emphasize older systems that are at or near the end of their serviceable life, and the hydraulic performance of both newer and older OWTSs should be monitored. These insights can aid decision makers in equitably prioritizing wastewater infrastructure investments and policies.

INDEX WORDS: Onsite Wastewater Treatment Systems, Decentralized Wastewater Infrastructure, Risk Management, Environmental Layers

ASSESSING RELATIONSHIPS BETWEEN PATTERNS OF ONSITE WASTEWATER TREATMENT SYSTEM MAINTENANCE AND ENVIRONMENTAL VARIABLES IN ATHENS-CLARKE COUNTY, GEORGIA

by

KYLE NEIL CONNELLY

BS, West Virginia University, 2016

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2021

ASSESSING RELATIONSHIPS BETWEEN PATTERNS OF ONSITE WASTEWATER TREATMENT SYSTEM MAINTENANCE AND ENVIRONMENTAL VARIABLES IN ATHENS-CLARKE COUNTY, GEORGIA

by

KYLE NEIL CONNELLY

Major Professors: Krista Capps
Seth Wenger
Committee: Amy Rosemond

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2021

DEDICATION

To my father: an environmentalist, leader, and inspiration.

ACKNOWLEDGEMENTS

I am tremendously thankful for the support, flexibility, and critiques provided by my committee: Drs. Krista Capps, Seth Wenger, and Amy Rosemond. This study would not have been achievable without the insights and expertise they provided.

A debt of gratitude is also owed to the hard-working practitioners of the Athens-Clarke County Unified Government. The technical knowledge and time provided by Mike Occhipinti in support of this research was irreplaceable and his commitment to his work and this field of study is extraordinary. The guidance and foresight of Dave Bloyer and Cheryl Shaw in compiling the data used in this study was also instrumental.

I am also thankful to Drs. Nandita Gaur and Jacob McDonald for sharing their specialized knowledge to help formulate the scope and methods of this study.

Finally, I would like to acknowledge the unwavering support of my friends, family, and beloved partner, Sarah. Without you, none of this would have been possible.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

ASSESSING RELATIONSHIPS BETWEEN PATTERNS OF ONSITE WASTEWATER TREATMENT SYSTEM MAINTENANCE AND ENVIRONMENTAL VARIABLES IN ATHENS-CLARKE COUNTY, GEORGIA

Introduction

Approximately one-fifth of American homes rely on onsite wastewater treatment systems (OWTSs) to safely dispose of their sewage.[1,2] In total, these aging systems process approximately 4 billion gallons of wastewater per day in the United States (US).[3] Conventional subsurface OWTSs are the most common onsite wastewater treatment option, as they offer an affordable alternative to central sewer system expansion[4] and thus tend to service rural and suburban residential areas beyond the extent of sewer networks. When properly designed, installed, and maintained, OWTSs can effectively remove pathogenic[5,6] and nutrient[7,8] pollution from wastewater. However, these systems can also become sources of nonpoint pollution in watersheds where they are malfunctioning or maintained improperly.[9–11]

Installation and site-specific factors can affect OWTS performance and lead to increased risk of hydraulic failure (i.e., surface ponding, system backup). Characteristics including edaphic and geologic conditions, water table dynamics, landscape position and slope, and lot area influence OWTS system function.[12,13] Conventional subsurface OWTSs have three primary components that are integral to their performance: (i) the septic tank, which collects and retains a large portion of the influent biosolids and grease, (ii) the subsurface wastewater infiltration system (also called a drainfield, infiltration trench, or leach field), which transmits and distributes wastewater effluent from the septic tank to the receiving soils, and (iii) the receiving soils, which absorb water and process contaminants.[12,14] Drainfields rely on spatially heterogeneous soil biogeochemical processes to treat OWTS effluent, but the US Environmental Protection Agency (US EPA) estimates that only one-third of the nation's land area has soils suitable for conventional OWTSs.[14] Therefore, it is likely that a significant portion of OWTSs are not functioning effectively due to improper siting and subsequent performance issues (e.g. ref 15). Three well-recognized and potentially interrelated subsurface site characteristics that may hydrologically impede drainfield effectiveness are the receiving soil's hydraulic conductivity, the presence of low-conductivity or impermeable restrictive layers, and depth to the water table.[12,16,17] Additionally, other topographic site characteristics that control drainage patterns such as slope, curvature (i.e., convexity versus concavity), and upslope contributing area may influence OWTS performance at the site level.[14,17–19] Because of the potential for partially treated wastewater to leach into ground or surface waters, the siting, design, and maintenance of OWTSs has significant implications for community health and drinking water supplies.[16,20,21]

Regular maintenance, including desludging (or pumping) in which biosolids are removed from the OWTS tank, is critical to ensure these systems function properly. The US EPA recommends OWTS owners get their systems inspected every three years and pumped either every three to five years or when biosolids in the OWTS tank exceed 30% of its volume.[14,22] However, these guidelines are not often followed because the costs of OWTS maintenance can be substantial.[23,24] Unless system failure leads to sewage backflow into the house or ponding within the drainfield, improperly functioning systems may go unnoticed. The diffuse nature of pollution from failing OWTSs and a general lack of understanding about subsurface flow within drainfields[25,26] means it is unlikely that any single OWTS owner could be held accountable for downstream water degradation; therefore, the environmental costs of these malfunctioning systems are borne by society as a whole, rather than individual owners.[27,28] Furthermore, programs that replace failing systems (e.g. ref 29) act to subsidize the costs of OWTS repairs or replacement, but jurisdictions can face substantial resistance from homeowners when trying to administer these programs.[30] Such initiatives also retroactively address OWTS failure instead of supporting proactive planning to prevent future contamination threats. Finally, and remarkably, some homeowners may not even be aware that they have an OWTS,[15] and thus homeowner action, including system inspections and maintenance, would not occur unless hydraulic failure ensues.

In general, OWTS infrastructure in the US is aging[31] and failures are increasingly common as system components deteriorate. Studies of drainfield effluent acceptance rates suggest that conventional subsurface OWTSs can remain hydraulically operational for 11 to more than 30 years,[16] but approximately half of OWTSs 19 to 27 years old can be expected to show evidence of failure.[32,33] Because OWTS installations tend to occur in clusters when housing developments are built, failures also may be clustered in space and time. Though realized OWTS lifespans are dependent upon factors such as maintenance and effluent loading rates, it is also likely that site characteristics contribute to OWTS service life. It remains unclear how interactions among siting, installation, maintenance, and use characteristics determine the effective lifespan of an OWTS.

Due to their decentralized nature and the general lack of available data for privately-owned OWTSs,[34] research has generally focused on the environmental effects of OWTSs above density thresholds at the watershed scale (e.g. refs 35–37). There are far fewer studies on the potential impact of an individual faulty or ill-maintained OWTS on water quality. Yet many state[38,39] and federal[14,40] guidelines highlight the potential impact poorly-maintained systems may have on surface and groundwater resources. Work by Macintosh and colleagues[10] in three rural catchments in north Ireland found that replacing specific malfunctioning OWTSs could result in reduced phosphorus concentrations downstream. Identifying the site-level characteristics of OWTSs that are at higher risk of failure is especially valuable to local resource managers who are tasked with ensuring environmental and community health.

To better understand the drivers of OWTS maintenance and failure, we explored relationships between OWTS age, environmental setting, and repair and pumping records for OWTSs in Athens-Clarke County, Georgia, USA. We collaborated with the local government to leverage a 46-year OWTS repair dataset (May 1972 – March 2018), and a 38-month (January 2017 – February 2020) dataset of OWTS tank pumping within the county. Our overarching goal was to use these records to identify where and when OWTSs fail to support municipal efforts to equitably prioritize investments in wastewater treatment infrastructure. We identified OWTSs that exhibited pumping patterns that are indicative of failing systems (hereafter “anomalous pumping”) based on whether (i) the OWTS had been pumped twice or more, thereby exceeding the US EPA’s guideline of one pump per three to five years, and/or (ii) the volume pumped from the system exceeded the tank capacity listed in county installation records. First, we tested the ability of OWTS age and site characteristics to explain variability in repair, pumping, and anomalous pumping. Using best-fit regression models, we developed probability curves to estimate the likelihood of OWTSs to be repaired or to exhibit anomalous pumping. Secondly, we assessed relationships between systems that had repair records and those that exhibited anomalous pumping to test whether anomalous pumping may be a correlate of necessary repairs.

We also calculated the proportion of OWTSs that were pumped and then subsequently repaired within three years, which could indicate that the homeowner knew that their system might need additional maintenance.

Data and Methods

Athens-Clarke County (ACC) is located in the Appalachian Piedmont region of Georgia. With a land area of approximately 116 square miles, ACC is Georgia's smallest county by area and the state's sixth most populous metropolitan area, with an estimated 126 913 residents.[41] Approximately 75% of ACC residents are serviced by the central sewer system, with approximately 32 000 residents dependent on OWTSs to dispose of and treat their wastewater.[42] To guide wastewater infrastructure decisions, ACC has invested substantial resources into developing and maintaining a county-wide geospatial dataset of all OWTSs, their ages, volumetric tank capacities, and repair and desludging histories.[43]

The county provided spatial and tabular data consisting of all registered OWTSs in the study area (n = 9802), from which we filtered out occurrences of missing data, abandoned systems, those with accountable damage (e.g., house fires, windthrown trees), and records with uncertain geolocations (Figure 1). Methods similar to Capps et al.[43] were used to verify the locational accuracy of a subset of the remaining OWTSs. Installation dates of OWTSs in this dataset with high locational accuracy (n = 8826) ranged from January 1940 – March 2018. With these cleaned data, we filtered out points with missing predictor values, which were mostly OWTSs on the periphery of the study area (refined dataset n = 8786; Figure 2). Subsequently, we calculated system age in years as the number of years between system installation and October 28, 2020 (the date of data retrieval). Repair records for OWTSs (May 1972 – March

2018), which were submitted to ACC's Department of Public Health when repair work was conducted on any OWTS in the jurisdiction, were used to calculate a binary response of whether each OWTS was repaired (n = 690; Figure 3).

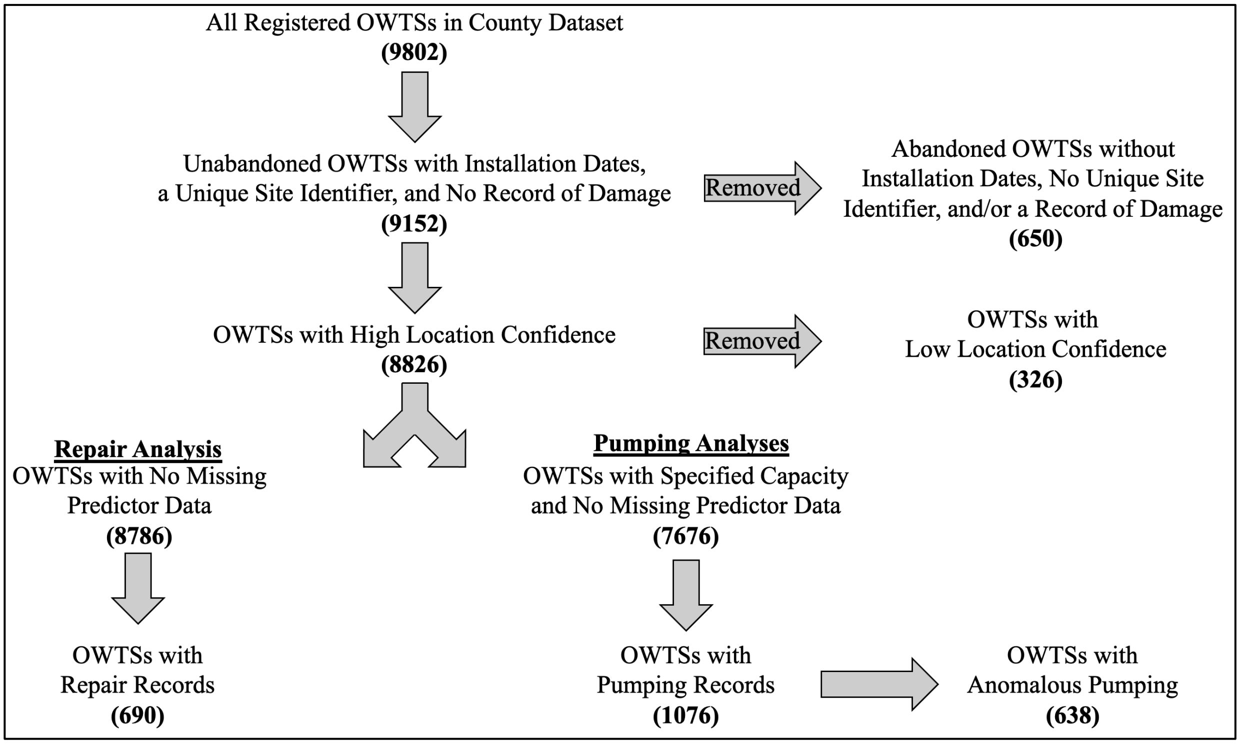


Figure 1. Conceptual diagram of processes used to filter, refine, and divide the dataset for the repair and pumping analyses. *Date of installations for the cleaned dataset with high location confidence ranged from January 1940 – March 2018. OWTS repair records (May 1972 – March 2018) were submitted to the Athens-Clarke County Department of Public Health when repair work was conducted at the location. Pumping records (January 2017 – February 2020) were collated from pumping manifests submitted to the Public Utilities Department when pumping companies disposed of septage at the Cedar Creek Water Reclamation Facility. Data included in the repair and pumping analyses are not exclusive, as all data points in the pumping analyses are a subset of data in the repair analysis, with the additional restriction of requiring a specified OWTS tank capacity. The corresponding number of records in each dataset or removal are indicated by the bold number in parentheses.*

To examine OWTS pumping behaviors, the point locations and attributes of the cleaned county dataset were filtered to remove records without specified septic tank capacities and those with missing predictor values (refined dataset n = 7676). This subset of OWTSs was combined with OWTS pumping information from more than 3000 OWTS tank pumping manifests (January

2017 – February 2020), which were recorded when septic pumping trucks unloaded biosolids at ACC's Cedar Creek Water Reclamation Facility. The scale used to estimate septic truck load in gallons at Cedar Creek was intermittently out of service during the period of record. During these times, discharged volumes were estimated by the OWTS pumpers, though as the pumping companies are charged by gallon,[44] it may have been beneficial for them to report lower, more conservative, volumes. Thus, some of these values may have been underestimated. There were also instances of pumping events at addresses that were not yet entered into the county's OWTS location dataset. These manifest records from addresses without registered OWTS point locations may be the result of illegally installed systems, exceptionally old units that predate record-keeping, missing county records, or incorrectly entered addresses. As registered locations could not be confirmed for these records, these pumping events were not included in the pumping analysis of this study. A total of 1605 pumping events were associated with points in the refined dataset. Address, pumped volume, and date from each manifest were associated with the OWTS locations in the refined dataset and a binary response of pumped/not pumped was determined for each of the 7676 tanks in the refined dataset.

