Methods for Estimating Locations of Housing Units Served by Private Domestic Wells in the United States Applied to 2010

Andrew Murray,
Oak Ridge Associated Universities, Cincinnati, Ohio, USA

Alexander Hall,
Office of Research and Development, United States Environmental Protection Agency Cincinnati, Ohio, USA

James Weaver,
Office of Research and Development, United States Environmental Protection Agency Cincinnati, Ohio, USA

Fran Kremer
Office of Research and Development, United States Environmental Protection Agency Cincinnati, Ohio, USA

Abstract

In 1990, the last time the decennial census included a question on domestic drinking water source, it was estimated that private domestic water wells (PDWs) supplied household water to about 15.1 million housing units (15% of the population) in the United States (U.S.). PDWs are not regulated by theSafe Drinking Water Act, and with few exceptions, are not subject to the water quality testing required of public water suppliers. We expanded two methods in estimating housing units reliant on PDWs from an Oklahoma pilot study (Weaver et al. 2017), nationally. Both use 1990 census data on drinking water sources as a baseline. The first method uses housing unit change and private well drilling logs for 20 states. This allows for the rate of well use to change between 1990 and 2010 in these states. The second, based solely on housing unit change, assumes a constant rate of well use. Ordinary least squares regression demonstrated ($R^2 = 0.78$) that the methods yield similar estimates for nationwide well use. Using the housing unit change method, it is estimated that in 2010, 23 million housing units were reliant on PDWs (17% of the population). We provide these estimates at the census block group and census block resolution. This dataset will assist in...
a better understanding of the reliance on PDWs in the U.S., and position local, tribal, state, and national groups to better protect this water resource from contaminant sources.

**Research Impact Statement:** The work provides improved estimates of the spatial distribution of housing units reliant on private domestic wells in the United States and a foundation to protect this water supply at all levels of government.

**Keywords**
- private domestic wells; geospatial analysis; drinking water; dasymetric; GIS

**INTRODUCTION**

**Private Domestic Wells and Regulations**

It has been estimated that roughly 15% of the population in the United States (U.S.) provided their own household water via domestic wells in 1990 (U.S. Census 1990). Estimates for 2010 have included 14% (Maupin et al. 2014) and 12% (Johnson et al. 2019). In 2010, an estimated 3.6 billion gallons of freshwater per day were being pumped by private domestic well (PDW) owners in the U.S. (Maupin et al. 2014). Unlike public water supplies, domestic wells generally draw from shallow sources of groundwater and are not regulated under the Safe Drinking Water Act. The Center for Disease Control has highlighted the public health challenges associated with private drinking water wells (Fox et al. 2016). These challenges included existing and emerging contaminants, advancing the need for private well testing, better data to characterize the contamination in private wells and the populations at risk, and improving stewardship and outreach. The American Academy of Pediatrics has indicated that illness stemming from the consumption of contaminated water from PDWs can be severe, including methemoglobinemia or blue baby syndrome (American Academy of Pediatrics 2009). They have specifically recommended that there should be readily accessible information on groundwater conditions and means to determine where private wells are located.

Under the Public Notification Rule in the Safe Drinking Water Act, public water supply consumers must be notified if there is a problem with their drinking water, where public sources are defined as water systems which serve more than 25 people or have more than 15 connections (regardless of ownership). However, there is no required testing nationwide of private wells and less testing is done for private wells, in general, than for public wells. In addition to the Safe Drinking Water Act not requiring testing of privately owned domestic wells, similarly, state governments generally do not require regular testing of private wells. A few states require testing under certain circumstances, such as property transfers (Oregon Revised Statue 448.271; New Jersey Statutory Authority 58:12A-26 et seq.) although some states have considered and not passed similar requirements (Flanagan and Zheng 2018).

In a recent American Water Works Association survey (American Water Works Association 2018), the two major issues confronting the U.S. water industry are water infrastructure and financing for capital improvements. In particular, areas with disperse housing or small towns present infrastructural and economic challenges in providing a public water supply (Humphreys et al. 2018). A recent study by Kane and Tomer (2019) assessed trends in water
infrastructure investments in the U.S., investments that are principally supported by state
and local governments. State and local governments have worked to stay apace of needed
maintenance in their water infrastructure; they have increased spending just to maintain
current systems. However, there has been a precipitous decline in capital investments for
the expansion of existing water systems or for new systems (Kane and Tomer 2019). Given
these infrastructural and economic challenges, the expansion of public water systems may be
limited, especially in lesser populated areas, resulting in continued reliance on PDWs.

