Climate Risk Management 40 (2023) 100499

Contents lists available at ScienceDirect

Climate Risk Management

journal homepage: www.elsevier.com/locate/crm

Predicting groundwater contamination to protect the storm-exposed vulnerable

Jacob Hochard [a,*], Nino Abashidze [a], Ranjit Bawa [b], Grace Carr [a], Bailey Kirkland [c], Yuanhao Li [d], Kayla Matlock [a], Wai Yan Siu [a]

[a] University of Wyoming, Haub School of Environment and Natural Resources, Laramie, WY 82072, United States
[b] University of New Hampshire, Department of Natural Resources & the Environment, Durham, NH 03824, United States
[c] University of Wyoming, Department of Economics, Laramie, WY 82072, United States
[d] SNF – Centre for Applied Research, Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway

ARTICLE INFO

Keywords:
Groundwater contamination
Predictive analysis
Vulnerable populations
Flooding events
Contaminated sites

ABSTRACT

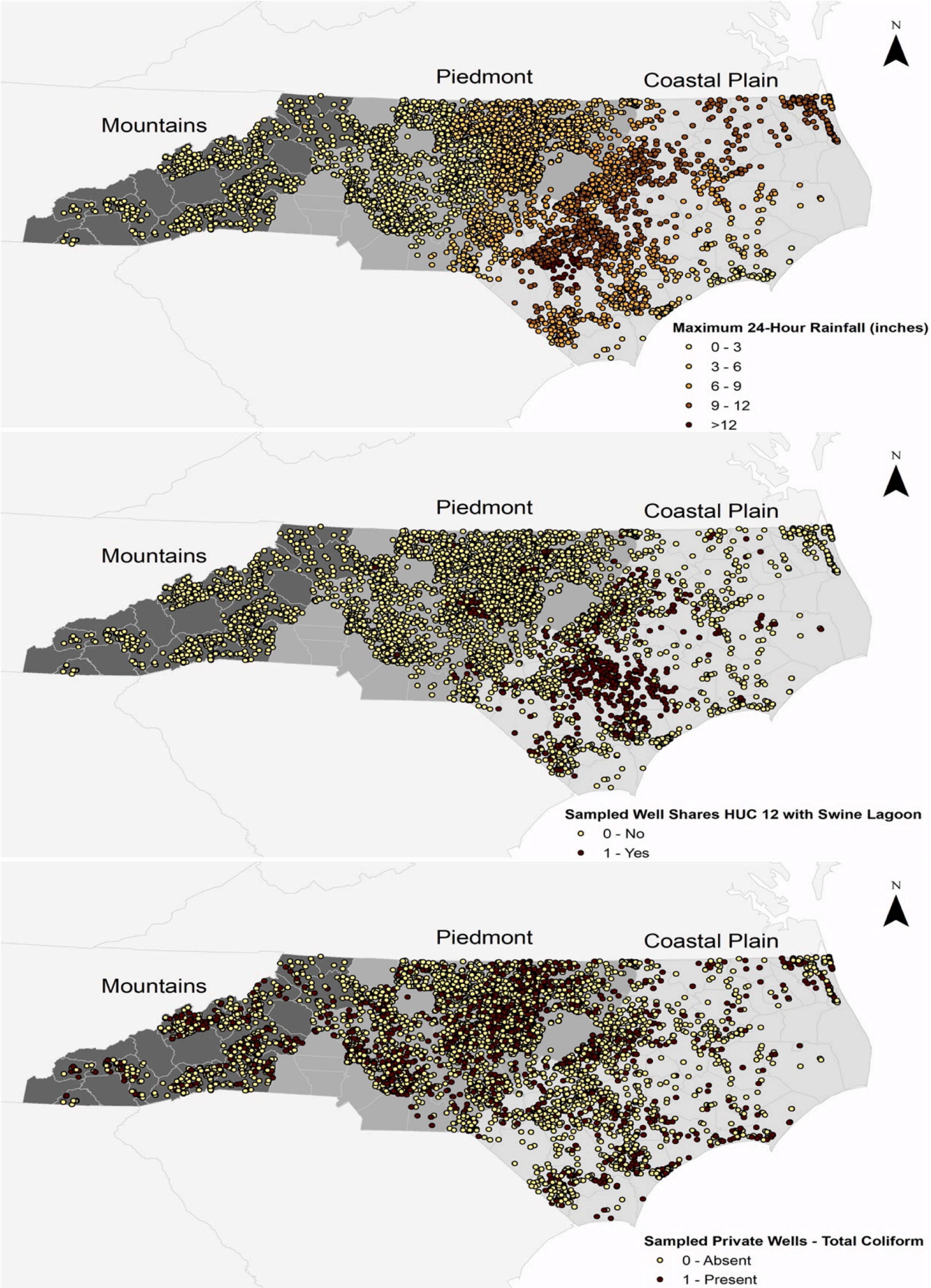
Domestic wells provide drinking water to 44 million people nationwide. Many of these wells, which remain federally unregulated and rarely tested for pollutants, serve rural populations clustered near surface-contaminated sites (e.g., hazardous waste sites, animal agriculture operations, coal ash ponds, etc.). The potential for natural disasters to deteriorate drinking water quality is well documented. Less understood is whether opportunistic post-disaster sampling might underrepresent vulnerable populations. When disaster strikes, well water sampling campaigns offer a glimpse into the quality of water for exposed residents. We examined over 8,000 well water samples from 2016 and 2017 to measure Hurricane Matthew's impact on the presence of indicator bacteria. Bacteria presence was predicted at the household level following Hurricane Matthew's landfall. The residential addresses associated with birth records as well as clinically estimated dates of conception and birth dates were used to predict the likelihood of indicator bacteria in drinking water sources that were unsampled but likely to have served pregnant women. We estimate that opportunistic well water sampling captures the average predicted contamination rates among households with pregnant women. Our approach documents a distribution of contamination risk where 2.7% of the vulnerable sample (670 unsampled households) have a 75% likelihood of total coliform presence. The predicted likelihood of indicator bacteria is elevated for a small share of households nearby swine lagoons that experienced the most torrential rainfall. However, the gap between sampled and unsampled households cannot otherwise be explained by the storm event or proximity to surface-contaminated sites. Findings suggest that sophisticated and holistic water quality prediction models may support post-disaster sampling campaigns by targeting individual households within vulnerable groups that are likely to experience higher risks from groundwater contamination.

1. Introduction

Most households in the United States source their drinking water from public water systems (PWS), which, under the Safe Drinking Water Act of 1974, are subject to federal regulation and required testing for over 90 different waterborne contaminants (Allaire et al.,

* Corresponding author.
E-mail address: JHochard@uwyo.edu (J. Hochard).

https://doi.org/10.1016/j.crm.2023.100499
Received 1 January 2023; Received in revised form 15 February 2023; Accepted 15 March 2023
Available online 22 March 2023



(caption on next page)

Fig. 1. Maximum 24-hour rainfall during Hurricane Matthew, total coliform outcomes and swine lagoon relationships with all residential addresses that received well water testing and were included in our baseline sample. Counties are shown by outline and shades represent the three physiographic provinces of North Carolina (mountains, piedmont and coastal plain). Wells sampled up to 280 days before or after Hurricane Matthew, which equals the generally accepted expected pregnancy duration, are included in the baseline sample. A total of 8,146 observations are included in the baseline sample and represented as points precise to the residential address level on each map. Points are intentionally kept large to preserve viewability and the privacy of households represented in the sample.