The Cedar Creek Water Reclamation Facility in ACC charges relatively low prices[45] for septage disposal and is one of the only treatment plants in a multicounty area that accepts septage.[46] Hence, we assumed that if an OWTS tank in ACC was pumped in this 38-month period, there would be a record of the pumping event in the county's manifests. We also assumed that all OWTSs were maintained legally. In other words, we have no records for homes that straight pipe raw sewage off their property or for pumpers who illegally dispose of their septage. To identify systems with anomalous pumping, we subset the total population of OWTSs with manifest records based upon binary arguments of whether (i) the OWTS had been pumped twice or more, thereby exceeding the US EPA's guideline of one pump per three to five years, and/or (ii) the pumped volume recorded on the manifest exceeded the tank capacity listed in county installation records. We assumed these instances of volume exceedance could be indicative of hydraulic drainfield failure and subsequent backup of effluent into the OWTS tank.

To characterize OWTS site conditions, we used widely available elevation, soil, and stream channel data. We used 30-meter resolution digital elevation model data from the US Geological Survey[47] to calculate the topographic wetness index in the study area. The topographic wetness index (Equation 1) is a unitless approximation of soil moisture as a function of local topography and contributing area[48] and has been used to assess pluvial flooding risk in land use planning.[49] We used the Terrain Analysis Using Digital Elevation Models (TauDEM) ArcGIS toolbox[50] to calculate slope, flow direction and contributing area using the D-infinity method (we added a value of 1.0e-6 to slopes with 0 to prevent division by zero).

$$W = \ln\left(\frac{\alpha}{\tan\beta}\right)$$

Equation 1. Topographic wetness index as defined by Beven and Kirby.[48] *W is the wetness index, α is the upslope contributing area divided by the flow width, and tanβ is the steepest local slope.*

We used Esri's ArcMap desktop application[51] for spatial processing and the R statistical environment[52] for statistical analysis. At the time of this study, we did not have access to county soil assessments, which are used to determine OWTS drainfield suitability at the time of installation. Therefore, we retrieved Gridded Soil Survey Geographic Database (gSSURGO) data for the study area using the Natural Resources Conservation Service's Soil Data Development Toolbox plugin for ArcMap.[53,54] The soil data used to permit OWTSs are based on high resolution soil maps or in-situ observations, while gSSURGO is an interpolated dataset that lacks onsite observations in many areas. Although soil values in this nationally available dataset are likely correlated with local conditions, gSSURGO data can differ slightly from the higher resolution data that OWTS permitters often have access to. Gridded soil values for saturated hydraulic conductivity were calculated using a weighted average of soil conditions between 30-183cm below the surface, as Georgia Department of Public Health regulations necessitate drainfield pipe installations to occur 15-30cm (6-12in) below the soil surface.[39] Gridded values for average depth to seasonal water table were calculated using the gSSURGO monthly (January-December) water table depth estimates. Values for depth to restrictive layer were calculated across the study area using the gSSURGO database. A raster layer of distance to stream was calculated using stream channel data from the National Hydrography Dataset Plus (NHD Plus) High Resolution dataset[55] to estimate each OWTS's potential exposure to fluvial flooding. A summary of predictor variables is found in Table 1.

To test whether site-level environmental conditions could predict OWTS maintenance occurrence, we developed and subsequently tested candidate and null models for each binary response variable (repaired/no repairs, pumped/no pumps, and anomalous pumping/no anomalous pumping) using logistic regression. We defined models to test combinations of the following variables: OWTS age, topographic wetness index, distance to the nearest stream, and three soil variables: saturated hydraulic conductivity, depth to restrictive layer, and depth to seasonal water table (Table 2). We tested a quadratic term for age because we hypothesized that younger OWTSs may exhibit signs of hydraulic failure due to installation or usage error at similar rates to older systems, which may be expected to fail due to system deterioration and soil clogging. A quadratic age term would allow the model to detect higher incidences of maintenance at either end of the dataset's age distribution, while moderately aged systems might be expected to have lower maintenance rates when compared to these temporal extremes. We only tested the three soil variables in combination because we believe these predictors together contribute to documented OWTS repair and pumping patterns and that no single soil condition is likely driving these patterns. Models were assessed and weighted using the Bayesian information criterion (BIC), due to the tendency of this method to favor parsimony in large datasets.[56] Top performing models were verified for discrimination performance using the area under the receiver operating characteristic curve (AUC) calculation. Semivariogram plots and residual maps were used to assess spatial autocorrelation and residual clustering in candidate models with the highest BIC weights.

We also evaluated how anomalous pumping in OWTSs might be indicative of necessary repairs using linear regressions between recent repairs and anomalous pumping. To do this, we restricted the date range of the repair dataset to five years prior to the earliest pumping record (January 2012) and assessed this single model's performance based on its p-value. We also identified the number OWTSs with pumping events that were subsequently followed by a repair event. Due to the short temporal overlap of the pumping and repair datasets (January 2017 – March 2018) we did not specify a date lag between pumping and repair event (i.e., all OWTSs that met these conditions were included).

Table 1. Predictor variables included in candidate models for the repair, pumping, and anomalous pumping analyses and variable abbreviations used in Tables 2 and 4.

Predictor Category	Predictor Variable	Abbreviation
Age	Years since OWTS installation	AGE
	Quadratic age of OWTS	AGE^2
Environmental	Topographic wetness index	TWI
	Distance to NHD Plus stream	DIST.S
Soils	Saturated hydraulic conductivity	KSAT
	Depth to restrictive layer	DEP.R
	Depth to seasonal water table	DEP.WT

__Table 2. Response variables and candidate models for the repair, pumping, and anomalous pumping analyses.__ The presence of anomalous pumping was also used as a predictor of recent repairs. "X" indicates which candidate models were tested to predict each of the three response variables. Abbreviations as in Table 1.

Response			Candidate Models
Repair	Pumping	Anomalous Pumping	
X	X	X	AGE
X		X	AGE + AGE^2
X	X	X	DIST.S
X	X	X	TWI
X	X	X	AGE + DIST.S
X	X	X	AGE + TWI
X	X	X	AGE + DIST.S + TWI
X	X	X	AGE + KSAT + DEP.R + DEP.WT
X	X	X	AGE + KSAT + DEP.R + DEP.WT + TWI
X*			Anomalous Pumping

*Repairs in this analysis were restricted to five years prior to the earliest pumping record.
AGE = Years since OWTS installation; AGE^2 = Quadratic age of OWTS; DIST.S = Distance to NHD Plus stream; TWI = Topographic wetness index; KSAT = Saturated hydraulic conductivity; DEP.R = Depth to restrictive layer; DEP.WT = Depth to seasonal water table.

Results

The installation dates for OWTSs with high location confidence (n = 8826) ranged from January 1940 – March 2018 and the median age of OWTSs in the county was 35 years. The median age of OWTSs in the repair dataset (n = 8786) was also 35 years, and the median age of repaired OWTSs was 65 years. The median age of OWTSs in the pumping dataset (i.e., those with s high location confidence and specified capacities; n = 7676) was slightly younger, at 33 years, and the median age of a pumped system was 34 years. On average, approximately 0.5% of the 8826 OWTSs in the study area were pumped per month over the 38-month period of pumping records. December of 2018 was the month with the highest percent of pumped systems (0.8%), and September of 2019 had the lowest single-month proportion of pumped systems during the study period (0.2%). Additional summary statistics for OWTSs included in each of the analyses are detailed in Table 3.

Table 3. Summary statistics for OWTSs in Athens-Clarke County, Georgia. *The size of each respective dataset can be found in Figure 1.*

Summary Statistic	Value	Units
Median age of OWTSs in county	35	years
Median age of OWTSs in repair dataset	35	years
Median age of repaired OWTSs	65	years
Percent of OWTSs repaired†	7.8	percent
Median age of OWTSs in pumping dataset	33	years
Median age of pumped OWTSs	34	years
Percent of OWTSs pumped†	12.2	percent
Average volume of an individual pump	1202	gallons
Percent of systems pumped annually*†	5.7	percent
Percent of pumping events which resulted in volume exceedances†	43.4	percent
Percent of OWTSs which had volume exceedances†	6.5	percent
Percent of OWTSs which were pumped more than once†	2.5	percent
Total volume of septage pumped in 38 months	1 933 307	gallons
Average monthly volume of septage pumped in the county	57 196.5	gallons

† Percent of OWTSs out of the 8826 OWTSs with high location confidence (see Figure 1)
*Percent calculated using monthly averages over a 38-month period of pumping records, multiplied by 12

Repair records existed for 690, or 7.8%, of the 8826 OWTSs with high location confidence. Though candidate models that included distance to the nearest stream, topographic wetness index, and the three soil variables had low BIC weights, we found that OWTS age alone was a good predictor of whether a system was repaired (Table 4). This model identified incidents of OWTS repairs slightly better than a random model (AUC = 0.589) and no spatial autocorrelation was present in this model's residuals (Appendix A).

There were 1605 pumping manifests from 1076 OWTS tanks during the period of record, meaning 12.2% of the 8826 OWTSs were pumped in the 38-month period of record. Of these pumping manifests, 697 of the 1605 manifests had reported volumes greater or equal to the OWTS tank from which they came. We identified anomalous pumping for 638 OWTSs of the 1076 with manifest records, 218 of which were pumped two or more times per three years. The most frequently pumped OWTS was desludged 40 times. Additionally, 576 out of the 1076 OWTSs with pumping records had pumping events where OWTS tank pumped volume was greater or equal to the tank's total capacity (Figure 4).

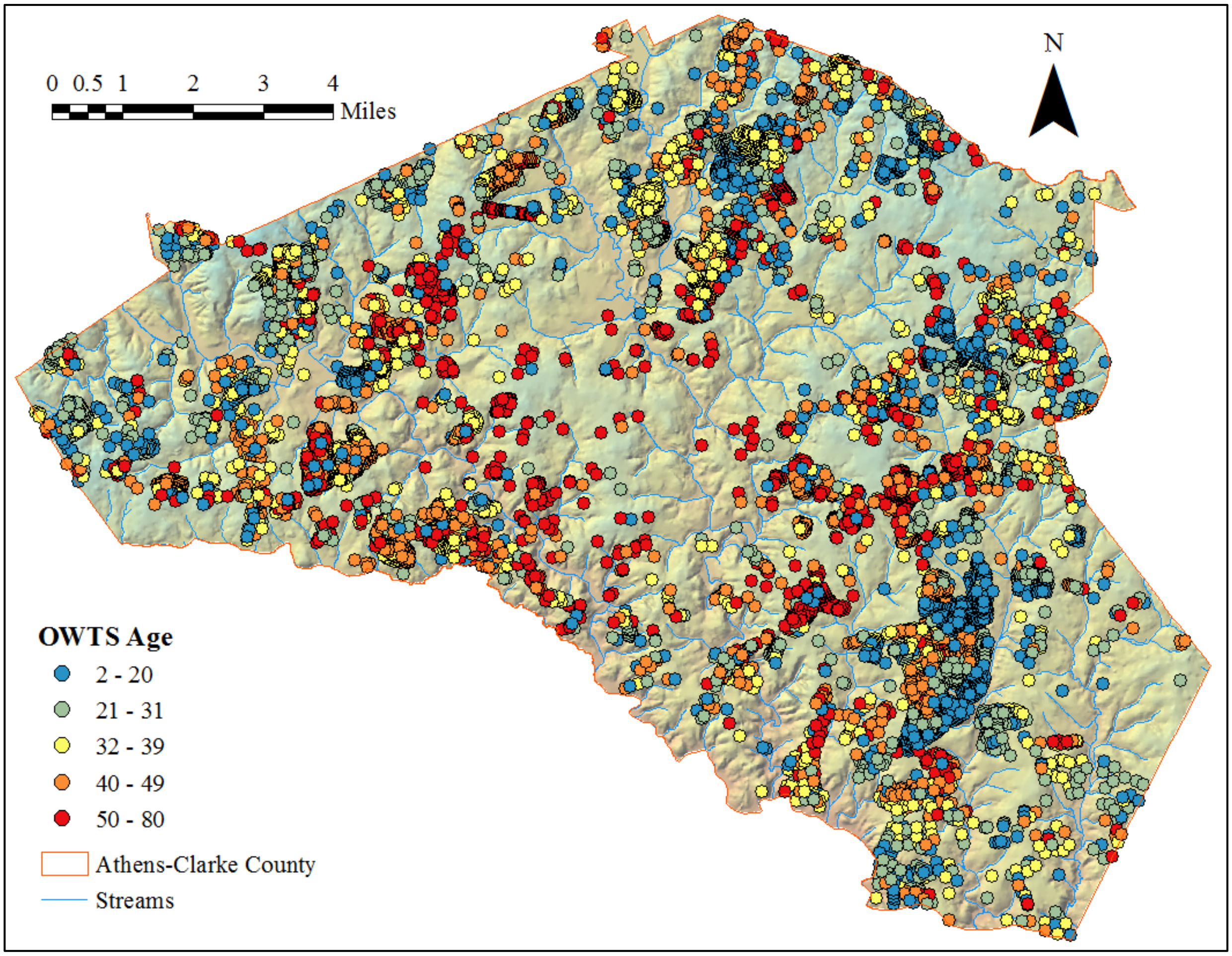


Figure 2. Age (years since OWTS installation) and spatial distribution of Athens-Clarke County OWTSs included in the repair analysis (n = 8786). *County installation records range from January 1940 – March 2018 and data were retrieved in October 2020. The US EPA suggests a serviceable OWTS life of approximately 15-40 years depending on siting and maintenance conditions.*[57] *The dataset of all registered OWTS locations provided by the county were filtered for quality control and assurance (see Figure 1). A shaded relief layer and stream networks were included to emphasize the spatial heterogeneity of these systems.*

Table 4. Top performing models per response, as assessed by their Bayesian information criterion (BIC) weight. *The model for predicting system repairs predicted positive incidents slightly better than a random model (AUC = 0.589). Likewise, the*

quadratic model that predicted volume exceedances performed slightly better than a random model (AUC = 0.576). Abbreviations as in Table 1.