**Potential Vulnerabilities**

Possible sources of contamination to private wells include, but are not limited to:
leaking underground storage tanks, seepage through landfills, failed septic tanks, mining,
agricultural activities (e.g., use of fertilizer, pesticide applications, or runoff from feedlots),
rainfall from urban areas (U.S. Environmental Protection Agency 2016), and geogenic
sources such as naturally occurring arsenic. The obstacles to understanding the vulnerability
of private wells range in scale from the local to the national. For the former, PDW owners
may not understand the importance of water quality, the potential sources contaminating the
wells, how and where to conduct water quality testing, or how to implement treatment. For
public agencies at the local, state, and national level, they may not know where wells are
located and the relative reliance on private wells as a water supply. Additionally, the location
of wells is critical to assist public agencies in ensuring the protection of these water sources
from point and nonpoint source contamination, including industrial and agricultural sources.

Identifying the location of wells and housing units relying on them is vital to developing an
approach to managing the water quality of the wells. Contaminant sources such as leaking
underground storage tanks associated with fueling facilities and abandoned mines are two
illustrative examples. In the U.S., there are nearly 65,000 leaking underground storage tanks
remaining to be cleaned up (U.S. Census Bureau 2019; U.S. Environmental Protection
Agency 2019) with groundwater being the key route of exposure from these releases. In the
U.S. there are over 140,000 abandoned hard rock mines and the prospect of hundreds of
thousands more undocumented ones exist (GAO 2020). These mines are generally sealed
but have unsecured tunnels and toxic waste piles. Rural and tribal areas with little public
water infrastructure are reliant on PDWs. Geospatial data on uranium and arsenic mines
in the Navajo Nation showed arsenic (13%) and uranium (15%) concentrations exceeding
drinking water standards for unregulated water sources (Hoover et al. 2017). Unregulated
water sources within 6 km of the abandoned uranium mines showed significantly higher
concentrations of uranium and arsenic than distant ones. Clearly, a better geospatial
definition of PDWs in relation to these types of contaminant sources will advance public
health.

Related to the importance of understanding the location of wells is testing of the wells
to assess water quality. Knowing the location of wells in context to potential sources of
contamination can provide critical information to the homeowner and public agencies. A
survey of private well owners in Wisconsin, for example, identified a limited understanding
of the need to test private wells and also identified inadequate access to information
on testing and treatment (Malecki et al. 2017). Recent studies in Ontario, Canada have
evidenced the need to have improved spatiotemporal data to better inform health risk assessments, especially for Escherichia coli detection rates (Latchmore et al. 2020). Other microbiological contaminants affecting PDWs include opportunistic pathogens such as Legionella that have been detected in PDWs, especially after flooding. Legionella gene markers were detected in PDWs for over three-fourths of a subset of 40 flooded homes (Dai et al. 2019).

With respect to emerging contaminants, improved understanding of the density and geospatial location of wells, and housing units relying on them, will improve detection and response efforts. As an example, Hu et al. (2016) while evaluating the presence of poly- and perfluoroalkyl substances in public water supplies, realized the shortcomings of available data on these contaminants in PDWs.

Studies focusing on arsenic, in particular, have found that in highly affected areas, more than half of households in central Maine (Flanagan, Marvinney, and Zheng 2015) and northern New Jersey (Flanagan et al. 2016) had never had their well water tested for arsenic. A follow-up study found that 27% of people who received high arsenic test results took no corrective action such as filtration or switching to bottled water (Flanagan, Marvinney, Johnston, et al. 2015). A study in Quebec on the effectiveness of interventions on arsenic testing rates found a preintervention testing rate of 4% and a postintervention rate of 16% (Renaud et al. 2011). At a global scale, Podgorski and Berg (2020) used average arsenic concentrations of nearly 60,000 wells at <100 ft depths in 1 km² grids to develop a machine learning model to estimate exposed populations. This work demonstrates the incomplete data available on PDWs and serves as a call to action to test domestic wells especially in highrisk areas identified by the models. A recent study by Lombard et al. (2021) that identified the potential for increasing concentrations of arsenic in PDWs with increasing drought duration further demonstrates this need. Lastly, Yin et al. (2021) define a geospatially adapted approach to refine sampling of PDWs to more readily detect contamination.