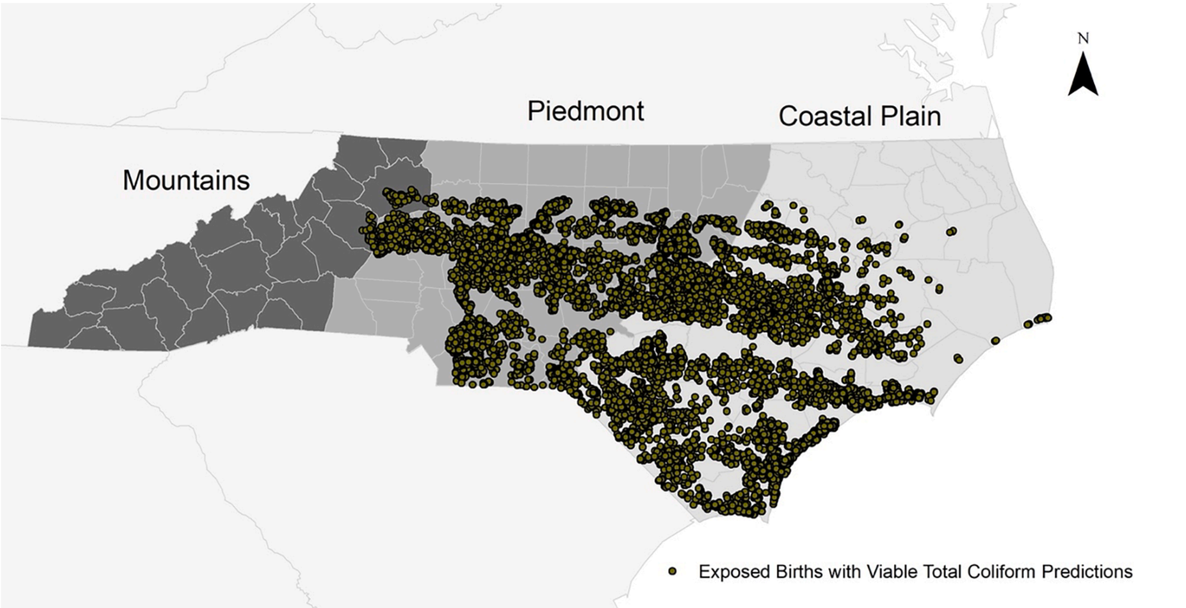


Fig. 2. Births occurring between October 8th 2016 and July 15th 2017 (n = 24,493) that are candidates for total coliform contamination of groundwater and exist in locations (e.g., zip codes, counties, etc.) where the empirical groundwater contamination model can make viable predictions. Note: Birth records are comprehensive for the relevant period and are not included based on the water within these households having been sampled. The purpose of focusing on these households is to predict potential groundwater contamination events for vulnerable households that did not necessarily receive testing. Inspection of our data reveals that < 1% of households occupied by pregnant women sampled tested their water within our sample window.

2018). In rural areas, where sparse populations are unable to justify public investments in water transmission, treatment, and distribution, households depend on private drinking water wells. Such wells serve 44 million people nationwide and remain federally unregulated, leaving the onus of well water testing on homeowners (Schaider et al., 2016). Despite private wells often being shallower and more vulnerable to contamination than those wells serving PWS (Schaider et al., 2016), routine testing consistent with federal guidelines is rare (Jones et al., 2005; Lewandowski et al., 2008; Malecki et al., 2017; Colley et al., 2019). There are several cases that do trigger widespread testing of private wells for contaminants, including flooding events (Dai et al., 2019; Gilliland et al., 2021; Mapili et al., 2022), observed changes in the smell and taste of water (Colley et al., 2019; Stillo et al., 2019; Murti et al., 2016) and discovery of groundwater "contamination plumes" (Bräunig et al., 2017; Newell et al., 2020; Loch-Caruso et al., 2022). Yet, such testing campaigns tend to adopt opportunistic sampling approaches, which may not represent vulnerable populations' risks associated with these environmental exposures. Herein we leverage the uptick in private well water tests following Hurricane Matthew's 2016 North Carolina landfall to predict groundwater contamination risks for exposed pregnant women whose drinking water was unlikely to be tested for contaminants.

Convenience, or opportunistic, well water sampling campaigns are common practice (Jones et al., 2006; Invik et al., 2017; Gilliland et al., 2021), especially when there is a suspected change in the likelihood of groundwater contamination. While these samples can be "checked" for alignment with general population demographics (Gilliland et al., 2021), the ability of such an approach to represent the most vulnerable groups of citizens remains unclear. For example, survey and sampling campaigns attempting to achieve broad enrollment often recruit participants from community organizations, such as churches (Heaney et al., 2013) and town halls (Jones et al., 2006). However, if pregnant women vulnerable to *in utero* environmental exposures are the target group of interest (Hochard et al., 2022; Hochard et al., 2023), these recruitment venues might be less representative of the target population than other locations, such as maternity wards or prenatal care clinics. Expanding environmental sampling efforts to target vulnerable populations is often cost prohibitive or impractical. For private wells, which tend to be in rural and sparsely populated areas, identifying vulnerable populations, and encouraging their widespread enrollment into such campaigns is challenging if not impractical.

The high cost of environmental monitoring in rural areas has led to calls for passive surveillance (Rahman et al., 2021; Vadapalli et al., 2020). Here, the promise is that disparate but relevant primary and secondary datasets, when analyzed together, contain enough predictive power to guide interventions where they might be most effective. Such approaches have gained traction for applications in environmental enforcement - e.g., identifying water pollution discharge violations (Hino et al., 2018); detecting poachers (de Knegt et al., 2021); discovering illegal deforestation operations (Vorotyntsev et al., 2021). A natural extension of rural environmental monitoring is the prediction of groundwater and drinking water contamination in rural areas that may otherwise be excluded or underrepresented in direct sampling campaigns. Such predictive models show promise in detecting arsenic (Winkel et al., 2008; Podgorski and Berg 2020, Erickson et al., 2021), microbiological (O'Dwyer et al., 2018; Díaz-Alcaide, and Martínez-Santos 2019), manganese (Erickson et al., 2021), nitrate (Knoll et al., 2019, Rahmati et al., 2019) and pesticides (Teso et al., 1996) in groundwater. Despite advances in the predictive power of groundwater contamination modeling, rarely are these models used in a household-level assessment of exposure risks to vulnerable groups.

1.1. Hurricane Matthew's exposure and data

We focus on North Carolina where Hurricane Matthew covered the state's coastal plain with as much as 16 in. of rain on October 8–9, 2016. The exposed region, which is home to the nation's two largest pork producing counties of Sampson and Duplin, NC, experienced devastating flooding with total statewide damages exceeding $1.5 billion. Ranking third nationally, North Carolina is a notable study area with >2.1 million residents, or approximately 22.7% of the state's population, depending on private drinking wells (Johnson et al., 2019). North Carolina's coastal plain aquifers are also particularly shallow (Winner and Coble 1996) where vulnerability hotspots arise when private wells source drinking water from the surficial aquifer that is less protected from surface sources of contaminants (Nolan and Hitt 2006).