Response		Model	BIC Weight
Repair		AGE	0.81
Pumping		Null	0.90
Anomalous Pumping	Frequency	Null	0.81
	Volume	AGE + AGE^2	0.99
	Either	Null	0.69

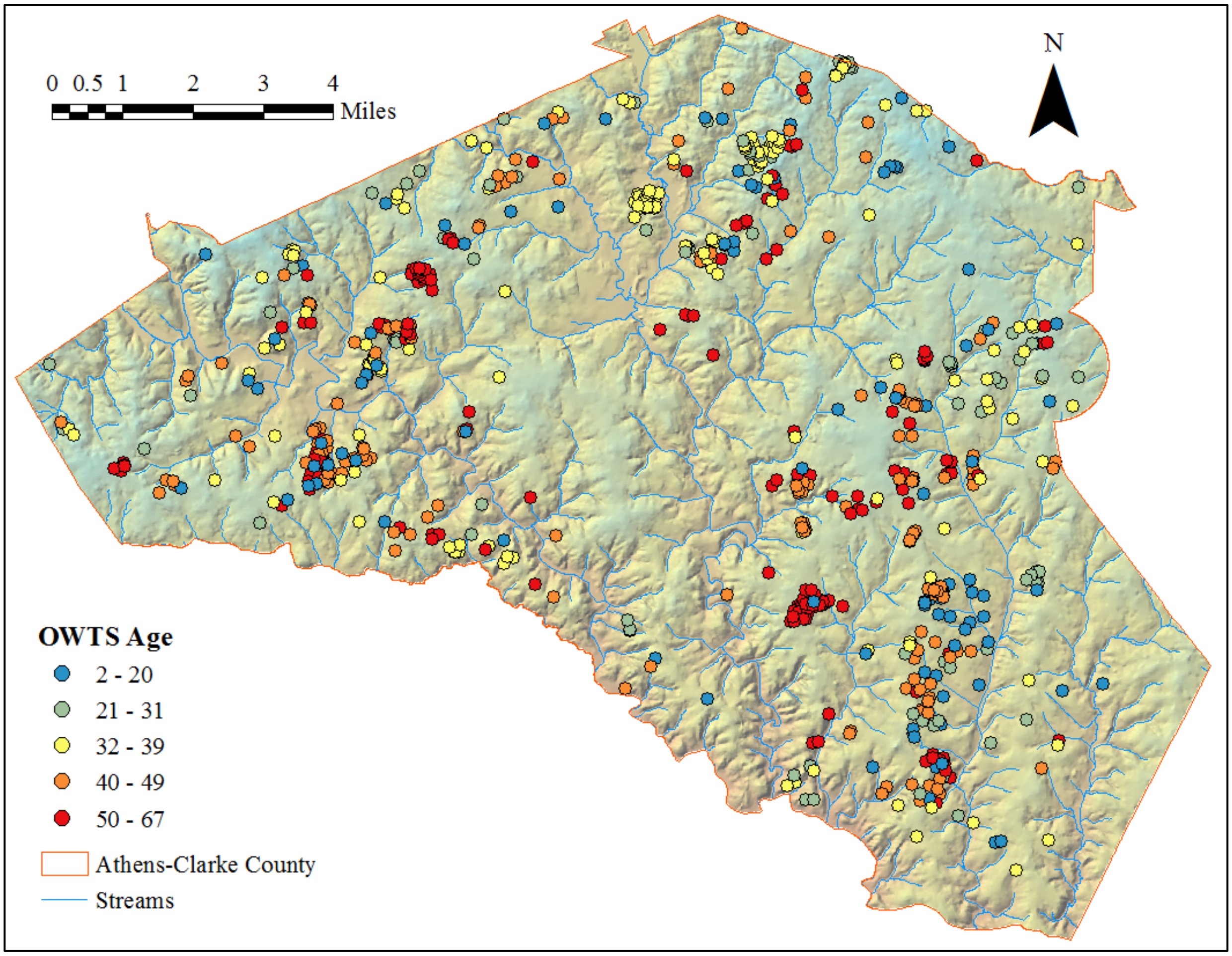


Figure 3. Age and spatial distribution of repaired OWTSs in Athens-Clarke County (n = 690). *County installation records range from January 1940 – March 2018 and repair dates range from May 1972 – March 2018. Data were retrieved in October 2020. The dataset of all registered OWTS locations provided by the county were filtered for quality control and assurance (see Figure 1). A shaded relief layer and stream networks were included to emphasize the spatial heterogeneity of these systems with repair records.*

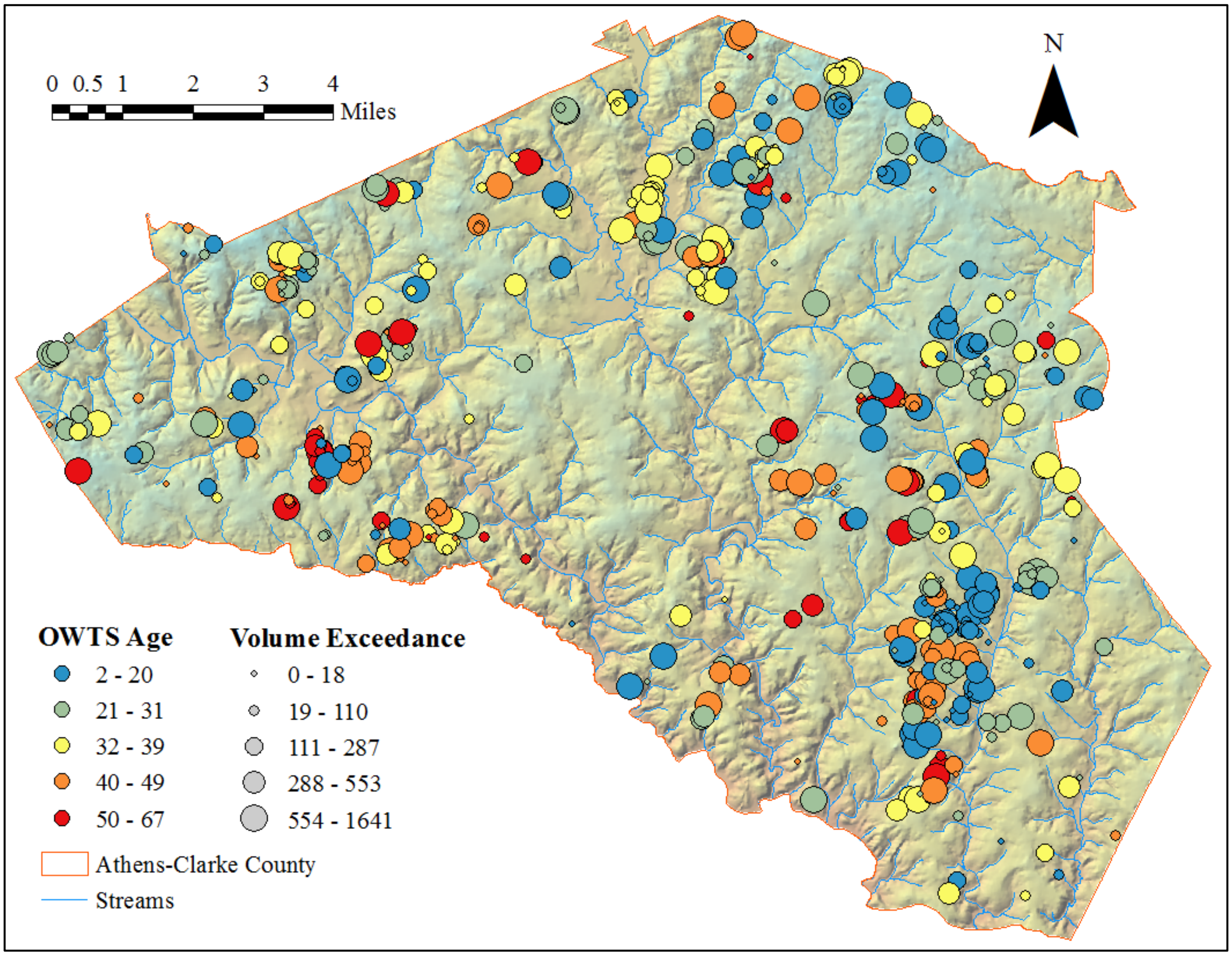


Figure 4. Age, spatial distribution, and maximum volume exceedance (gallons) of OWTS tanks in a 38-month period of pumping records (January 2017 – February 2020; n = 576). *Maximum volume exceedance was calculated as the difference between the recorded septage volume and the installed OWTS tank size. A dataset of all registered OWTS locations was provided by the county and these data were filtered for quality control and assurance (see Figure 1). A shaded relief layer and stream networks were included to emphasize the spatial heterogeneity of these systems with volume exceedances.*

No candidate models in the pumping/no pumping analysis outperformed the null model (Table 4). Candidate models also failed to outperform a null model in the anomalous pumping analysis both for OWTSs that were frequently pumped and in the pooled frequently pumped or volume exceedance pump analysis. However, we found that a model with a quadratic age term performed well at predicting the occurrence of OWTS tanks with volume exceedance (Table 4; Figure 5). This model indicates that tanks during this period of record had a higher probability of pumped septage exceeding the installed tank capacity when they were either relatively new or relatively old. The top performing volume exceedance model identified positive incidences of this type of anomalous pumping slightly better than a random model (AUC = 0.576). No spatial autocorrelation was present in this model's residuals (Appendix A).

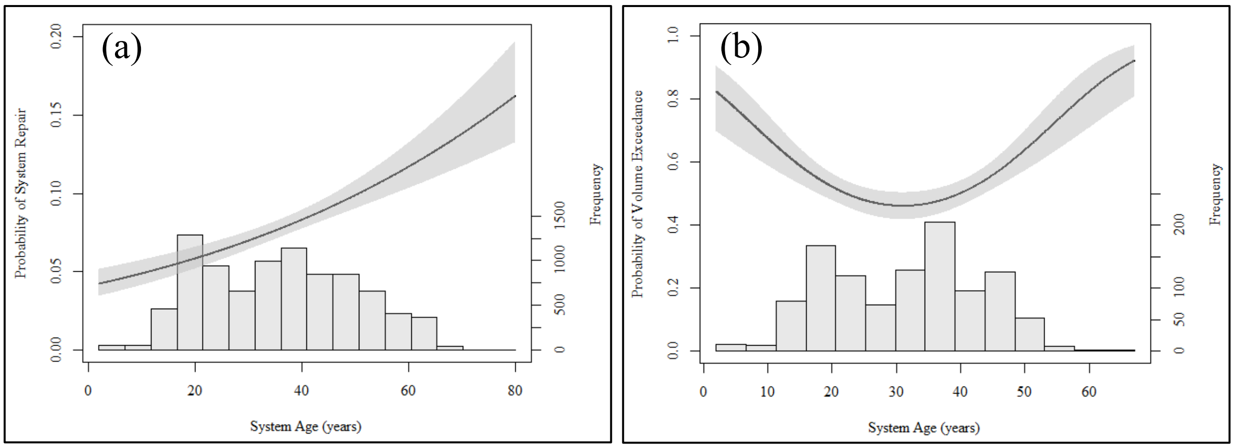


Figure 5. (a) Probability of system repair per OWTS age (years between installation and October 2020), (b) probability of volume exceedance per quadratic OWTS age. *Gray shading represents 95% confidence intervals. Secondary y-axes indicate frequency of binned ages per analysis, which are displayed as underlying histograms.*

After restricting the repair dataset by the date of the repair event, we found that 100 OWTSs were repaired during or five years prior to the pumping data record period. Additionally, there was repair and anomalous pumping concurrence in 18 OWTSs, and we documented a weak statistical relationship between repairs and anomalous pumping ($p = 0.106$). Only 18 OWTSs were repaired during the pumping data record period. Of these 18 systems that were repaired between January 2017 and March 2018, five of them had a prerequisite pump, ranging from eight months to three days before the repair event.

Discussion

Decentralized wastewater treatment infrastructure is a critical component of water management in many regions of the US,[58] and based on land development projections for suburban areas,[59,60] this is likely to remain true well into the future. Scientists and natural resource managers at the local,[61] state,[62] and federal[63] levels have highlighted the threats that poorly sited or maintained OWTSs present to human health and the environment. However, data describing relationships between system characteristics, including age and edaphic conditions, and maintenance records, including pumping and repair, are exceptionally limited. Here, we used an exceptionally complete county-wide dataset to assess relationships between site-level OWTS conditions and pumping and repair patterns. We found that OWTS age was a strong predictor of pumping and repairs and that older OWTSs were more likely to receive repair work. Our analysis of OWTS anomalous pumping patterns suggests that both newer systems, which may have been improperly sited or designed, and older systems that are nearing the end of their service life are most likely to exhibit signs of hydraulic failure. Our findings also suggest that widely available spatial data likely do not have high enough spatial resolution to capture environmental conditions at the local (i.e., drainfield) scale. Collectively, our results indicate that younger OWTSs should be monitored for hydraulic performance within the first few years of installation, while areas with older OWTSs should be targeted for sewer network expansion or proactive repair and/or replacement programs.

Our study suggests that as an OWTS ages, it is increasingly likely to need to be repaired. This finding is supported by both empirical studies and anecdotal evidence. Using similar repair records, Noss and Billa[64] calculated that OWTSs in Amherst, Massachusetts had an expected half-life of approximately 25 years, though the probability of failure quickly increased after this age. In contrast to our findings, the US EPA reported an OWTS lifespan of 15 to 40 years,[57] which is the age range of OWTSs in our dataset that were least likely to show evidence of hydraulic failure. There is general agreement that OWTS monitoring, maintenance, and replacement efforts should be focused on systems that are approaching the end of their service life (e.g. refs 16,38,63). However, the realized lifespans of individual OWTSs are variable, so local practitioners would need to develop objective OWTS performance definitions to impartially implement such programs.

Our unique dataset of pumping records also provided some valuable insights into how these systems are maintained and when potential problems may arise. Unfortunately, we were unable to predict the occurrence of a binary pumped/not pumped response with our candidate models. There are several possible reasons for this. First, the resolution of spatial data in the study area may not have been sufficient to detect meaningful differences among sites, suggesting that these widely available spatial predictors provide little insight into site-level OWTS management. Second, the socio-economic status of homeowners very likely plays a role in whether a household decides to get their OWTS pumped.[65] Finally, homeowners may lack the knowledge that their home is serviced by an OWTS or that regular maintenance is required for the system to remain functional.[15,66] The annualized pumping rate of 5.7% we found for ACC systems is significantly lower than the annualized pumping rate of 11% that Silverman reported in a study of OWTS maintenance in northwest Ohio.[67] In a similar study of OWTS owners in King County, Washington, 47% of survey respondents indicated that they had pumped their system in the last five years.[68] The annual percent of pumped OWTSs we found is markedly low when compared to these other studies, though due to the date range restrictions of the dataset we used, we are uncertain of the number of OWTS owners who serviced their systems in the preceding years.

The broad-scale phenomenon of anomalous OWTS pumping reported in this study, particularly in young systems (2-10 years), has not been previously reported. We posit three explanations for these unexpected results. First, the high failure rate of recently installed OWTSs may indicate incorrectly constructed systems. In a study of OWTS drainfield performance in North Carolina Piedmont soils, Coulter et al.[69] showed that over half of OWTSs on certain clayey soils failed within two years of installation. The researchers attributed these failures to one of three causes: (i) the presence of a shallow impervious layer, (ii) poor surface drainage, or (iii) construction damage to the soils during installation. Second, new homeowners may not know that system loading is limited by soil absorption rates and overload the system. Third, it is well established that the biologically active layer of soil directly below the drainfield (i.e., the biomat) undergoes rapid development within the first years of installation.[70] Soil clogging from fine particulate matter or elevated loading rates can negatively affect this layer's development and may result in reduced water acceptance rates.[71] Any of these three site-level attributes may cause drainfields to become saturated and backup into the OWTS tank, resulting in volume exceedance events. The presence of garbage disposals can also dramatically increase the suspended solid loading rate to OWTS tanks and subsequently the drainfield, which may cause faster than expected clogging and reduced acceptance rates.[12] The use of these appliances may be contributing in part to volume exceedances at OWTSs of any age. We also documented the expected pattern of increasing anomalous pumping with system age. This suggests that systems that had exceeded the recommended life expectancy of OWTSs (in this study >50 years) may be failing due to system deterioration and the reduction in absorption rates due to soil clogging over time.