In addition to water quality, understanding the locations of wells is key in assessing water demand, especially under drought conditions, to prevent well failure. Recent work by Pauloo et al. (2020) in California’s Central Valley highlights the importance of determining well locations and well depth as over 2,000 well failures occurred between 2012 and 2016 during the drought. This work underscores the importance of improving groundwater management regimes that is inclusive of public and private water wells — improved geospatial data will significantly advance this.

Lastly, improving environmental health literacy can assist in protecting and maintaining PDWs, especially understanding contaminant sources in relation to wells. Screening tools such as the Water Environmental Literacy Level Scale (Irvin et al. 2019) can help in determining populations that may benefit from focusing resources in needed areas. Providing improved geospatial information on the location of wells and the potential sources and persistence of contamination and providing this information in a medium that is accessible to homeowners and public agencies can provide a better context for making decisions on protecting water quality in PDWs.
## Existing Data on PDWs

Data establishing that 15 million housing units in the U.S. used private wells for domestic water use are derived from the 1990 decennial long-form census. This represents the last time that a robust nationwide survey was conducted to assess the source of household water and provided the most complete assessment to date of domestic well use for the 50 states and the District of Columbia.

Presently, the locations of private wells are inconsistently recorded and maintained across the U.S. Not all states require a well drilling permit and most did not prior to the 1990s (available records are in Supporting Information). Furthermore, required reporting of well logs in states with permitting requirements may not occur due to drillers failing to properly submit them (Oklahoma Water Resources Board 2014). Given the 24% population growth in the U.S. between 1990 and 2010, better methods are needed to augment the 1990 census data for improved accuracy on the reliance of PDWs.

Some current estimates of PDW use rely on county-level estimates, which are made by subtracting the estimated population on public water supply from the total population (Hutson 2007; Maupin et al. 2014). Johnson and Belitz (2015, 2017) have previously proposed two methods to refine the spatial accuracy of the 1990 census data relating to domestic well use by removing areas that are known to have zero population and redefining the areas thought to be using wells based on a 1,000 m distance from named roads. While this is an effective approach for refining and narrowing the geographic areas most likely to be reliant on domestic wells in 1990, it does not fully account for any additional wells constructed since 1990. Additionally, Johnson and Belitz (2015) considered constructed wells in California, which were based on a sample (6%/41,671) of well completion reports provided (N = 741,262) by the California Department of Water Resources, relative to the 464,621 housing units using domestic wells in the 1990 census. In 2019, Johnson et al., advanced that the number of people on domestic water supply by block group could be inferred from a relationship between 1990 well use and population density. They used a domestic ratio, households supplied by PDWs divided by the total households in a block group, derived from the 1990 census data as the starting point and adjusted the ratio for 2000 and 2010 depending on the household density in those years. Additionally, they maintain that as housing density increases, well use decreases; however, this does not consistently occur, especially with significant declines in capital spending for water infrastructure as stated earlier. For example, in the urban fringe, private wells can facilitate scattered development where quality aquifers are accessible. Where development density is low, the cost of private well installation might be much lower than the cost of extending infrastructure and paying municipal connection fees (Burchfield et al. 2006). Thus, additional housing units in a census block group do not always eliminate the possibility of private well use. Additionally, this domestic ratio may not be maintained temporally. For example, this would suggest that within a block group that is 100% reliant on wells and has a housing unit density >14.2/km², if additional houses were built, the number of wells would decrease instead of increase.

In this paper, we apply two methods developed by Weaver et al. (2017): The reported wells (RW) method to 20 states with available well log data and the net housing unit (NHU)
method to all 50 states for estimating domestic well use. Both of these extrapolate from the 1990 census estimates to 2010 using the most recent decennial census data. The RW method relies on state agency records to consider the expansion of domestic well use over time and uses datasets of domestic well drilling logs (N = 1.4 million) obtained from 20 states, which have required well drilling reports since 1990. The second method, NHU, is based on the net change in housing units. In the absence of a nationwide dataset, the RW method provides an empirical benchmark for comparison of conceptually based methods such as NHU. These two approaches allow us to not only consider existing well logs but also changes in housing units temporally at the census block level in order to estimate the total population served by PDWs in the U.S.