The analysis centers on Hurricane Matthew's North Carolina rainfall exposures as estimated by Parameter-elevation Regressions on Independent Slopes Model data. (PRISM Climate Group, 2020). The most intensive rainfall (12–16 in.) was experienced by residents east of the state's coastal plain-piedmont interface (Fig. 1), which contains a disproportionate number of the swine lagoons associated with confined animal feeding operations (CAFOs). Here, hydrological proximity to a CAFO was based on whether the well was sampled within the same hydrological unit code 12 (HUC 12) of at least one lagoon (Fig. 1.). Of the 8,146 well water samples included in the analysis, 10.1% were taken from wells that shared a HUC 12 with a lagoon (see Fig. 2).

Total coliform presence or absence is a useful and commonly accepted (Dai et al., 2019; Pieper et al., 2021; Mapili et al., 2022) indicator for microbiological contamination of groundwater. However, because the presence of total coliform is common and stems from diverse sources (e.g., leaking septic tanks, migrating or domestic animals, etc.), its mere presence alone is not enough to draw a linkage between groundwater contamination and specific nearby surface-contaminated sites, such as animal agriculture. Following Hochard et al. (2023), we control for a variety of surface contaminated sites that might also lead to contamination of nearby private drinking water during storm events. These alternative contaminated sites are georeferenced at the address level and include poultry farms (Formuzis, 2016), hazardous waste sites (NC DEQ, 2016) and dry cleaner facilities (NC DEQ, 2019).

Comprehensive birth records are obtained from the North Carolina Department of Health and Human Services (NC DHHS) vital statistics dataset. We focus on the 280-day window of births that occurred between October 8, 2016, and July 15, 2017. While the precision of this window is not critical for our analysis, a standard pregnancy term (Persson and Rossin-Slater, 2018) that began when Hurricane Matthew made North Carolina landfall will conservatively identify the residential addresses associated with known pregnancies and an elevated risk of water contamination. Henceforth we call this window a "selection into sample" criteria. To ensure that our empirical predictions of groundwater contamination are relevant for our out-of-sample pregnant populations, we use the same selection in the sample window to identify the 8,146 samples that are used within this analysis. We predict total coliform contamination likelihoods for 24,493 such residences known to be occupied by pregnant women.

1.2. Empirical strategy

We extend the empirical approach used by Hochard et al. (2022) to predict a linear probability model. In its most parsimonious form, we estimate:

$$I(bacteria = 1)_{i,c,z,m} = \beta_o + \beta_1 E_{i,c,z,m} + \beta_2 \ln(R_{i,c,z}) + \beta_3 \mathrm{E}_{\mathrm{i,c,z,m}} \ln(R_{i,c,z}) + \beta_4 H_{i,c,z} + \beta_5 H_{i,c,z} \mathrm{E}_{\mathrm{i,c,z,m}} + \beta_6 H_{i,c,z} \mathrm{E}_{\mathrm{i,c,z,m}} \ln(R_{i,c,z}) + \gamma_c + \zeta_z + \eta_m + \epsilon_{i,c,z,m}$$

where $I(\cdot)$ is an indicator variable equal to 1 if total coliform bacteria are detected in sample i located in county c and *zip* code z, which was collected during month m. E represents a selection into sample exposure variable equal to 1 if the sample was collected within 280 days following Hurricane Matthew's North Carolina landfall and 0 if the sample was collected in the 280 days prior. Here, the exposure variable leverages Hurricane Matthew as a natural experiment where water samples collected within the same counties, zip codes, and months serve as a baseline of comparison for samples that were "treated" by Hurricane Matthew. The intensity of Hurricane Matthew's exposure for a given sample is measured using the natural log of the maximum rainfall experienced during the storm within a 24-hour period at the sample location ($R_{i,c,z,}$). The variable $H_{i,c,z}$ equals 1 when sample i was taken from a HUC 12 watershed occupied by at least 1 swine lagoon and 0 otherwise.

The regression intercept β_0 captures a common background rate of total coliform presence across all sampled wells and β_1 adjusts this rate for any systematic difference in total coliform presence rates unique to post-Hurricane Matthew samples, but unrelated to rainfall activity. The coefficient β_2 serves as a control factor. Because $\ln(R_{i,c,z})$ is uninteracted with the exposure variable, $\mathrm{E}_{\mathrm{i,c,z,m}}$, β_2 detects whether total coliform presence in wells is spatially correlated with areas that were exposed to Hurricane Matthew's rainfall – i.

Table 1
Ordinary least squares (OLS) and probit estimation of Hurricane Matthew on total coliform presence in sampled wells. Regressions are presented including and excluding alternative contaminated sites (dry cleaning solvent, hazardous waste sites and poultry operations) and for a full and half "selection into sample" window of 280 days and 140 days following the storm's North Carolina landfall on October 8, 2016.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	Probit	Probit	Probit
Dependent Variable: Binary Total Coliform Sample Presence (=1)						
Exp.	0.0180	0.0131	−0.0651	0.0520	0.0372	−0.2025
	(0.0211)	(0.0225)	(0.0650)	(0.0650)	(0.0690)	(0.1844)
Ln (rain)	0.0175	0.0164	0.0631	0.0562	0.0538	0.2221
	(0.0433)	(0.0440)	(0.0742)	(0.1309)	(0.1332)	(0.2119)
$Exp.Ln(rain)$	−0.0079	−0.0069	−0.0010	−0.0219	−0.0190	−0.0154
	(0.0143)	(0.0157)	(0.0224)	(0.0450)	(0.1332)	(0.0717)
H	0.0378	0.0372	0.0476	0.1345	0.1307	0.1765*
	(0.0246)	(0.0241)	(0.0327)	(0.0831)	(0.0806)	(0.1012)
$Exp.H$	−0.1506*	−0.1543*	−0.3162**	−0.5455**	−0.5617**	−1.2729**
	(0.0860)	(0.0856)	(0.1216)	(0.2775)	(0.2800)	(0.5647)
$Exp.Ln(rain)H$	0.0833**	0.0902**	0.1584***	0.3090**	0.3371**	0.6624**
	(0.0407)	(0.0417)	(0.0597)	(0.1376)	(0.1440)	(0.3131)
N	8,146	8,146	3,761	7,754	7,754	3,304
R^2	0.1164	0.1173	0.1842			
Pseudo R^2				0.0794	0.0802	0.0969
Alt. Cont. Sites	N	Y	Y	N	Y	Y
Window (days)	280	280	140	280	280	140

Notes: All specifications include monthly, zip code and county fixed effects and standard errors clustered at the county level, which is the administrative unit for county health offices that facilitate private well testing programs in North Carolina. All explanatory and outcome variables are binary indicators except rainfall that is measured as the natural log of the max in, of rainfall experienced within a 24-hour period during the storm event. Coefficients on fixed effects and alternative contamination sites are not presented. Warmer months do show positive seasonality in total coliform presence, which is a well-documented finding (see Hochard et al., 2023) and no relationship is found between alternative contaminated sites and total coliform presence as it relates to Hurricane Matthew's exposures. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

e., rainfall levels from Hurricane Matthew are mapped onto wells that were sampled after the storm event where physical impacts were experienced as well as those that were sampled before the storm and did not experience these physical effects. Effectively, this variable removes spatial trends in total coliform contamination that are correlated with Hurricane Matthew's physical impacts but are otherwise unrelated to the storm itself because they were not unique to the treatment group. The rainfall impacts of Hurricane Matthew on total coliform presence in treated samples is captured by β_3.