Our finding that OWTSs with anomalous pumping records are weak, yet potentially consequential, correlates of recent repairs provides additional insights into how these decentralized systems were maintained. This relationship suggests many systems that were pumped at frequencies or volumes indicative of hydraulic failure were also the same OWTSs which received repair work. We speculate that these systems may have been sited inappropriately. In a survey of OWTSs on hydraulically sensitive soils in Connecticut, Groff and Obeda[72] found that homeowners continued to experience signs of hydraulic system failure, even after repair work was done. Owning an OWTS that concurrently exhibits signs of hydraulic failure, and results in anomalous pumping, and also necessitates repairs, would result in exceptionally high costs for homeowners. Data on OWTS repair expenses from the Boulder County Public Health Department, Colorado showed that 60% of repairs had an average cost of $14 866 per event,[73] while regular pumping usually costs several hundred dollars per event.[57] As few homeowners are likely to invest this much into their wastewater system, these prohibitive expenses suggest that there may be many OWTSs that need repair work, but whose homeowners choose to only get their systems pumped as a cheaper, short-term fix. Alternatively, homeowners may initially invest in repair work, but if their system continues to fail, they may feel they have reached the limit of how much they can or are willing to invest in their OWTS and opt to let the system fail. Therefore, our analysis only captures a subset of OWTSs where owner(s) have invested substantial capital into both pumping and repairs.

The socioeconomic status of OWTS owners may also contribute to many of the pumping and repair phenomena we report in this study. The costs of OWTS pumping, inspection, and repairs can be economically prohibitive for many people, even when they know that their system is failing. Our data indicate that in ACC, older systems are more likely to be repaired and are approximately equally as likely to exhibit signs of hydraulic failure as a young system. This finding may be relevant to class and racial divides in housing. Using a long-term, nationally-representative dataset of pre-retirement Americans, Flippen[74] found that on average, African

Americans are more likely to own older houses. Previous work in ACC has documented that older OWTSs are disproportionately located in predominately non-white and/or impoverished census blocks in ACC.[43] Hence, the deleterious effects of failing decentralized wastewater infrastructure may be impacting vulnerable communities to a greater degree than what is currently recognized. Future work should couple census and cadastral data with wastewater infrastructure to provide insights into community wealth and resilience to help identify areas of high management priority.

Municipal governments throughout the US are challenged to manage decentralized wastewater infrastructure with relatively limited information. With a national median housing age of 37 years[31] and the understanding that only a small portion of malfunctioning OWTSs are ever repaired or replaced by homeowners,[14,68] it is very likely that aging and obsolete decentralized wastewater infrastructure is commonplace. However, ACC is unique in that it has invested in a robust and continually updated dataset of OWTS locations and maintenance. Yet, there are still many unknowns in the county dataset. For instance, we have only a portion of the "positive" incidences of repairs, and we cannot quantify the number of OWTSs that need repairs but have not received any work. Additionally, the analysis of volume exceedance pumps in this study most likely documents situations where hydraulic failure was evident. Nevertheless, these kinds of records do not provide insight about drainfield failures where groundwater or surface water contamination is likely. Further site-specific assessments would be needed to identify the characteristics of those OWTSs that are failing to completely treat wastewater.

By coupling Public Utilities Department pumping records with installation and repair records from the Georgia Department of Public Health, our study indicates that older OWTSs (> 50 years) are subject to greater failure rates than younger systems. Thus, policies should be considered to support OWTS maintenance or replacement of older OWTSs. Notably, our analysis also identified a markedly low OWTS pumping rate during the period of record. Education initiatives are relatively inexpensive and if the OWTS data were combined with property and housing records that document building age, this information could be used in targeted outreach efforts to notify homeowners when their OWTSs are due for maintenance. However, the effectiveness of outreach programs alone to change homeowner OWTS maintenance behavior is likely low,[67] and should be only one component in watershed management efforts. Regulations to monitor and incentivize specific OWTS management outcomes, such as more robust installation guidelines and appropriate pumping intervals,[64,75] may also lead to increased homeowner participation in regular OWTS maintenance. Surprisingly, our results suggest that younger systems are also subject to hydraulic failure. These instances may be attributed to poor placement or installation practices, suggesting local governments may want to invest in more training for and oversight of OWTS installation. However, more work is needed to understand the causes of system failure. The findings from this study highlight that widely available spatial data may not be granular enough to account for site-level phenomenon associated with decentralized wastewater treatment. Municipalities should continue investing in site assessments before approving OWTS installations. By specifying the maintenance and risk characteristics of OWTSs, managers may be better informed to implement programs to monitor system performance and address the externalities introduced by malfunctioning OWTSs. Focusing efforts on both newly installed OWTSs as well as aging ones may help managers identify systems that are not performing adequately and guide decision makers in prioritizing wastewater infrastructure investments and policies.

REFERENCES

(1) Eggers, F. J.; Thackeray, A. *32 Years of Housing Data*; Project No. 017-002; Bethesda, MD, 2007.

(2) US Census Bureau. *American Housing Survey for the United States: 2007*; H150/07; U.S. Government Printing Office: Washington, D.C., 2008.

(3) US Environmental Protection Agency. *Annual Report 2013: Decentralized Wastewater Management Program Highlights*; EPA-832-R-140006; 2013.

(4) US Environmental Protection Agency. *Response to Congress on Use of Decentralized Wastewater Treatment Systems*; 1997.

(5) Bouma, J.; Ziebell, W. A.; Walker, W. G.; Olcott, P. G.; McCoy, E.; Hole, F. D. *Soil Absorption of Septic Tank Effulent: A Field Study of Some Major Soils in Wisconsin*; Madison, WI, 1972.

(6) Van Cuyk, S.; Siegrist, R. L.; Lowe, K.; Harvey, R. W. Evaluating Microbial Purification during Soil Treatment of Wastewater with Multicomponent Tracer and Surrogate Tests. *J. Environ. Qual.* **2004**, *33* (1), 316–329. https://doi.org/10.2134/jeq2004.3160.

(7) Stewart, L. W.; Reneau, R. B. Shallowly Placed, Low Pressure Distribution System to Treat Domestic Wastewater in Soils with Fluctuating High Water Tables. *J. Environ. Qual.* **1988**, *17* (3), 499–504. https://doi.org/10.2134/jeq1988.00472425001700030026x.

(8) Walker, W. G.; Bouma, J.; Keeney, D. R.; Magdoff, F. R. Nitrogen Transformations During Subsurface Disposal of Septic Tank Effluent in Sands: I. Soil Transformations. *J. Environ. Qual.* **1973**, *2* (4), 475–480.

(9) Arnscheidt, J.; Jordan, P.; Li, S.; McCormick, S.; McFaul, R.; McGrogan, H. J.; Neal, M.; Sims, J. T. Defining the Sources of Low-Flow Phosphorus Transfers in Complex Catchments. *Sci. Total Environ.* **2007**, *382* (1), 1–13. https://doi.org/10.1016/j.scitotenv.2007.03.036.

(10) Macintosh, K. A.; Jordan, P.; Cassidy, R.; Arnscheidt, J.; Ward, C. Low Flow Water Quality in Rivers; Septic Tank Systems and High-Resolution Phosphorus Signals. *Sci. Total Environ.* **2011**, *412–413*, 58–65. https://doi.org/10.1016/j.scitotenv.2011.10.012.

(11) Katz, B. G.; Griffin, D. W.; McMahon, P. B.; Harden, H. S.; Wade, E.; Hicks, R. W.; Chanton, J. P. Fate of Effluent-Borne Contaminants beneath Septic Tank Drainfields Overlying a Karst Aquifer. *J. Environ. Qual.* **2010**, *39* (4), 1181–1195. https://doi.org/10.2134/jeq2009.0244.

(12) Otis, R. J.; Boyle, W. C.; Clements, E. V.; Schmidt, C. J. *Design Manual: Onsite Wastewater Treatment and Disposal Systems*; Cincinnati, OH, 1980.

(13) Dunne, T.; Leopold, L. B. Water in the Soil. In *Water in Environmental Planning*; W. H. Freeman and Company: San Francisco, CA, 1978; pp 163–191.

(14) US Environmental Protection Agency. *Onsite Wastewater Treatment System Manual*; US EPA/625/R-00/008; 2002.

(15) Goonetilleke, A.; Dawes, L. A. Audit of Septic Tank Performance. In *Proceedings of the Conference On-site ’01 Advancing on-site wastewater systems*; Patterson, R. A., Ed.; Lanfax Laboratories: Armidale, Australia, 2001; pp 155–162.

(16) Siegrist, R. L.; Tyler, E. J.; Peter D. Jenssen. Design and Performance of Onsite Wastewater Soil Absorption Systems. In *National Research Needs Conference Proceedings: Risk- Based Decision Making for Onsite Wastewater Treatment.*; St. Louis,

MO, 2000; p 51.

(17) Collick, A. S.; Easton, Z. M.; Montalto, F. A.; Gao, B.; Kim, Y.-J.; Day, L.; Steenhuis, T. S. Hydrological Evaluation of Septic Disposal Field Design in Sloping Terrains. *J. Environ. Eng.* **2006**, *132* (10), 1289–1297. https://doi.org/10.1061/(ASCE)0733-9372(2006)132:10(1289).

(18) Dawes, L.; Goonetilleke, A. An Investigation into the Role of Site and Soil Characteristics in Onsite Sewage Treatment. *Environ. Geol.* **2003**, *44* (4), 467–477. https://doi.org/10.1007/s00254-003-0781-6.

(19) Hassett, J. M.; Chan, A.; Martin, J. S.; DeJohn, T. Effectiveness of Alternative On-Site Wastewater Treatment Technologies in the Catskill/Delaware Watershed. In *Eleventh Individual and Small Community Sewage Systems Conference Proceedings*; American Society of Agricultural and Biological Engineers: Warwick, RI, 2007; pp 1–11. https://doi.org/10.13031/2013.24020.

(20) Gold, A. J.; Sims, J. T. Research Needs in Decentralized Wastewater Treatment and Management: A Risk-Based Approach To Nutrient Contamination. In *National Research Needs Conference Proceedings: Risk-based decision making for onsite wastewater treatment*; Electric Power Research Institute, U.S. Environmental Protection Agency, and National Decentralized Water Resources Capacity Development Project: St. Louis, MO, 2000; pp 122–151.

(21) Van Cuyk, S.; Siegrist, R. L. Pathogen Fate in Wastewater Soil Absorbtion Systems as Affected by Effluent Quality and Soil Clogging Genesis. In *Proceedings of the 9th National Symposium on Individual and Small Community Sewage Systems*; Mancl, K., Ed.; American Society of Agricultural Engineers: Fort Worth, TX, 2001; pp 125–136.

(22) US Environmental Protection Agency. *A Homeowner's Guide to Septic Systems*; EPA-832-B-02-005; Cincinnati, OH, 2005.

(23) Kohler, L. E.; Silverstein, J.; Rajagopalan, B. Risk-Cost Estimation of On-Site Wastewater Treatment System Failures Using Extreme Value Analysis. *Water Environ. Res.* **2016**, *89* (5), 406–415. https://doi.org/10.2175/106143016x14609975747289.

(24) Williams, J. R.; Powell, G. M.; Jones, R. D.; Murat, A.; Perry, W. R. Costs of Individual Onsite Wastewater Treatment Systems. In *On-Site Wastewater Treatment X*; Mankin, K. R., Ed.; American Society of Agricultural and Biological Engineers: Sacramento, CA, 2004; Vol. 701P0104, pp 130–139.

(25) Cardona, M. E. *Nutrient and Pathogen Contributions To Surface and Subsurface Waters from On-Site Wastewater Systems - A Review*; Raleigh, NC, 1998.

(26) Beal, C. D.; Gardner, E. A.; Menzies, N. W. Process, Performance, and Pollution Potential: A Review of Septic Tank - Soil Absorption Systems. *Soil Res.* **2005**, *43* (7), 781. https://doi.org/10.1071/SR05018.

(27) Mohamed, R. Why Households in the United States Do Not Maintain Their Septic Systems and Why State-Led Regulations Are Necessary: Explanations from Public Goods Theory. *Int. J. Sustain. Dev. Plan.* **2009**, *4* (2), 143–157. https://doi.org/10.2495/SDP-V4-N2-143-157.

(28) Pugel, K. Policy Reform Alternatives to Combat Failure of Onsite Wastewater Treatment Systems. *J. Sci. Policy Gov.* **2019**, *14* (1).

(29) US Environmental Protection Agency. Funding for septic systems https://www.epa.gov/septic/funding-septic-systems (accessed Jan 2, 2021).

(30) Hughes, J.; Simonson, A. Government Financing for On-Site Wastewater Treatment

Facilities in North Carolina. *Pop. Gov.* **2005**, No. Fall, 37–45.

(31) American Society of Civil Engineers. *Wastewater*; 2021.

(32) Clayton, J. W. An Analysis of Septic Tank Survival Data From 1952 to 1972 in Fairfax County, Virginia. *J. Environ. Health* **1974**, *36* (6), 562–567.

(33) Winneberger, J. T. Sanitary Surveys and Survival Curves of Septic Tank Systems. *J. Environ. Health* **1975**, *38* (1), 36–39.

(34) Withers, P. J. A.; Jordan, P.; May, L.; Jarvie, H. P.; Deal, N. E. Do Septic Tank Systems Pose a Hidden Threat to Water Quality? *Front. Ecol. Environ.* **2014**, *12* (2), 123–130. https://doi.org/10.1890/130131.

(35) Yates, M. V. Septic Tank Density and Ground-Water Contamination. *Groundwater* **1985**, *23* (5), 586–591.

(36) Oliver, C. W.; Risse, L. M.; Radcliffe, D. E.; Habteselassie, M.; Clarke, J. Evaluating Potential Impacts of On-Site Wastewater Treatment Systems on the Nitrogen Load and Baseflow in Streams of Watersheds in Metropolitan Atlanta, Georgia. *Trans. ASABE* **2014**, *57* (4), 1121–1128. https://doi.org/10.13031/trans.57.10329.

(37) Borchardt, M. A.; Chyou, P. H.; DeVries, E. O.; Belongia, E. A. Septic System Density and Infectious Diarrhea in a Defined Population of Children. *Environ. Health Perspect.* **2003**, *111* (5), 742–748. https://doi.org/10.1289/ehp.5914.

(38) Georgia Department of Public Health. *Background and Use of Onsite Wastewater Treatment Systems in Georgia*; 2019.

(39) Georgia Department of Public Health. *Manual for On-Site Sewage Management Systems*; O.C.G.A. Sect.12-8-1, 31-2A-6, and 31-2A-11, 2019.

(40) US Environmental Protection Agency. *Voluntary National Guidelines for Management of*

Onsite and Clustered (Decentralized) Wastewater Treatment Systems; US EPA 832-B-03-001; 2003.

(41) US Census Bureau. QuickFacts: Athens-Clarke County (balance), Georgia https://www.census.gov/quickfacts/athensclarkecountybalancegeorgia (accessed Apr 4, 2020).

(42) Athens-Clarke County Public Utilities Department. Water Reclamation https://www.accgov.com/1235/Water-Reclamation (accessed Dec 10, 2020).