The coefficient β_4 measures the difference in total coliform presence rates for wells that share a HUC 12 watershed with swine lagoon relative to those that do not. Like β_1, the adjustment factor, β_5, captures differences in total coliform presence rates between the treatment and control groups that are otherwise unrelated to Hurricane Matthew's rainfall impacts. The impact of rainfall from Hurricane Matthew on total coliform presence in samples taken within the same HUC 12 as swine lagoons is reflected in β_6, which is one of the coefficients of interest because it documents a potential pathway between surface contaminated sites, extreme weather, and groundwater contamination.

A variety of robustness checks are performed to assess the sensitivity of our results to different estimators, alternative local surface sources of contamination, and impacts of the storm event on alternative inorganic analytes that are less associated with animal agriculture – i.e., falsification tests. We also examine how the magnitudes of our estimated effects change when the selection into sample criteria is narrowed from 280 to 140 days when groundwater contamination events that were triggered by Hurricane Matthew are less likely to have been resolved. Finally, we choose our preferred specification to predict the likelihood of total coliform presence in the groundwater at out-of-sample residential addresses known to be occupied by pregnant women. Because few of these households containing pregnant women tested their water following Hurricane Matthew (we find $< 1\%$), we cannot be certain that all these households rely on private wells as their primary source of drinking water. We do adopt the same selection into sample criteria to ensure that the target population (e.g., vulnerable residences occupied by pregnant women) represents closely the population that received post-storm well water testing. We then verify the representativeness of our approach by showing that predicted coliform contamination rates do not vary systematically with known block group rates of private well use (US EPA, 2018), within which the in-sample residents reside.

2. Results

Among HUC 12 basins that do not contain a swine lagoon, total coliform detection rates are generally unrelated to Hurricane Matthew's impact or the intensity of rainfall that the storm brought. The effect of each of the first three covariates (Table 1) is statistically indistinguishable from zero across all three robustness checks (not controlling for alternative contaminated sites, controlling for alternative contaminated sites and changing the selection-into-sample window from 280 days to 140 days), which corresponds to

Table 2
Ordinary least squares (OLS) estimation of Hurricane Matthew on alternative inorganic analytes that were tested in sampled wells. Regressions are presented controlling for alternative contaminated sites (dry cleaning solvent, hazardous waste sites and poultry operations). Each dependence variable is an indicator variable equal to 1 if the level exceeded EPA drinking water maximum contaminant load (MCL) standard and 0 otherwise.

	(1)	(2)	(3)	(5)	(6)
	Arsenic	Chromium	Lead	Manganese	Nitrate
Exp.	0.0056*	0.0007	0.0040	−0.0001	−0.0032
	(0.0032)	(0.0014)	(0.0043)	(0.0160)	(0.0024)
Ln (rain)	0.0259	0.0019**	−0.0092	−0.0980***	0.0064
	(0.0175)	(0.0015)	(0.0138)	(0.0360)	(0.0058)
Exp.Ln(rain)	−0.0002	−0.0014	0.0029	−0.0002	0.0026*
	(0.0018)	(0.0006)	(0.0030)	(0.0115)	(0.0014)
H	0.0070	−0.0003	−0.0001	0.0662	−0.0173**
	(0.0120)	(0.009)	(0.0095)	(0.0439)	(0.0080)
Exp.H	0.0494	0.0002	−0.0182	0.0324	0.0020
	(0.0651)	(0.0012)	(0.0149)	(0.1233)	(0.0147)
Exp.Ln(rain)H	−0.0293	0.0002	−00030	−0.0725	0.0014
	(0.0280)	(0.0005)	(0.0080)	(0.0727)	(0.0095)
N	7,536	7,530	7,957	7,535	6,042
R^2	0.1204	0.0471	0.1219	0.1775	0.1152
Alt. Cont. Sites	N	Y	Y	Y	Y
Window (days)	280	280	280	280	280

Notes: All specifications include monthly, zip code and county fixed effects and standard errors clustered at the county level, which is the administrative unit for county health offices that facilitate private well testing programs in North Carolina. All explanatory and outcome variables are binary indicators except rainfall that is measured as the natural log of the max in, of rainfall experienced within a 24-hour period during the storm event. Coefficients on fixed effects and alternative contamination sites are not presented. The following thresholds are used following the U.S. EPA's maximum contaminant load thresholds and guidance for drinking water: arsenic (0.01 mg/L), chromium (0.1 mg/L), lead (0.015 mg/L), manganese (0.05 mg/L) and nitrate (10 mg/L). *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

columns 1–3. These effects remain null when repeating each estimation using the Probit estimation model (Table 1 columns 4–5). While not statistically significant, it is notable to mention that the β_3 coefficient decreases in absolute magnitude towards "zero" when limiting the estimation window from 280 days to 140 days (comparing columns 1–2 with 3 and columns 4–5 with 6). Here, we would expect the true effect of Hurricane Matthew's rainfall to increase in magnitude as the selection window is shortened and the contamination signal following the storm event is strengthened – i.e., groundwater contamination events following intensive rainfall and flooding are likely to subside over time suggesting samples collected immediately after the storm are more likely to reveal an elevated presence of total coliform.

Hurricane Matthew does alter water quality sampled within the same HUC 12 basins as swine lagoons, which represents 820 observations or approximately 10% of our total sample. Of these sampled wells, we find that 20% of wells are located within 1 km of its nearest swine lagoon and 67% are located within 3 km of its nearest swine lagoon. While there is anecdotal evidence that these areas have generally higher rates of total coliform presence, β_4 ranging from 3.7% to 17.7% depending on the estimation method, these effects are measured with a high degree of imprecision. Evidence connecting Hurricane Matthew's exposure and rainfall with total coliform detection rates is stronger. In the treated sample, we find that low rainfall HUC 12 areas shared with a swine lagoon have lower rates of total coliform detected in well water samples. For the HUC 12 areas that contained a swine lagoon *and* experienced torrential rainfall, we measure a sharp increase in total coliform detection rates, β_5 and β_6. These findings reveal a switching point, identified as around six to eight in. of rainfall within a 24-hour period, where hydrological proximity to a swine lagoon begins to deteriorate drinking water quality for households that were impacted by Hurricane Matthew. This switching point and the overall essence of the findings do not change when controlling explicitly for the distance between a sampled well and its nearest swine lagoon (Table 1a).

We further inspect the relationship between swine lagoons and drinking water quality by measuring the same spatial and temporal impacts of Hurricane Matthew on inorganic analytes that are not particularly associated with swine lagoons.[1] These falsification outcomes help to validate the link between contaminated groundwater and a specific surface-contaminated site, which in this case is nearby swine lagoons. If upstream but unobserved contaminated sites, such as coal ash ponds or other industrial sites, were co-located with swine lagoons, we might expect associated contaminants to also be detected. We find no compelling associations linking Hurricane Matthew's exposure and rainfall intensity with observed rates of arsenic, chromium, lead, manganese, or nitrate contamination (Table 2).

We choose the Probit model with alternative contaminated sites to predict groundwater contamination rates (Table 1 column 5) as our preferred specification because it is the most comprehensive and bounds predictions between 0 and 1, which can be interpreted as

[1] The exception here is groundwater nitrate levels, which would be unsurprising to find elevated nearby swine lagoons and following a heavy rainfall event.

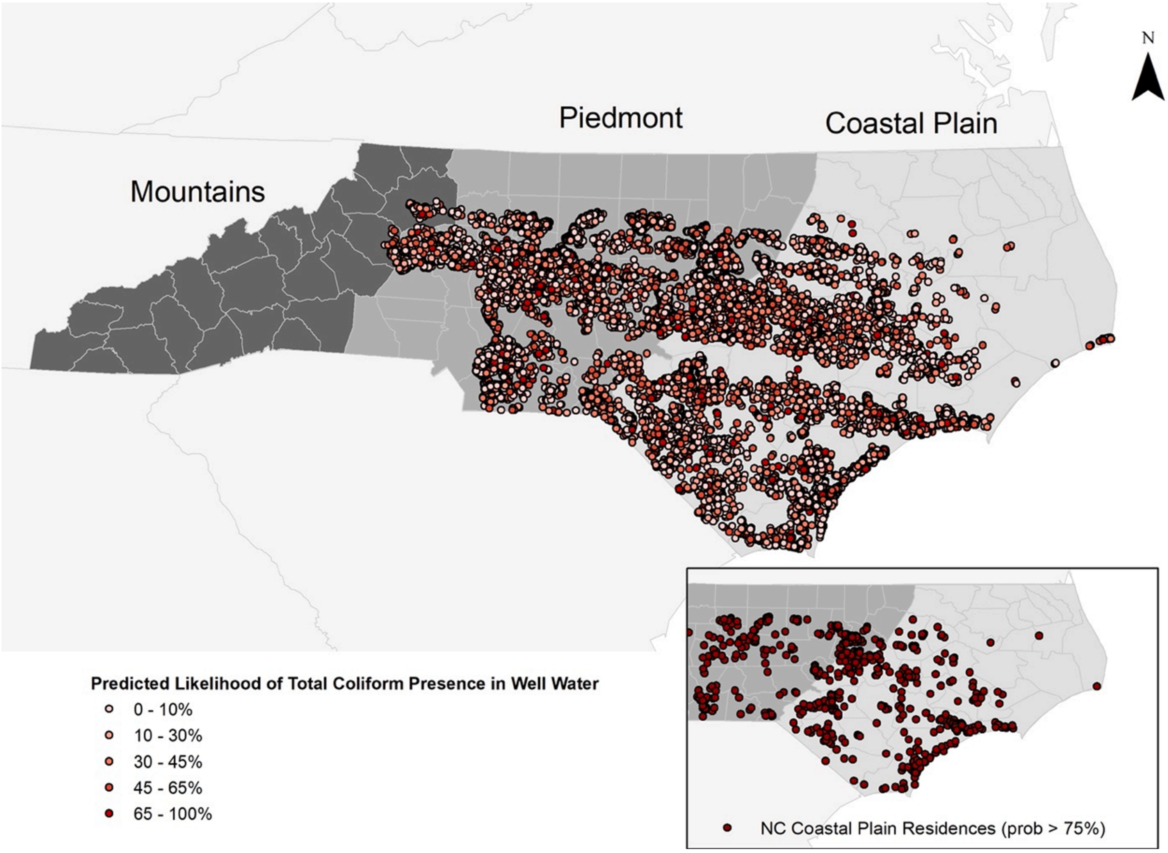


Fig. 3. Predicted likelihood of total coliform contamination in households where pregnant women were known to be living during the 280 days following Hurricane Matthew. Total coliform likelihoods are predicted using the specification shown in Table 1 Column 5, which adopts a Probit estimator and includes control variables for alternative contaminated sites. The predicted likelihood of total coliform presence assumes that the household uses a private well as their primary source of drinking water and that the well was tested in the "highest risk" month throughout the pregnancy but following Hurricane Matthew's North Carolina landfall on October 8th, 2016. Here, the highest risk months are chosen based on the estimated monthly fixed effects that reveal strong seasonality in total coliform presence in private well water during warm months. The sample of viable out-of-sample residential addresses includes 24,493 observations.

"contamination likelihoods". Using this model, we predict the likelihood of detecting the total coliform presence in the drinking water of homes known to have been occupied by pregnant women in the 280 days following Hurricane Matthew's North Carolina landfall. We inspect that > 99% of these addresses were untested during this period. As a hypothetical exercise in prediction, we do not know in which month the sample would be taken but do recognize that there is strong seasonality in total coliform detection. For each woman in our sample, we select the month that overlaps with their pregnancy and maximizes the likelihood of detecting total coliform. This assumption is consistent with how a data driven and targeted water sampling program might be structured – i.e., optimizing sampling efforts in households that have elevated vulnerability and high risk based on spatially and temporally varying factors.

The likelihood of total coliform presence is predicted for a total of 24,493 residences. Of these, the average predicted rate of contamination is 27.6%, which is slightly lower than the measured rate from post-Hurricane Matthew sampling campaigns (29.1%). This finding suggests that average findings from opportunistic sampling do not necessarily underrepresent households occupied by pregnant women. However, among this group of vulnerable households, we find that 17.3% of households (4,228) are predicted to have > 50% likelihood of total coliform presence in their drinking water (Fig. 4). We also find that 2.7% of the sample (670 households) and 0.8% of the sample (190 households) are predicted to have > 75% and > 90% likelihood of total coliform presence in their drinking water, respectively (Fig. 4). To the extent that post-disaster sampling campaigns are constrained by limited resources, leveraging the findings from opportunistic sampling may enable targeted sampling of households that are identified to be vulnerable *and* subject to an elevated risk of groundwater contaminant exposures (see Fig. 5).