(43) Capps, K. A.; Bateman McDonald, J. M.; Gaur, N.; Parsons, R. Assessing the Socio-Environmental Risk of Onsite Wastewater Treatment Systems to Inform Management Decisions. *Environ. Sci. Technol.* **2020**, *54* (23), 14843–14853. https://doi.org/10.1021/acs.est.0c03909.

(44) Shaw-Burgess, C.; Bloyer, D. Getting a Handle on It. *Georg. Oper.* **2020**, *57* (1), 55–60.

(45) Athens-Clarke County Public Utilities Department. Sewer/Wastewater/Septage Rates https://www.accgov.com/1223/Sewer-Wastewater-Septage-Rates (accessed Jun 11, 2021).

(46) Georgia Environmental Protection Division. County Environmental Protection Division Permitted Public Wastewater Treatment Facilities Accepting Septage. 2017, pp 1–2.

(47) US Geological Survey. National Elevation Dataset - 30 Meter: The National Map Viewer. US Geological Survey 2013.

(48) Beven, K. J.; Kirby, M. J. A Physically Based, Variable Contributing Area Model of Basin Hydrology / Un Modèle à Base Physique de Zone d'appel Variable de l'hydrologie Du Bassin Versant. *Hydrol. Sci. J.* **1979**, *24* (1), 43–69. https://doi.org/10.1080/02626667909491834.

(49) Pourali, S. H.; Arrowsmith, C.; Chrisman, N.; Matkan, A. A.; Mitchell, D. Topography

Wetness Index Application in Flood-Risk-Based Land Use Planning. *Appl. Spat. Anal. Policy* **2016**, *9* (1), 39–54. https://doi.org/10.1007/s12061-014-9130-2.

(50) Tarboton, D. Terrain Analysis Using Digital Elevation Models (TauDEM). Utah Water Research Laboratory, Utah State University 2016.

(51) Esri. ArcMap. Redlands, CA 2019.

(52) R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria 2020.

(53) Soil Survey Staff. Soil Data Development Toolbox. United States Department of Agriculture, Natural Resources Conservation Service 2020.

(54) Soil Survey Staff. Gridded Soil Survey Geographic (GSSURGO) Database for Clarke and Oconee Counties, Georgia. United States Department of Agriculture, Natural Resources Conservation Service 2020.

(55) US Geological Survey. National Hydrography Dataset Plus High Resolution (NHDPlus HR): The National Map Viewer. US Geological Survey, National Geospatial Program 2018.

(56) Aho, K.; Derryberry, D.; Peterson, T. Model Selection for Ecologists: The Worldviews of AIC and BIC. *Ecology* **2014**, *95* (3), 631–636.

(57) US Environmental Protection Agency. *New Homebuyer's Guide to Septic Systems*; EPA-832-F-17-009; Washington, D.C., 2017.

(58) US Census Bureau. *Characteristics of New Housing: Type of Sewage System by Water Supply Source in New Single-Family Houses Completed*; Washington, D.C., 2020.

(59) Hamidi, S.; Ewing, R. A Longitudinal Study of Changes in Urban Sprawl between 2000 and 2010 in the United States. *Landsc. Urban Plan.* **2014**, *128*, 72–82.

https://doi.org/10.1016/j.landurbplan.2014.04.021.

(60) Homer, C.; Dewitz, J.; Jin, S.; Xian, G.; Costello, C.; Danielson, P.; Gass, L.; Funk, M.; Wickham, J.; Stehman, S.; Auch, R.; Riitters, K. Conterminous United States Land Cover Change Patterns 2001–2016 from the 2016 National Land Cover Database. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162* (March), 184–199. https://doi.org/10.1016/j.isprsjprs.2020.02.019.

(61) Walker, R. L.; Cotton, C. F.; Payne, K. A. *A GIS Inventory of On-Site Septic Systems Adjacent to the Coastal Waters of McIntosh County, Georgia*; Athens, GA, 2003.

(62) Macrellis, A.; Douglas, B. *Update of the Advanced On-Site Wastewater Treatment and Management Market Study: State Reports*; 05-DEC-3SGd; Gloucester, MA, 2009.

(63) US Environmental Protection Agency. *Decentralized Wastewater Treatment Systems: A Program Strategy*; Cincinnati, OH, 2005.

(64) Noss, R. R.; Billa, M. Septic System Maintenance Management. *J. Urban Plan. Dev.* **1988**, *114* (2), 73–90. https://doi.org/10.1061/(ASCE)0733-9488(1988)114:2(73).

(65) Flowers, C. F.; Ward, J. K.; Winkler, I. *Flushed and Forgotten: Sanitation and Wastewater in Rural Communities in the United States*; Lowndes County, AL, 2019.

(66) Harrison, M.; Stanwyck, E.; Beckingham, B.; Starry, O.; Hanlon, B.; Newcomer, J. Smart Growth and the Septic Tank: Wastewater Treatment and Growth Management in the Baltimore Region. *Land use policy* **2012**, *29* (3), 483–492. https://doi.org/10.1016/j.landusepol.2011.08.007.

(67) Silverman, G. S. The Effectiveness of Education as a Tool to Manage Onsite Septic Systems. *J. Environ. Health* **2005**, *68* (1), 17–22.

(68) Gomez, A.; Taylor, M.; Nicola, R. M. Development of Effective On-Site Sewage Disposal

Surveys in King County, Washington. *J. Environ. Health* **1992**, *54* (5), 20–25.

(69) Coulter, J. B.; Bendixen, T. W.; Edwards, A. B. *Appendices to Study of Seepage Beds*; Cincinnati, OH, 1960.

(70) Siegrist, R. L.; Boyle, W. C. Wastewater-Induced Soil Clogging Development. *J. Environ. Eng.* **1987**, *113* (3), 550–566.

(71) Siegrist, R. L. Soil Clogging during Subsurface Wastewater Infiltration as Affected by Effluent Composition and Loading Rate. *J. Environ. Qual.* **1987**, *16* (2), 181–187. https://doi.org/10.2134/jeq1987.00472425001600020016x.

(72) Groff, D. W.; Obeda, B. A. Septic Performance in Hydrologically Sensitive Soils. *Wetlands* **1982**, *2* (1), 286–302. https://doi.org/10.1007/BF03160562.

(73) Kohler, L. E.; Silverstein, J.; Rajagopalan, B. Predicting Life Cycle Failures of On-Site Wastewater Treatment Systems Using Generalized Additive Models. *Environ. Eng. Sci.* **2016**, *33* (2), 112–124. https://doi.org/10.1089/ees.2015.0275.

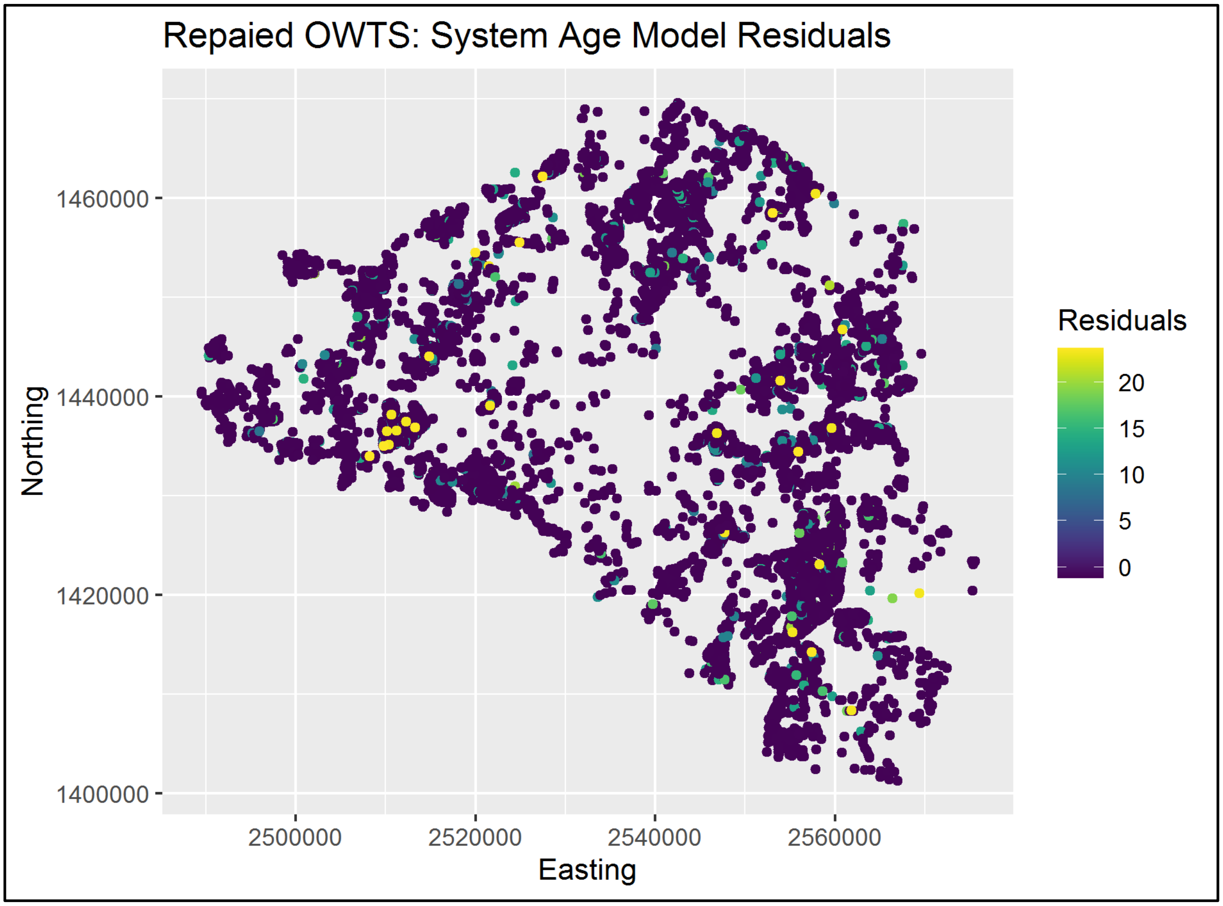
(74) Flippen, C. Unequal Returns to Housing Investments? A Study of Real Housing Appreciation among Black, White, and Hispanic Households. *Soc. Forces* **2004**, *82* (4), 1523–1551. https://doi.org/10.1353/sof.2004.0069.

(75) Goehring, D. R.; Carr, F. R. Septic System Problems on an Urban Fringe. *J. Water Resour. Plan. Manag. Div.* **1980**, *106* (1), 89–102. https://doi.org/10.1061/JWRDDC.0000158.