Overlaying the block group rates of private well water use (US EPA, 2018 and Fig. 4) reveals that predicted total coliform contamination rates do not vary systematically with deciles of private well dependence (Table 3). This finding is instructive for how a "passive surveillance" approach to private well sampling might be implemented. Because we cannot verify which individual households use private wells as a drinking water source, it is possible that PWS infrastructure has already been expanded into areas with the highest levels of predicted contamination risk. In such a case, our targeting campaign.

would be systematically focused on households with high levels of predicted risk but low levels of exposure because an alternative

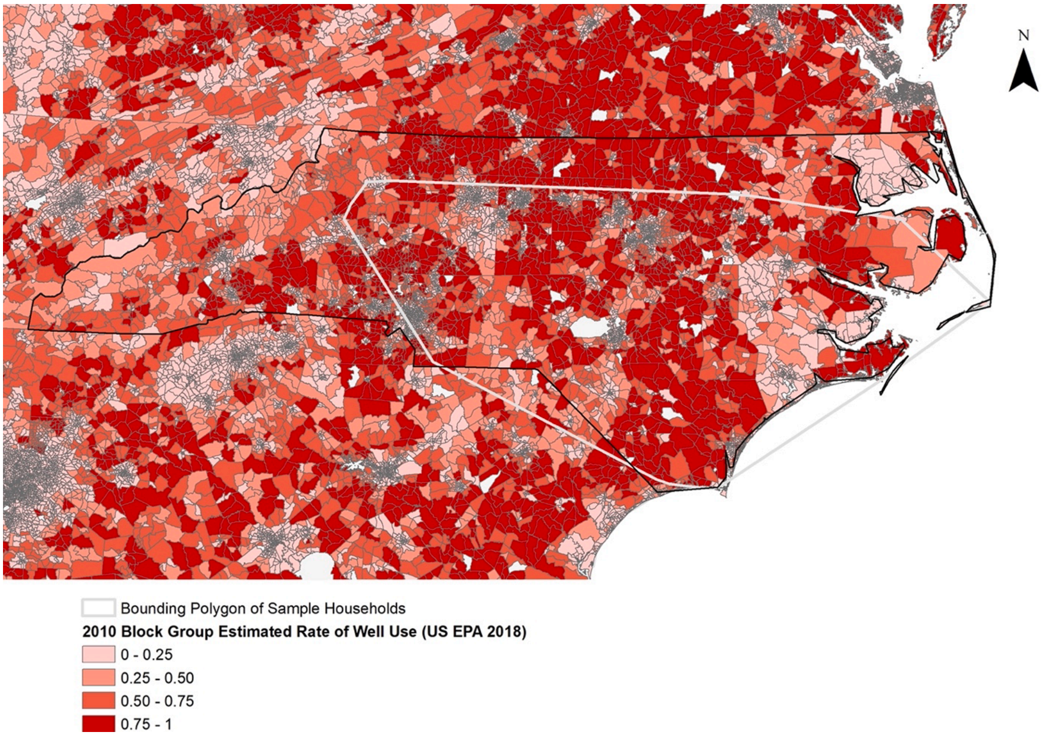


Fig. 4. Minimum bounding geometry of in-sample households and 2010 block group estimates of private well water use rates (US EPA, 2018).

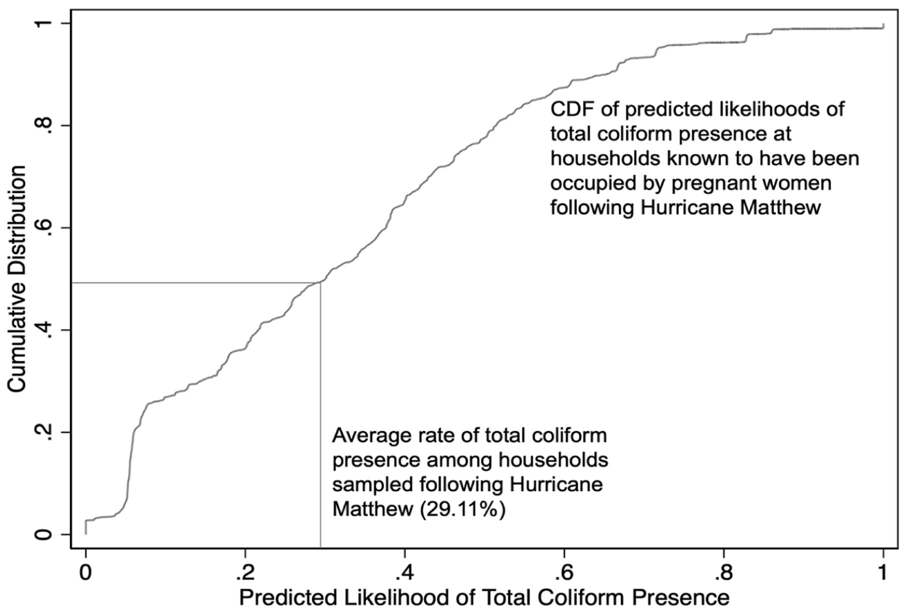


Fig. 5. Cumulative distribution function of the likelihood of total coliform contamination in households where pregnant women were known to be living during the 280 days following Hurricane Matthew. Total coliform likelihoods are predicted using the specification shown in Table 1 Column 5, which adopts a Probit estimator and includes control variables for alternative contaminated sites. The predicted likelihood of total coliform presence assumes that the household uses a private well as their primary source of drinking water and that the well was tested in the "highest risk" month throughout the pregnancy but following Hurricane Matthew's North Carolina landfall on October 8th, 2016. Here, the highest risk months are chosen based on the estimated monthly fixed effects that reveal strong seasonality in total coliform presence in private well water during warm months. The sample of viable out-of-sample residential addresses includes 24,493 observations.

Table 3
Decile relationship between average block group rates of private well use (US EPA, 2018) across study area and average predicted total coliform contamination rates presented in Fig. 3 and Tables 1 column 5.

Block Group Private Well Use (%)	Number of Observations	Average Predicted Total Coliform Contamination Rate
$\leq$ 10%	7,529	0.22
$>$ 10% to $\leq$ 20%	3,024	0.22
$>$ 20% to $\leq$ 30%	2,165	0.23
$>$ 30% to $\leq$ 40%	1,577	0.22
$>$ 40% to $\leq$ 50%	1,790	0.22
$>$ 50% to $\leq$ 60%	1,154	0.22
$>$ 60% to $\leq$ 70%	1,351	0.22
$>$ 70% to $\leq$ 80%	841	0.21
$>$ 80% to $\leq$ 90%	1,459	0.22
$\geq$ 90%	2,457	0.22

water source has been secured – i.e., committing a type I targeting error that would undermine the efficiency of the water sampling program. We do not find such a relationship between predicted risk and private well dependence, which suggests that predicted risk and predicted dependence on private wells could be used together to support a targeted private well sampling program. Designing a well sampling programing in this way recognizes the importance of targeting vulnerable populations (e.g., pregnant women) who are at risk (e.g., high predicted rates of total coliform contamination) and likely exposed (e.g., high local dependence on private wells as a source of drinking water).

3. Conclusions

Rural households that depend on private wells are among a vast population underserved by access to common public services, such as public water systems. These communities also tend to be co-located near surface contaminated sites that are prone to leaking and risk the contamination of nearby groundwater (Paul et al., 2019) and surface water (Sousan et al., 2021) upon which rural livelihoods depend. To date, private wells remain federally unregulated in the United States and the burden of monitoring water quality falls on the homeowner. Yet, widespread evidence suggests that homeowners rarely test their private well water for contaminants in a manner consistent with U.S. Environmental Protection Agency guidelines. The persistence of rural consumption of untested groundwater is of increasing concern with growing evidence that natural disasters trigger surface-contamination events that may jeopardize the potability of water drawn from nearby private wells. North Carolina is particularly concerning where tropical storms lead to frequent and extensive flooding in a coastal plain known for intensive animal agriculture, heavy dependence on private wells and shallow aquifers underneath porous soils subject to infiltration from surface borne contaminants.