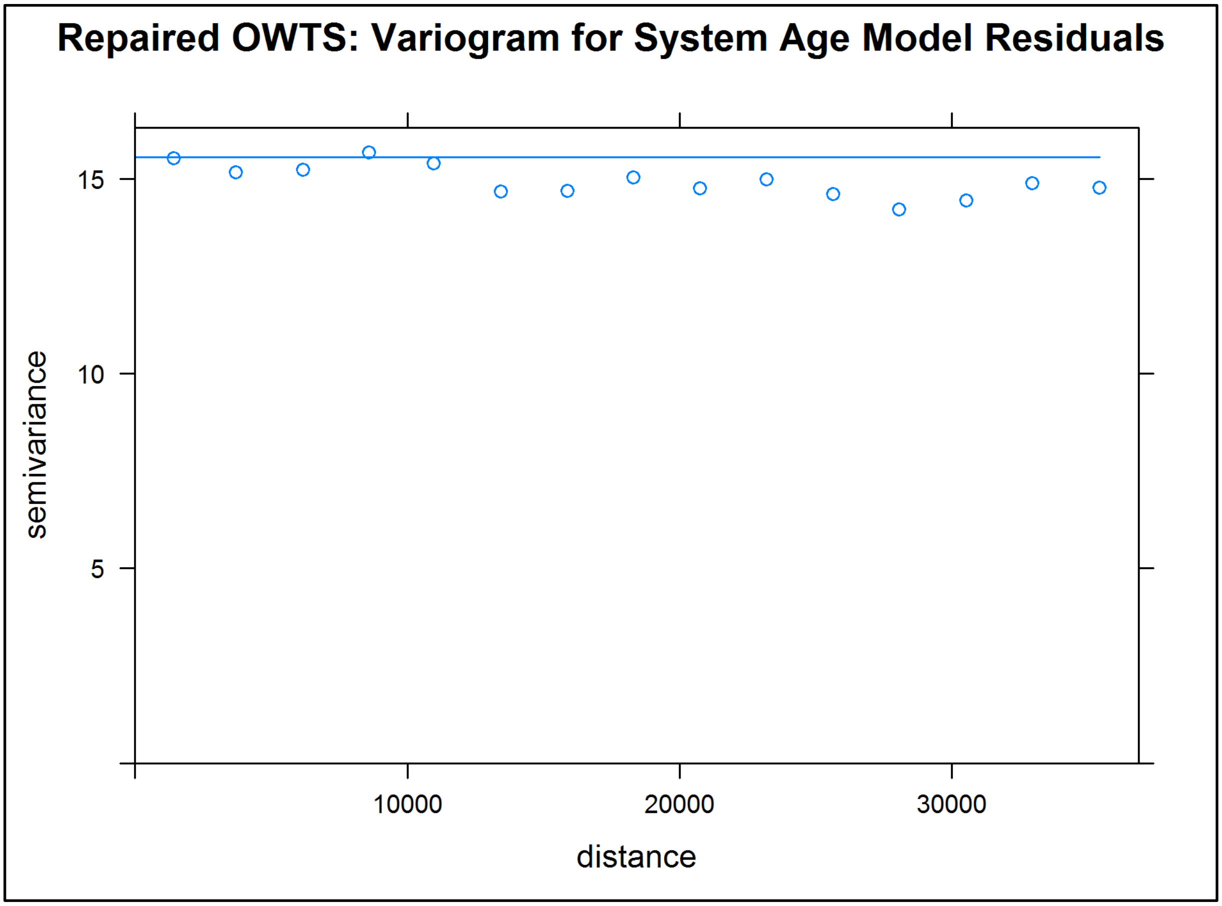
APPENDIX A

SUPPLEMENTAL FIGURES

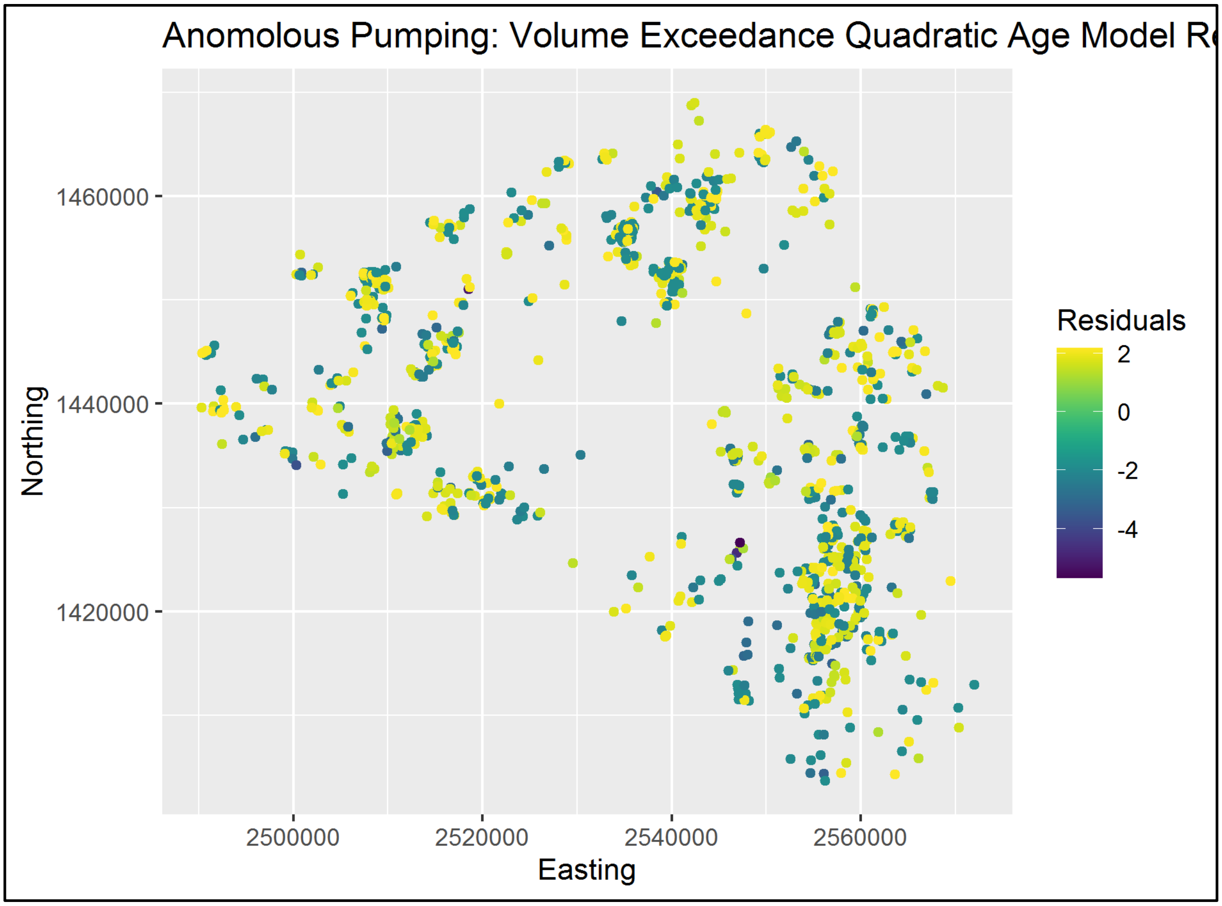
Residual maps and semivariograms were developed and assessed for each top model in the study. Residual maps help to identify any occurrences of spatial clustering in model residuals. Semivariogram plots show the dissimilarity of spatial points across varying lag distances. By plotting the semi-variance of the model residuals, semivariograms help us determine if spatial autocorrelation remains in the model residuals or if most of the variance is explained by our candidate model.



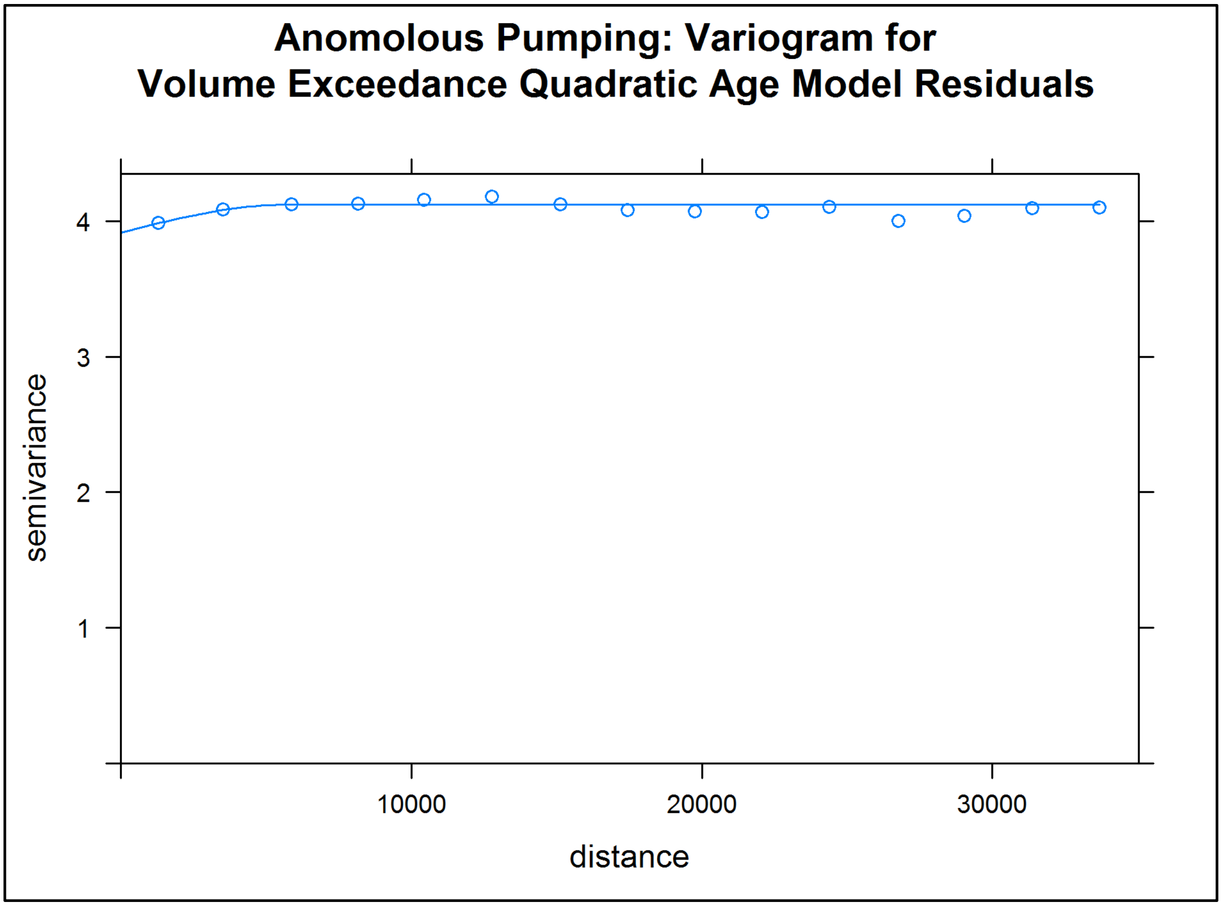
Supplemental 1. Spatial distribution of model residuals for the top performing "AGE" model in the repair analysis. *Coordinate system is NAD 83 - State Plane Georgia West FIPS 1002 Feet (EPSG:2240).*



Supplemental 2. Semivariogram for top performing "AGE" model in the repair analysis. *Distance units are in feet.*



Supplemental 3. Spatial distribution of model residuals for the top performing "AGE + AGE²" model in the volume exceedance anomalous pumping analysis. *Coordinate system is NAD 83 - State Plane Georgia West FIPS 1002 Feet (EPSG:2240).*



Supplemental 4. Semivariogram for top performing "AGE + AGE^2" model in the volume exceedance anomalous pumping analysis. *Distance units are in feet.*

APPENDIX B

R CODE AND SELECT OUTPUTS

The following code was written and finalized in the R statistical environment. Raster predictor variables consist of the topographic wetness index, soil conditions (average saturated hydraulic conductivity, depth to water table, and depth to restrictive layer) and distance to stream. Response variables are coded as the binary presence/absence of OWTS pumping, anomalous pumping, or repairs at OWTS point locations. Spatial data for both predictor and response variables were mostly pre-processed in ArcMap 10.7. Some additional dataset sub-setting and quality assurance measures were conducted in R. For reproducibility and security, personal directory path names have been substituted by "./" or ".\" depending on the file or driver being accessed.