The purpose of our study is twofold. First, we assess the extent to which post-disaster but opportunistic well water sampling campaigns represent vulnerable populations in a region. Here, we find that the average risk of well water contamination does not vary systematically between households that received sampling and those households within the same communities that were occupied by pregnant women but did not receive sampling. Second, we examine the extent to which predictive drinking water quality models can be usefully operationalized to support post-disaster sampling campaigns. We discover strong heterogeneity in risks of private well water contamination among our sample of residences home to pregnant women. Specifically, we identify several hundred households that face an extreme likelihood of groundwater contamination (>75%) yet remain uninformed of that risk level because they were not captured by the opportunistic sampling efforts.

The primary contribution of our work is highlighting that water quality prediction models can be bridged with other large, disparate, and residence-level datasets to target post-disaster interventions that may have significant public health savings. Our model highlights a strong relationship between intensive rainfall and the microbiological contamination risk of wells near swine lagoons. However, most of the predictive power in our model comes from the careful control of local and temporal effects (e.g., zip code and county fixed effects, month fixed effects), which are otherwise unobserved but may be useful when predicting spatially and temporally varying risks of groundwater contamination. The approach highlighted in this manuscript should serve as a starting point for more sophisticated models that leverage diverse datasets to improve the predictive accuracy of contamination events. For example, future modeling and targeting could be tailored along socioeconomic and demographic lines to ensure that the most vulnerable communities (e.g., elderly, low-income, etc.) received pointed support. Risk factors themselves, such as the year a structure was built that will correlate with the age of the installed private well system, could also be included to improve the predictive accuracy of such efforts.

A shortcoming of our analysis is that we lack a comprehensive database on households that use private wells as their primary drinking water source. To this end, leveraging such a model in a passive water quality surveillance system might lead to the prediction of risks to vulnerable households that do not use a private well because they have access to a PWS or prefer to purchase bottled water. To the extent that households are aware of groundwater contamination risks (e.g., broad water quality advisories or notification from neighbors who had their water tested, etc.), such averting behavior (e.g., purchasing bottled water, treating water with shock chlorination, or installing a filtration or reverse osmosis system) might also respond to risk conditions. Moving forward, the model presented here serves as one potential basis for more sophisticated prediction and implementation of water quality advisories, which are common practices for PWS users but rare for private well water users. Models used to tailor such advisories for rural homeowners should be validated with on-the-ground testing of predicted outcomes, which we do not complete here, and may also consider

Table A1
Primary Probit specification from Table 1 column 5 of Hurricane Matthew on total coliform presence in sampled wells. Distance to nearest swine lagoon is included as an additional control variable. Dependent Variable: Binary Total Coliform Sample Presence (=1).

	(1)	(2)
	OLS	OLS
Exp.	0.0372	0.0370
	(0.0690)	(0.0688)
Ln (rain)	0.0538	0.0522
	(0.1332)	(0.1332)
Exp.Ln(rain)	−0.0190	−0.0188
	(0.1332)	(0.0484)
H	0.1307	0.1319
	(0.0806)	(0.0852)
Exp.H	−0.5617**	−0.5606*
	(0.2800)	(0.2812)
Exp.Ln(rain)H	0.3371**	0.3367**
	(0.1440)	(0.1444)
Dist. Nearest lagoon (m)		0.0000 (0.0000)
N	7,754	7,754
R^2		
Pseudo R^2	0.0802	0.0802
Alt. Cont. Sites	Y	Y
Window (days)	280	280

leveraging key covariates that are correlated with other known risk factors (e.g., home construction year that correlates with the likely age of a private well installed on the parcel) to enhance their predictive power.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jacob Hochard reports financial support was provided by US Environmental Protection Agency. Jacob Hochard reports financial support was provided by National Science Foundation.

Data availability

The data that has been used is confidential.

Acknowledgments

This work benefited from funding under National Science Foundation award #1902282, U.S. Environmental Protection Agency (EPA) award #RD836942 and #R84018 and support from the U.S. EPA's Office of Resource Conservation and Recovery Act/Oak Ridge Institute for Science and Education faculty fellowship program.

Appendix

See Table A1.

References

Allaire, M., Haowei, W.u., Lall, U., 2018. National trends in drinking water quality violations. Proceedings of the National Academy of Sciences 115 (9), 2078–2083.

Bräunig, J., Baduel, C., Heffernan, A., Rotander, A., Donaldson, E., Mueller, J.F., 2017. Fate and redistribution of perfluoroalkyl acids through AFFF-impacted groundwater. Science of the Total Environment 596, 360–368.

Colley, S.K., Kane, P.KM., Gibson, J.M., 2019. Risk communication and factors influencing private well testing behavior: A systematic scoping review. International Journal of Envrironmental Research and Public Health 16 (22), 4333.

Dai, D., Rhoads, W.J., Katner, A., Strom, L., Edwards, M.A., Pruden, A., Pieper, K.J., 2019. Molecular survey of Legionella and Naegleria fowleri in private well water and premise plumbing following the 2016 Louisiana flood. Environmental Science: Water Research & Technology 5 (8), 1464–1477.

de Knegt, H.J., Eikelboom, J.AJ., Frank van Langevelde, W., Spruyt, F., Prins, H.HT., 2021. Timely poacher detection and localization using sentinel animal movement. Scientific Reports 11 (1), 1–11.

Díaz-Alcaide, S., Martínez-Santos, P., 2019. Mapping fecal pollution in rural groundwater supplies by means of artificial intelligence classifiers. Journal of Hydrology 577, 124006.

Erickson, M.L., Elliott, S.M., Brown, C.J., Stackelberg, P.E., Ransom, K.M., Reddy, J.E., Charles, A., Cravotta III, C.A., 2021. Machine-learning predictions of high arsenic and high manganese at drinking water depths of the glacial aquifer system, northern continental United States. Environmental Science & Technology 55 (9), 5791–5805.

Formuzis, A., 2016. Fields of filth: landmark report maps feces-laden hog and chicken operations in North Carolina. Environmental Working Group https://www.ewg.org/release/fields-filth-landmark-report-maps-feces-laden-hog-and-chicken-operations-north-carolina (22 June 2016).

Gilliland, A.E., Pieper, K.J., Straif-Bourgeois, S., Rhoads, W.J., Dai, D., Edwards, M., Brisolara, K., Olexia, D., Katner, A., 2021. Evaluation of preparedness and recovery needs of private well users after the Great Louisiana Flood of 2016. Journal of Public Health Management and Practice 27 (6), 577–587.

Heaney, C.D., Wing, S., Wilson, S.M., Campbell, R.L., Caldwell, D., Hopkins, B., O'Shea, S., Yeatts, K., 2013. Public infrastructure disparities and the microbiological and chemical safety of drinking and surface water supplies in a community bordering a landfill. Journal of Environmental Health 75 (10), 24.

Hino, M., Benami, E., Brooks, N., 2018. Machine learning for environmental monitoring. Nature Sustainability 1 (10), 583–588.

Hochard, J., Li, Y., Abashidze, N., 2022. Associations of hurricane exposure and forecasting with impaired birth outcomes. Nature Communications 13 (1), 1–10.

Hochard, J., Abashidze, N., Bawa, R., Etheridge, R., Li, Y., Peralta, A., Sims, C., Vogel, T., 2023. Air temperature spikes increase bacteria presence in drinking water wells downstream of hog lagoons. Science of the Total Environment 867, 161426.

Invik, J., Barkema, H.W., Massolo, A., Neumann, N.F., Checkley, S., 2017. Total coliform and Escherichia coli contamination in rural well water: analysis for passive surveillance. Journal of Water and Health 15 (5), 729–740.

Johnson, T.D., Belitz, K., Lombard, M.A., 2019. Estimating domestic well locations and populations served in the contiguous U.S. for years 2000 and 2010. Science of the Total Environment 687, 1261–1273.

Jones, A.Q., Dewey, C.E., Doré, K., Majowicz, S.E., McEwen, S.A., David, W.-T., Eric, M., Carr, D.J., Henson, S.J., 2006. Public perceptions of drinking water: a postal survey of residents with private water supplies. BMC Public Health 6 (1), 1–11.

Lewandowski, A.M., Montgomery, B.R., Rosen, C.J., Moncrief, J.F., 2008. Groundwater nitrate contamination costs: A survey of private well owners. Journal of Soil and Water Conservation 63 (3), 153–161.

Loch-Caruso, R., Rayle, R., Caruso, V.P., Bailey, R.E., Collins, E., Knol, K.P., 2022. Michigan's Gelman Site 1, 4-Dioxane Groundwater Contamination: Still Spreading Decades after Detection. Current Opinion in Environmental Science & Health 100405.

Jones, A.Q., Dewey, C.E., Doré, K., Majowicz, S.E., McEwen, S.A., Waltner-Toews, D., Henson, S.J., Mathews, E., 2005. Public perception of drinking water from private water supplies: focus group analyses. BMC Public Health 5 (1), 1–12.

Knoll, L., Breuer, L., Bach, M., 2019. Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. Science of the Total Environment 668, 1317–1327.

Malecki, K.MC., Schultz, A.A., Severtson, D.J., Anderson, H.A., VanDerslice, J.A., 2017. Private-well stewardship among a general population-based sample of private well-owners. Science of the Total Environment 601, 1533–1543.

Mapili, K., Rhoads, W.J., Coughter, M., Pieper, K.J., Edwards, M.A., Pruden, A., 2022. Occurrence of opportunistic pathogens in private wells after major flooding events: A four state molecular survey. Science of The Total Environment 826, 153901.

Murti, M., Yard, E., Kramer, R., Haselow, D., Mettler, M., McElvany, R., Martin, C., 2016. Impact of the 2012 extreme drought conditions on private well owners in the United States, a qualitative analysis. BMC Public Health 16 (1), 1–10.

NC DEQ, 2016. "Hazardous Waste Sites".

NC DEQ, 2019. "Dry Cleaner Facilities" Avaialble at https://deq.nc.gov/about/divisions/waste-management/science-data-and-reports/dsca-site-listsfacility-inventories.

Newell, C.J., Adamson, D.T., Kulkarni, P.R., Nzeribe, B.N., Stroo, H., 2020. Comparing PFAS to other groundwater contaminants: Implications for remediation. Remediation Journal 30 (3), 7–26.

Nolan, B.T., Hitt, K.J., 2006. Vulnerability of shallow groundwater and drinking-water wells to nitrate in the United States. Environmental Science & Technology 40 (24), 7834–7840.

O'Dwyer, J., Hynds, P.D., Byrne, K.A., Ryan, M.P., Adley, C.C., 2018. Development of a hierarchical model for predicting microbiological contamination of private groundwater supplies in a geologically heterogeneous region. Environmental Pollution 237, 329–338.

Paul, Srikanto, Ghebreyesus, Dawit, Sharif, Hatim O., 2019. Brief communication: Analysis of the fatalities and socio-economic impacts caused by Hurricane Florence. Geosciences 9 (2), 58.

Persson, P., Rossin-Slater, M., 2018. Family ruptures, stress, and the mental health of the next generation. American Economic Review 108 (4–5), 1214–1252.

Pieper, K.J., Nathan Jones, C., Rhoads, W.J., Rome, McNamara, Gholson, D.M., Katner, A., Boellstorff, D.E., Edward Beighley, R., 2021. Microbial Contamination of Drinking Water Supplied by Private Wells after Hurricane Harvey. Environmental Science & Technology 55 (12), 8382–8392.

Podgorski, J., Berg, M., 2020. Global threat of arsenic in groundwater. Science 368 (6493), 845–850.

PRISM Climate Group, 2020. Oregon State University, Gridded Climate Data. http://prism.oregonstate.edu. Accessed 4/25/2020.

Rahman, M.M., Abu Reza Md Towfiqul Islam, Mashura Shammi, 2021. Emerging trends of water quality monitoring and applications of multivariate tools. Water Engineering Modeling and Mathematic Tools 271–283.

Rahmati, O., Choubin, B., Fathabadi, A., Coulon, F., Soltani, E., Shahabi, H., Mollaefar, E., Tiefenbacher, J., Cipullo, S., Bin Ahmad, B., Bui, D., 2019. Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. Science of the Total Environment 688, 855–866.

Schaider, L.A., Ackerman, J.M., Rudel, R.A., 2016. Septic systems as sources of organic wastewater compounds in domestic drinking water wells in a shallow sand and gravel aquifer. Science of the Total Environment 547, 470–481.

Stillo III, F., de Bruin, W.B., Zimmer, C., Gibson, J.M., 2019. Well water testing in African-American communities without municipal infrastructure: Beliefs driving decisions. Science of the Total Environment 686, 1220–1228.

Sousan, S., Iverson, G., Humphrey, C., Lewis, A., Streuber, D., Richardson, L., 2021. High-frequency assessment of air and water quality at a concentration animal feeding operation during wastewater application to spray fields. Environmental Pollution 288, 117801.

Teso, R.R., Poe, M.P., Younglove, T., McCool, P.M., 1996. Use of logistic regression and GIS modeling to predict groundwater vulnerability to pesticides. Journal of Environmental Quality 25 (3), 425–432.

U.S. Environmental Protection Agency. "US Estimated Private Domestic Wells Map Service" (2018) Available at https://edg.epa.gov/metadata/catalog/search/resource/details.page?uuid=%7B508BE1CB-B79B-4BA0-B632-63C95DCD4222%7D.

Vadapalli, V.R., Kumar, E.S., Dube, G., Coetzee, H., 2020. Mine Water Treatment and the Use of Artificial Intelligence in Acid Mine Drainage Prediction. Recovery of Byproducts from Acid Mine Drainage Treatment 23–48.

Vorotyntsev, P., Gordienko, Y., Alienin, O., Rokovyi, O., Stirenko, S., 2021. Satellite image segmentation using deep learning for deforestation detection. In: In *2021 IEEE 3rd Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, pp. 226–231.

Winkel, L., Berg, M., Amini, M., Hug, S.J., Annette Johnson, C., 2008. Predicting groundwater arsenic contamination in Southeast Asia from surface parameters. Nature Geoscience 1 (8), 536–542.

Winner Jr, M.D., Coble, R.W., 1996. Hydrogeologic framework of the North Carolina coastal plain no, 1404-I.