OWTS Logistic Regression

Kyle Connelly

June 26th, 2021

```
# Install and load packages
#install.packages("arcgisbinding", repos="https://r.esri.com", type="win.binary")
library(arcgisbinding); arc.check_product() #interact with gdb's

## *** Please call arc.check_product() to define a desktop license.

## product: ArcGIS Desktop (10.7.1.11595)
## license: Advanced
## version: 1.0.1.244

library(AICcmodavg)
library(raster)

## Loading required package: sp

library(rgdal)
```

```
## rgdal: version: 1.5-21, (SVN revision 1105)
## Geospatial Data Abstraction Library extensions to R successfully loaded


## Loaded GDAL runtime: GDAL 3.0.4, released 2020/01/28
## Path to GDAL shared files: C:/Users/knc51897/AppData/Local/RStudio-Desktop/R-4.0.3/library/
rgdal/gdal
## GDAL binary built with GEOS: TRUE
## Loaded PROJ runtime: Rel. 6.3.1, February 10th, 2020, [PJ_VERSION: 631]
## Path to PROJ shared files: C:/Users/knc51897/AppData/Local/RStudio-Desktop/R-4.0.3/library/
rgdal/proj
## Linking to sp version:1.4-5
## To mute warnings of possible GDAL/OSR exportToProj4() degradation,
## use options("rgdal_show_exportToProj4_warnings"="none") before loading rgdal.

library(rgeos)

## rgeos version: 0.5-5, (SVN revision 640)
##  GEOS runtime version: 3.8.0-CAPI-1.13.1
##  Linking to sp version: 1.4-4
##  Polygon checking: TRUE

library(ModelMetrics)

##
## Attaching package: 'ModelMetrics'

## The following object is masked from 'package:base':
##
##     kappa

library(spatialEco)

##
## Attaching package: 'spatialEco'

## The following object is masked from 'package:raster':
##
##     shift

# setwd(.)

# Load pre-processed raster data (i.e., predictor layers)
twi <- as.raster(arc.raster(arc.open("./ACC_twi_1.tif")))
wt_dep <- as.raster(arc.raster(arc.open("./ACC_Dep2WatTbl_WTA_1.tif")))
ksat <- as.raster(arc.raster(arc.open("./ACC_Ksat_WTA_30to183cm_1.tif")))
restr_dep <- as.raster(arc.raster(arc.open("./ACC_Dep2ResLyr_WTA_1.tif")))
dist_strm <- as.raster(arc.raster(arc.open("./ACC_dist_strm_1.tif")))


# Create raster stack (collection of rasters with same extent and resolution) and rename compo
nents
rs <- stack(twi,wt_dep,ksat,restr_dep,dist_strm)
names(rs) <- c('topo.wet','wtrtbl.dep','sat.cond','dep.to.restr','dist.strm')

# Convert stack to grid (preserves names)

rs_grd <- writeRaster(rs,"SDMStack.grd", format="raster", overwrite=TRUE)

# Read in repair, pumping, and anomalous pumping response point shapefiles (0/1 data of observ
ations and absences)
```

```
gdb_path <- "./ACC_Sewer_and_Septic_knc.gdb"
repair <- readOGR(dsn=gdb_path, layer="SepticTanks_Repaired1")

## OGR data source with driver: OpenFileGDB
## Source: ".\ACC_Sewer_and_Septic_knc.gdb", layer: "SepticTanks_Repaired1"


## with 8826 features
## It has 27 fields

repair1 <- readOGR(dsn=gdb_path, layer="SepticTanks_Repaired1")

## OGR data source with driver: OpenFileGDB
## Source: ".\ACC_Sewer_and_Septic_knc.gdb", layer: "SepticTanks_Repaired1"
## with 8826 features
## It has 27 fields

pump <- readOGR(dsn=gdb_path, layer="SepticTanks_Pump1")

## OGR data source with driver: OpenFileGDB
## Source: ".\ACC_Sewer_and_Septic_knc.gdb", layer: "SepticTanks_Pump1"
## with 7709 features
## It has 35 fields

pump.anom <- readOGR(dsn=gdb_path, layer="SepticTanks_Pump_Anom1")

## OGR data source with driver: OpenFileGDB
## Source: ".\ACC_Sewer_and_Septic_knc.gdb", layer: "SepticTanks_Pump_Anom1"
## with 1077 features
## It has 54 fields

# Extract raster values of predictor variables to OWTS points
repair_extr <- data.frame(extract(rs_grd, repair))

## Warning in .local(x, y, ...): Transforming SpatialPoints to the CRS of the
## Raster

pump_extr <- data.frame(extract(rs_grd, pump))

## Warning in .local(x, y, ...): Transforming SpatialPoints to the CRS of the
## Raster

pump.anom_extr <- data.frame(extract(rs_grd, pump.anom))

## Warning in .local(x, y, ...): Transforming SpatialPoints to the CRS of the
## Raster

# Append extracted values to the OWTS point tables
repair@data <- data.frame(repair@data, repair_extr[match(rownames(repair@data), rownames(repai
r_extr)),])
pump@data <- data.frame(pump@data, pump_extr[match(rownames(pump@data), rownames(pump_extr)),]
)

pump.anom@data <- data.frame(pump.anom@data, pump.anom_extr[match(rownames(pump.anom@data), ro
wnames(pump.anom_extr)),])

# Drop observations with N/A's in predictor variable columns (largely points outside or on edg
e of study extent)
repair <- sp.na.omit(repair, col.name = "topo.wet")
repair <- sp.na.omit(repair, col.name = "wtrtbl.dep")
repair <- sp.na.omit(repair, col.name = "sat.cond")
repair <- sp.na.omit(repair, col.name = "dep.to.restr")
```

```
pump <- sp.na.omit(pump, col.name = "topo.wet")
pump <- sp.na.omit(pump, col.name = "wtrtbl.dep")
pump <- sp.na.omit(pump, col.name = "sat.cond")
pump.anom <- sp.na.omit(pump.anom, col.name = "topo.wet")

# Center covariates
repair@data$Age.C <- repair@data$Age - mean(repair@data$Age)
repair@data$dist.strm.C <- repair@data$dist.strm - mean(repair@data$dist.strm)
repair@data$topo.wet.C <- repair@data$topo.wet - mean(repair@data$topo.wet)
repair@data$wtrtbl.dep.C <- repair@data$wtrtbl.dep - mean(repair@data$wtrtbl.dep)
repair@data$sat.cond.C <- repair@data$sat.cond - mean(repair@data$sat.cond)
repair@data$dep.to.restr.C <- repair@data$dep.to.restr - mean(repair@data$dep.to.restr)
pump@data$Age.C <- pump@data$Age - mean(pump@data$Age)
pump@data$dist.strm.C <- pump@data$dist.strm - mean(pump@data$dist.strm)
pump@data$topo.wet.C <- pump@data$topo.wet - mean(pump@data$topo.wet)
pump@data$wtrtbl.dep.C <- pump@data$wtrtbl.dep - mean(pump@data$wtrtbl.dep)
pump@data$sat.cond.C <- pump@data$sat.cond - mean(pump@data$sat.cond)
pump@data$dep.to.restr.C <- pump@data$dep.to.restr - mean(pump@data$dep.to.restr)
pump.anom@data$Age.C <- pump.anom@data$Age - mean(pump.anom@data$Age)
pump.anom@data$dist.strm.C <- pump.anom@data$dist.strm - mean(pump.anom@data$dist.strm)
pump.anom@data$topo.wet.C <- pump.anom@data$topo.wet - mean(pump.anom@data$topo.wet)
pump.anom@data$wtrtbl.dep.C <- pump.anom@data$wtrtbl.dep - mean(pump.anom@data$wtrtbl.dep)
pump.anom@data$sat.cond.C <- pump.anom@data$sat.cond - mean(pump.anom@data$sat.cond)
pump.anom@data$dep.to.restr.C <- pump.anom@data$dep.to.restr - mean(pump.anom@data$dep.to.rest
r)

# Calculate correlations among predictor variables and display correlation matrices
(repr_spmn_R <- cor(repair@data[,c(33:38)], method="spearman"))

##                     Age.C dist.strm.C  topo.wet.C wtrtbl.dep.C  sat.cond.C
## Age.C           1.00000000  0.09246566  0.03721647  -0.00357821  0.04833696
## dist.strm.C     0.09246566  1.00000000  0.16365787   0.14160730  0.01247559
## topo.wet.C      0.03721647  0.16365787  1.00000000  -0.01319048  0.07364810
## wtrtbl.dep.C   -0.00357821  0.14160730 -0.01319048   1.00000000 -0.23582587
## sat.cond.C      0.04833696  0.01247559  0.07364810  -0.23582587  1.00000000
## dep.to.restr.C  0.05437435  0.09999574  0.08092591   0.05265809  0.62250946
##                dep.to.restr.C
## Age.C              0.05437435
## dist.strm.C        0.09999574
## topo.wet.C         0.08092591
## wtrtbl.dep.C       0.05265809
## sat.cond.C         0.62250946
## dep.to.restr.C     1.00000000

(pump_spmn_R <- cor(pump@data[,c(41:46)], method="spearman"))

##                      Age.C  dist.strm.C  topo.wet.C wtrtbl.dep.C   sat.cond.C
## Age.C           1.000000000  0.058378028  0.01240769 -0.005213071  0.049533773
## dist.strm.C     0.058378028  1.000000000  0.15408125  0.154928263 -0.004652192
## topo.wet.C      0.012407691  0.154081248  1.00000000 -0.013589027  0.075739536
## wtrtbl.dep.C   -0.005213071  0.154928263 -0.01358903  1.000000000 -0.199440454
## sat.cond.C      0.049533773 -0.004652192  0.07573954 -0.199440454  1.000000000
## dep.to.restr.C  0.056863774  0.085783875  0.08270727  0.061944760  0.623173379
##                dep.to.restr.C
## Age.C              0.05686377
## dist.strm.C        0.08578387
## topo.wet.C         0.08270727
## wtrtbl.dep.C       0.06194476
## sat.cond.C         0.62317338
## dep.to.restr.C     1.00000000
```

```r
(pump.a_spmn_R <- cor(pump.anom@data[,c(60:65)], method="spearman"))

##                     Age.C dist.strm.C   topo.wet.C wtrtbl.dep.C   sat.cond.C
## Age.C          1.000000000 0.110886952  0.004430343  0.024608380  0.007092071
## dist.strm.C    0.110886952 1.000000000  0.180596040  0.129442315  0.002743354
## topo.wet.C     0.004430343 0.180596040  1.000000000 -0.003316183  0.083171035
## wtrtbl.dep.C   0.024608380 0.129442315 -0.003316183  1.000000000 -0.205337531
## sat.cond.C     0.007092071 0.002743354  0.083171035 -0.205337531  1.000000000
## dep.to.restr.C 0.068011399 0.106130962  0.076144265  0.039096620  0.551456952
##                dep.to.restr.C
## Age.C              0.06801140
## dist.strm.C        0.10613096
## topo.wet.C         0.07614426
## wtrtbl.dep.C       0.03909662
## sat.cond.C         0.55145695
## dep.to.restr.C     1.00000000

# Run candidate models with centered predictor variables
##  Repaired
lr01_r <- glm(Repaired ~ Age.C, family=binomial(link="logit"), repair)
lr02_r <- glm(Repaired ~ dist.strm.C, family=binomial(link="logit"), repair)
lr03_r <- glm(Repaired ~ topo.wet.C, family=binomial(link="logit"), repair)
lr04_r <- glm(Repaired ~ Age.C + I(Age.C^2), family=binomial(link="logit"), repair)
lr05_r <- glm(Repaired ~ Age.C + dist.strm.C, family=binomial(link="logit"), repair)
lr06_r <- glm(Repaired ~ Age.C + topo.wet.C, family=binomial(link="logit"), repair)
lr07_r <- glm(Repaired ~ Age.C + dist.strm.C + topo.wet.C, family=binomial(link="logit"), repa
ir)
lr08_r <- glm(Repaired ~ Age.C + sat.cond.C + wtrtbl.dep.C +dep.to.restr.C, family=binomial(li
nk="logit"), repair)
lr09_r <- glm(Repaired ~ Age.C + sat.cond.C + wtrtbl.dep.C +dep.to.restr.C + topo.wet.C, famil
y=binomial(link="logit"), repair)
lr00_r <- glm(Repaired ~ 1, family=binomial(link="logit"), repair)
##  Pumped
lr01_p <- glm(Pumped ~ Age.C, family=binomial(link="logit"), pump)
lr02_p <- glm(Pumped ~ dist.strm.C, family=binomial(link="logit"), pump)
lr03_p <- glm(Pumped ~ topo.wet.C, family=binomial(link="logit"), pump)
lr04_p <- glm(Pumped ~ Age.C + dist.strm.C, family=binomial(link="logit"), pump)
lr05_p <- glm(Pumped ~ Age.C + topo.wet.C, family=binomial(link="logit"), pump)
lr06_p <- glm(Pumped ~ Age.C + dist.strm.C + topo.wet.C, family=binomial(link="logit"), pump)
lr07_p <- glm(Pumped ~ Age.C + sat.cond.C + wtrtbl.dep.C +dep.to.restr.C, family=binomial(link
="logit"), pump)
lr08_p <- glm(Pumped ~ Age.C + sat.cond.C + wtrtbl.dep.C +dep.to.restr.C + topo.wet.C, family=
binomial(link="logit"), pump)
lr00_p <- glm(Pumped ~ 1, family=binomial(link="logit"), pump)
## Anomalous Pumping
###  Pumping frequency (>=2 pumps in 3 year period)
lr01_p_a_f <- glm(Pump_Freq ~ Age.C, family=binomial(link="logit"), pump.anom)
lr02_p_a_f <- glm(Pump_Freq ~ dist.strm.C, family=binomial(link="logit"), pump.anom)
lr03_p_a_f <- glm(Pump_Freq ~ topo.wet.C, family=binomial(link="logit"), pump.anom)
lr04_p_a_f <- glm(Pump_Freq ~ Age.C + I(Age.C^2), family=binomial(link="logit"), pump.anom)
lr05_p_a_f <- glm(Pump_Freq ~ Age.C + dist.strm.C, family=binomial(link="logit"), pump.anom)
lr06_p_a_f <- glm(Pump_Freq ~ Age.C + topo.wet.C, family=binomial(link="logit"), pump.anom)
lr07_p_a_f <- glm(Pump_Freq ~ Age.C + dist.strm.C + topo.wet.C, family=binomial(link="logit"),
pump.anom)
lr08_p_a_f <- glm(Pump_Freq ~ Age.C + sat.cond.C + wtrtbl.dep.C +dep.to.restr.C, family=binomi
al(link="logit"), pump.anom)
lr09_p_a_f <- glm(Pump_Freq ~ Age.C + sat.cond.C + wtrtbl.dep.C +dep.to.restr.C + topo.wet.C,
family=binomial(link="logit"), pump.anom)

lr00_p_a_f <- glm(Pump_Freq ~ 1, family=binomial(link="logit"), pump.anom)
###  Pumping volume (manifest recorded volume >= OWTS capacity)
```

```r
lr01_p_a_v <- glm(Max_Over_Cap ~ Age.C, family=binomial(link="logit"), pump.anom)
lr02_p_a_v <- glm(Max_Over_Cap ~ dist.strm.C, family=binomial(link="logit"), pump.anom)
lr03_p_a_v <- glm(Max_Over_Cap ~ topo.wet.C, family=binomial(link="logit"), pump.anom)
lr04_p_a_v <- glm(Max_Over_Cap ~ Age.C + I(Age.C^2), family=binomial(link="logit"), pump.anom)
lr05_p_a_v <- glm(Max_Over_Cap ~ Age.C + dist.strm.C, family=binomial(link="logit"), pump.anom
)
lr06_p_a_v <- glm(Max_Over_Cap ~ Age.C + topo.wet.C, family=binomial(link="logit"), pump.anom)
lr07_p_a_v <- glm(Max_Over_Cap ~ Age.C + dist.strm.C + topo.wet.C, family=binomial(link="logit
"), pump.anom)
lr08_p_a_v <- glm(Max_Over_Cap ~ Age.C + sat.cond.C + wtrtbl.dep.C +dep.to.restr.C, family=bin
omial(link="logit"), pump.anom)
lr09_p_a_v <- glm(Max_Over_Cap ~ Age.C + sat.cond.C + wtrtbl.dep.C +dep.to.restr.C + topo.wet.
C, family=binomial(link="logit"), pump.anom)
lr00_p_a_v <- glm(Max_Over_Cap ~ 1, family=binomial(link="logit"), pump.anom)
###  Any anomalous pumping (either frequently pumped OR volume exceedance)
lr01_p_a_a <- glm(Pump_Anom ~ Age.C, family=binomial(link="logit"), pump.anom)
lr02_p_a_a <- glm(Pump_Anom ~ dist.strm.C, family=binomial(link="logit"), pump.anom)
lr03_p_a_a <- glm(Pump_Anom ~ topo.wet.C, family=binomial(link="logit"), pump.anom)
lr04_p_a_a <- glm(Pump_Anom ~ Age.C + I(Age.C^2), family=binomial(link="logit"), pump.anom)
lr05_p_a_a <- glm(Pump_Anom ~ Age.C + dist.strm.C, family=binomial(link="logit"), pump.anom)
lr06_p_a_a <- glm(Pump_Anom ~ Age.C + topo.wet.C, family=binomial(link="logit"), pump.anom)
lr07_p_a_a <- glm(Pump_Anom ~ Age.C + dist.strm.C + topo.wet.C, family=binomial(link="logit"),
pump.anom)
lr08_p_a_a <- glm(Pump_Anom ~ Age.C + sat.cond.C + wtrtbl.dep.C + dep.to.restr.C, family=binom
ial(link="logit"), pump.anom)
lr09_p_a_a <- glm(Pump_Anom ~ Age.C + sat.cond.C + wtrtbl.dep.C +dep.to.restr.C + topo.wet.C,
family=binomial(link="logit"), pump.anom)
lr00_p_a_a <- glm(Pump_Anom ~ 1, family=binomial(link="logit"), pump.anom)

# Create lists of all the models and name them
repr_Mdls <- list(lr01_r, lr02_r, lr03_r, lr04_r, lr05_r, lr06_r, lr07_r, lr08_r, lr09_r, lr00
_r)
pump_Mdls <- list(lr01_p, lr02_p, lr03_p, lr04_p, lr05_p, lr06_p, lr07_p, lr08_p, lr00_p)
pump_anom_freq_Mdls <- list(lr01_p_a_f, lr02_p_a_f, lr03_p_a_f, lr04_p_a_f, lr05_p_a_f, lr06_p
_a_f, lr07_p_a_f, lr08_p_a_f, lr09_p_a_f, lr00_p_a_f)
pump_anom_vol_Mdls <- list(lr01_p_a_v, lr02_p_a_v, lr03_p_a_v, lr04_p_a_v, lr05_p_a_v, lr06_p_
a_v, lr07_p_a_v, lr08_p_a_v, lr09_p_a_v, lr00_p_a_v)
pump_anom_all_Mdls <- list(lr01_p_a_a, lr02_p_a_a, lr03_p_a_a, lr04_p_a_a, lr05_p_a_a, lr06_p_
a_a, lr07_p_a_a, lr08_p_a_a, lr09_p_a_a, lr00_p_a_a)

lrNames_r <- c("Age","Strm.Dist","TWI","Age.Quad","Age.Strm", "Age.TWI", "Age.Fld","Age.Soil",
"Age.Soil.TWI","Null")
lrNames_p <- c("Age","Strm.Dist","TWI","Age.Strm","Age.TWI","Age.Fld","Age.Soil","Age.Soil.TWI
","Null")
lrNames_p_a_f <- c("Age","Strm.Dist","TWI","Age.Quad","Age.Strm", "Age.TWI", "Age.Fld","Age.So
il","Age.Soil.TWI","Null")
lrNames_p_a_v <- c("Age","Strm.Dist","TWI","Age.Quad","Age.Strm", "Age.TWI", "Age.Fld","Age.So
il","Age.Soil.TWI","Null")
lrNames_p_a_a <- c("Age","Strm.Dist","TWI","Age.Quad","Age.Strm", "Age.TWI", "Age.Fld","Age.So
il","Age.Soil.TWI","Null")


# Calculate and print BIC tables
(bicWt_r <- bictab(cand.set = repr_Mdls, modnames = lrNames_r, sort = TRUE, c.hat=1))

##
## Model selection based on BIC:
##
##                K     BIC Delta_BIC BICWt Cum.Wt       LL
## Age            2 4805.58      0.00  0.81   0.81 -2393.71
## Age.Strm       3 4808.82      3.24  0.16   0.97 -2390.79
```

```
## Age.Quad      3 4813.82      8.24  0.01   0.99 -2393.29
## Age.TWI       3 4814.45      8.87  0.01   1.00 -2393.60
## Age.Fld       4 4817.16     11.58  0.00   1.00 -2390.42
## Age.Soil      5 4831.21     25.63  0.00   1.00 -2392.90
## Age.Soil.TWI  6 4840.05     34.47  0.00   1.00 -2392.78
## Strm.Dist     2 4843.60     38.02  0.00   1.00 -2412.72
## Null          1 4844.44     38.86  0.00   1.00 -2417.68
## TWI           2 4853.51     47.93  0.00   1.00 -2417.67
```

```
(bicWt_p <- bictab(cand.set = pump_Mdls, modnames = lrNames_p, sort = TRUE, c.hat=1))
```

```
##
## Model selection based on BIC:
##
##                K     BIC Delta_BIC BICWt Cum.Wt       LL
## Null           1 6234.51      0.00  0.90   0.90 -3112.78
## Strm.Dist      2 6239.28      4.77  0.08   0.98 -3110.69
## Age            2 6243.43      8.92  0.01   0.99 -3112.77
## TWI            2 6243.45      8.95  0.01   1.00 -3112.78
## Age.Strm       3 6248.15     13.65  0.00   1.00 -3110.66
## Age.TWI        3 6252.37     17.86  0.00   1.00 -3112.77
## Age.Fld        4 6257.03     22.52  0.00   1.00 -3110.62
## Age.Soil       5 6266.90     32.39  0.00   1.00 -3111.09
## Age.Soil.TWI   6 6275.85     41.34  0.00   1.00 -3111.09
```

```
(bicWt_p_a_f <- bictab(cand.set=pump_anom_freq_Mdls, modnames=lrNames_p_a_f, sort=TRUE, c.hat=
1))
```

```
##
## Model selection based on BIC:
##
##                K     BIC Delta_BIC BICWt Cum.Wt      LL
## Null           1 1069.27      0.00  0.81   0.81 -531.15
## Age            2 1073.54      4.27  0.10   0.90 -529.79
## TWI            2 1075.11      5.84  0.04   0.95 -530.57
## Strm.Dist      2 1075.67      6.39  0.03   0.98 -530.85
## Age.Quad       3 1078.25      8.97  0.01   0.99 -528.65
## Age.TWI        3 1079.23      9.96  0.01   1.00 -529.14
## Age.Strm       3 1080.17     10.90  0.00   1.00 -529.61
## Age.Fld        4 1086.04     16.77  0.00   1.00 -529.06
## Age.Soil       5 1089.36     20.09  0.00   1.00 -527.23
## Age.Soil.TWI   6 1094.87     25.60  0.00   1.00 -526.49
```

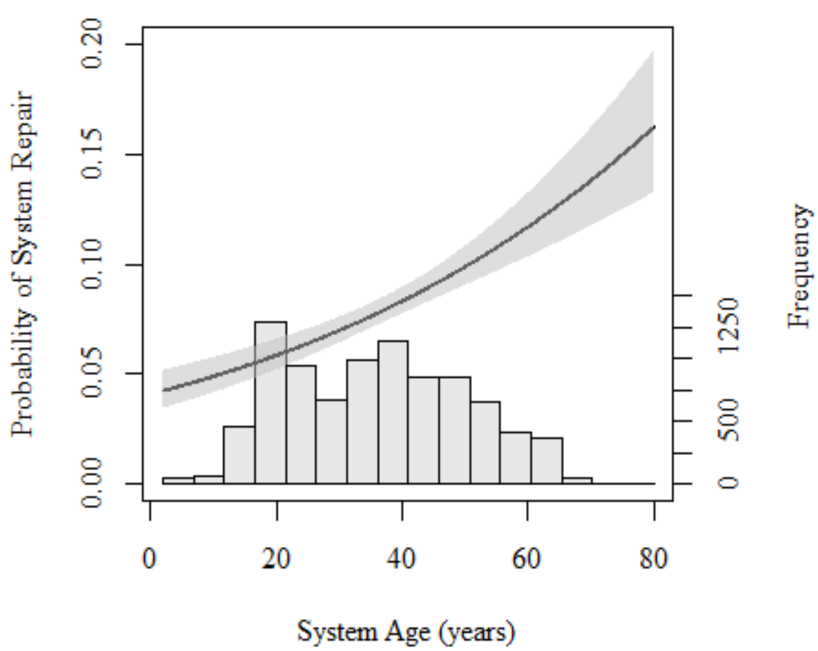
```
(bicWt_p_a_v <- bictab(cand.set=pump_anom_vol_Mdls, modnames=lrNames_p_a_v, sort=TRUE, c.hat=1
))
```

```
##
## Model selection based on BIC:
##
##                K     BIC Delta_BIC BICWt Cum.Wt      LL
## Age.Quad       3 1483.83      0.00  0.99   0.99 -731.45
## Null           1 1493.26      9.43  0.01   1.00 -743.14
## Strm.Dist      2 1496.85     13.01  0.00   1.00 -741.44
## TWI            2 1499.22     15.39  0.00   1.00 -742.63
## Age            2 1500.07     16.23  0.00   1.00 -743.05
## Age.Strm       3 1503.78     19.95  0.00   1.00 -741.42
## Age.TWI        3 1505.99     22.16  0.00   1.00 -742.52
## Age.Fld        4 1510.18     26.35  0.00   1.00 -741.13
## Age.Soil       5 1513.27     29.44  0.00   1.00 -739.18
## Age.Soil.TWI   6 1519.44     35.61  0.00   1.00 -738.78
```

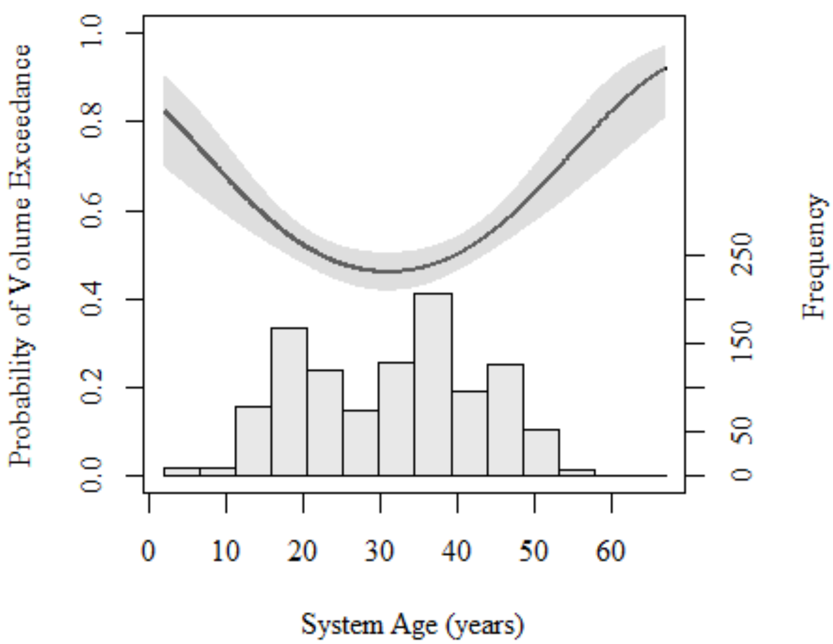
```r
(bicWt_p_a_a <- bictab(cand.set=pump_anom_all_Mdls, modnames=lrNames_p_a_a, sort=TRUE, c.hat=1
))

##
## Model selection based on BIC:
##
##              K     BIC Delta_BIC BICWt Cum.Wt      LL
## Null         1 1461.24      0.00  0.69   0.69 -727.13
## Strm.Dist    2 1464.60      3.35  0.13   0.82 -725.32
## Age.Quad     3 1464.76      3.52  0.12   0.93 -721.91
## TWI          2 1467.35      6.11  0.03   0.97 -726.69
## Age          2 1467.70      6.46  0.03   0.99 -726.87
## Age.Strm     3 1471.30     10.06  0.00   1.00 -725.18
## Age.TWI      3 1473.75     12.51  0.00   1.00 -726.41
## Age.Fld      4 1477.81     16.57  0.00   1.00 -724.94
## Age.Soil     5 1485.98     24.73  0.00   1.00 -725.54
## Age.Soil.TWI 6 1492.10     30.86  0.00   1.00 -725.11

# Inspect top models
summary(lr01_r)

##
## Call:
## glm(formula = Repaired ~ Age.C, family = binomial(link = "logit"),
##     data = repair)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5950  -0.4359  -0.3909  -0.3470   2.5158
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.494128   0.040813 -61.112  < 2e-16 ***
## Age.C        0.018969   0.002748   6.903 5.09e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4835.4  on 8785  degrees of freedom
## Residual deviance: 4787.4  on 8784  degrees of freedom

## AIC: 4791.4
##
## Number of Fisher Scoring iterations: 5

summary(lr04_p_a_v)

##
## Call:
## glm(formula = Max_Over_Cap ~ Age.C + I(Age.C^2), family = binomial(link = "logit"),
##     data = pump.anom)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8648  -1.1687   0.8856   1.1456   1.2456
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.1571172  0.0879798  -1.786   0.0741 .
## Age.C        0.0037342  0.0051816   0.721   0.4711
```

```
## I(Age.C^2)   0.0020253  0.0004388   4.616 3.91e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1486.3  on 1075  degrees of freedom
## Residual deviance: 1462.9  on 1073  degrees of freedom
## AIC: 1468.9
##
## Number of Fisher Scoring iterations: 4

# Check 90% confidence intervals to see if they cross zero
confint(lr01_r, level = 0.90)

## Waiting for profiling to be done...

##                    5 %        95 %
## (Intercept) -2.5619450 -2.42766361
## Age.C        0.0144546  0.02349611

confint(lr04_p_a_v, level = 0.90)

## Waiting for profiling to be done...

##                      5 %         95 %
## (Intercept) -0.302362123 -0.012838271
## Age.C       -0.004779256  0.012275578
## I(Age.C^2)   0.001315311  0.002759602

# Create new data tables to make predictions to based on top models
## Repair (Age)
n_rep_dat <- data.frame(Age.C = seq(min(repair@data$Age.C), max(repair@data$Age.C), length = 1
000))
Link_01_r <- predict(lr01_r, newdata = n_rep_dat, type="link", se.fit = TRUE)

n_rep_dat$Age <- seq(min(repair@data$Age), max(repair@data$Age), length = 1000)
n_rep_dat$p <- plogis(Link_01_r$fit)
n_rep_dat$lwr <- plogis(Link_01_r$fit - 1.96*Link_01_r$se.fit)
n_rep_dat$upr <- plogis(Link_01_r$fit + 1.96*Link_01_r$se.fit)

## Anomalous Pumping: Volume
n_vol_dat <- data.frame(Age.C=seq(min(pump.anom@data$Age.C), max(pump.anom@data$Age.C), length
=1000))
Link_04_vol <- predict(lr04_p_a_v, newdata=n_vol_dat, se.fit=TRUE, type="link")
n_vol_dat$Age <- seq(min(pump.anom@data$Age), max(pump.anom@data$Age), length=1000)
n_vol_dat$p <- plogis(Link_04_vol$fit)
n_vol_dat$lwr <- plogis(Link_04_vol$fit - 1.96*Link_04_vol$se.fit)
n_vol_dat$upr <- plogis(Link_04_vol$fit + 1.96*Link_04_vol$se.fit)

# AUC calculations
library(ROCR)

# Repair Model
## Age (lr01_r)
# Create predictions
prob_lr01_r <- predict(lr01_r, repair@data, se.fit = TRUE, type = "response")
pred_lr01_r <- prediction(prob_lr01_r[1], repair@data$Repaired)

# Calculate and print area under the curve
```

```r
auc_lr01_r <- performance(pred_lr01_r, "auc")
auc_lr01_r@y.values

## [[1]]
## [1] 0.5892685

# Pumping Models
## Volume Exceedance
### Age + Age^2 (lr04_p_a_v)
# Create predictions
prob_lr04_p_a_v <- predict(lr04_p_a_v, pump.anom@data, se.fit = TRUE, type = "response")
pred_lr04_p_a_v <- prediction(prob_lr04_p_a_v[1], pump.anom@data$Max_Over_Cap)

# Calculate and print area under the curve
auc_lr04_p_a_v <- performance(pred_lr04_p_a_v, "auc")
auc_lr04_p_a_v@y.values

## [[1]]
## [1] 0.5759514

# Plot predicted probabilities
windowsFonts(TNR = windowsFont("Times New Roman"))
## Repair
par(mar = c(5,5,2,5))
hist(repair@data$Age, main=NA, axes=F, xlab=NA, ylab=NA, ylim=range(0,3500),
     col =  adjustcolor("grey", alpha.f = 0.35))
axis(side = 4, family="TNR", seq(0,1500,250))
par(new = TRUE)
plot(p ~ Age, data = n_rep_dat, type="l", lwd = 2.5, xlab="System Age (years)", ylab="Probabil
ity of System Repair", family="TNR", ylim=c(0.0,0.2))
mtext(side = 4, line = 3, family="TNR", "Frequency")
polygon(x = c(n_rep_dat$Age, rev(n_rep_dat$Age)),
        y = c(n_rep_dat$lwr, rev(n_rep_dat$upr)),
        col =  adjustcolor("grey", alpha.f = 0.5), border = NA)
```



```r
## Anomalous pumping: Volume
par(mar = c(5,5,2,5))
hist(pump.anom@data$Age, main=NA, axes=F, xlab=NA, ylab=NA, ylim=range(0,500),
     col =  adjustcolor("grey", alpha.f = 0.35))
axis(side = 4, family="TNR", seq(0,250,50))
par(new = TRUE)
plot(p ~ Age, data = n_vol_dat, xlab="System Age (years)", ylab="Probability of Volume Exceeda
nce", family="TNR", ylim=c(0,1), type = "l", lwd = 2.5)
mtext(side = 4, line = 3, family="TNR", "Frequency")
polygon(x = c(n_vol_dat$Age, rev(n_vol_dat$Age)),
        y = c(n_vol_dat$lwr, rev(n_vol_dat$upr)),
        col =  adjustcolor("grey", alpha.f = 0.5), border = NA)
```




```r
# Regressions between OWTS repaired within 5 years of pumping records and those with any anoma
lous pumping
# Subset anomalous pumping SpatialPointsDataFrame by sites which have been repaired since 01/0
1/2012 (5 years prior to the earliest pumping record)
pump.anom.rep <- sp.na.omit(pump.anom, col.name = "Repaired5")

# Calculate correlations among variables and display correlation matrix
(pump.a.r_spmn_R <- cor(pump.anom.rep@data[,c(48,49,52)], method="spearman"))

##                Pump_Anom Max_Over_Cap  Repaired5
## Pump_Anom     1.00000000    0.8247245 0.06139246
## Max_Over_Cap 0.82472455    1.0000000 0.16376789
## Repaired5     0.06139246    0.1637679 1.00000000

# Run logistic regression models between binary anomalous pumping occurrence and repair occurr
ence
lr01_p_a_r <- glm(Repaired5 ~ Pump_Anom, family=binomial(link="logit"), pump.anom.rep)
lr02_p_a_r <- glm(Repaired5 ~ Max_Over_Cap, family=binomial(link="logit"), pump.anom.rep)

# Summarize models
summary(lr01_p_a_r)
```

```
##
## Call:
## glm(formula = Repaired5 ~ Pump_Anom, family = binomial(link = "logit"),
##     data = pump.anom.rep)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7910  -0.7910  -0.6905  -0.1126   1.7610
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.3122     0.4258  -3.082  0.00206 **
## Pump_Anom     0.3107     0.5072   0.613  0.54013
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 112.47  on 99  degrees of freedom
## Residual deviance: 112.08  on 98  degrees of freedom
## AIC: 116.08
##
## Number of Fisher Scoring iterations: 4

summary(lr02_p_a_r)

##
## Call:
## glm(formula = Repaired5 ~ Max_Over_Cap, family = binomial(link = "logit"),
##     data = pump.anom.rep)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.86205  -0.86205  -0.60386  -0.07045   1.89302
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.6094     0.4140  -3.887 0.000101 ***
## Max_Over_Cap   0.8109     0.5020   1.615 0.106211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 112.47  on 99  degrees of freedom
## Residual deviance: 109.69  on 98  degrees of freedom
## AIC: 113.69
##
## Number of Fisher Scoring iterations: 4

# Summary statistics
median(repair1@data$Age)

## [1] 35

median(repair@data$Age)

## [1] 35
```

```
with(repair@data, median(Age[Repaired = 1]))

## [1] 65

median(pump@data$Age)

## [1] 33

with(pump@data, median(Age[Pumped == 1]))

## [1] 34

sum(pump@data$FREQUENCY==0)

## [1] 6599

sum(pump@data$FREQUENCY>=2)

## [1] 211

mean(pump@data$FREQUENCY)

## [1] 0.2040125

with(pump@data, mean(FREQUENCY[FREQUENCY >= 1]))

## [1] 1.454039

sum(c(pump.anom.rep$Repaired5 == 1 & pump.anom.rep$Max_Over_Cap == 1))

## [1] 18
```