METHODS

Data Sources

Well drilling records for domestic wells were sought from each of the 50 states plus Washington, District of Columbia. These records were principally obtained from online databases (see Supporting Information). Specific definitions of what constitutes a PDW can vary among states but, in general, refer to household domestic water uses such as drinking, cleaning, and washing (specific definitions provided in Supporting Information).

In the 1990 census, data from the long-form describe the source of water as either public water supply, drilled well, dug well, or some other source. These data are available at the block group and larger levels (i.e., census tract, county, state). As defined by the U.S. Census Bureau: “Block Groups (BGs) are statistical divisions of census tracts, [and] are generally defined to contain between 600 and 3,000 people.” (U.S. Census Bureau 2010). Alternatively, a census block is “the smallest geographic area for which the Bureau of the Census collects and tabulates decennial census data, are formed by streets, roads, railroads, streams and other bodies of water, other visible physical and cultural features, and the legal boundaries shown on Census Bureau maps” (U.S. Census Bureau 2013). In order to preserve the security of personally identifiable information, relatively few data are available at resolutions finer than the census block group. Tapp (2010) asserts that personal confidentiality takes precedence over geographical accuracy when dealing with large census enumeration districts. Thus, census-based estimates of PDW usage have limited spatial resolution, and individual wells cannot be identified from such results.

Delineations of census areas (“geographies”) change from each decennial census to the next. Therefore, an approach was needed to account for changing areas. Widely considered the ideal method to use for dealing with changing boundaries over time, and also used by Johnson et al. (2019), dasymetric mapping enables the conversion of older census data into a format that is usable with current data. In dasymetric mapping, numerical data within geographic boundaries are converted into density values and then overlaid with new geographic boundaries. Total numerical values can then be recalculated using the new geographic boundaries to multiply area by density (i.e., Holt et al. 2004). This approach allows for the calculation of variables across time and boundaries at the cost of spatial errors dependent on area. A detailed analysis on the statistical accuracies of dasymetric mapping is available in Eicher and Brewer (2001). Because of the emphasis on privacy and the resulting
lack of determination of specific well locations, ancillary data sources must be used to refine geographical accuracy. If it is determined that a significant correlation exists between coarse data and ancillary data which exists at a higher resolution (i.e., the relationship between the source of water and housing units), it is possible to refine the coarse data. Dasymetric mapping of census data requires data that are “spatially extensive” (Goodchild and Lam 1980). As an example, population data and housing unit data exist down to the census block level for the entirety of the U.S.; therefore, if a relationship can be established between domestic well use and population or housing units, spatial resolution can be increased to the census block level.

In 2010, the average area of a census block was 0.826 km$^2$. This size represents a significant increase in spatial resolution by a factor of 50 from the census block group which covers an average of 42.12 km$^2$ per block group. In addition to the relatively large size of block groups, the 1990 census data are now 30 years old. With more than a quarter-century of growth in the U.S., the reliance on private wells has changed and would have changed spatially with population shifts. As spatial scales decrease (county, census tract, census block group, census block), PDW use resolution increases as the amount of undeveloped land included decreases (Weaver et al. 2017). Thus, increasing the resolution provides a more precise estimate of PDW reliance.

Spatial Analysis

Dasymetric mapping, a GIS-based method, was applied to extrapolate 1990 baseline PDW density estimates to 2010. Well counts were extrapolated from prior estimates using the “RW” (Equation 1) and “NHU” (Equation 2) methods developed by Weaver et al. (2017). We used housing units at the census block level as a predictive variable to account for a change in number of housing units reliant on wells over time and to spatially refine our results to the census block level. In both approaches, the 1990 census estimate of household reliance on PDWs was used to generate a baseline of domestic well use for each census block group in the 50 states using the NHU method and for 20 states using the RW method that had well records. The first method is based on the number of RW and housing units lost during a specified time period:

$$\rho_{pdw-est} = \rho_{pdw-init} + \Delta \frac{N_w}{A} - f_{pdw} \frac{N_{HU-lost}}{A},$$

where $\rho_{pdw-est}$ is the PDW density estimate, $\rho_{pdw-init}$ is the initial PDW density, $\Delta \frac{N_w}{A}$ is the change in the number of housing units reliant on wells ($N_w$), and $A$ is the area for analysis (km$^2$). $f_{pdw}$ is the fraction of PDW use to total water supply, and $\frac{N_{HU-lost}}{A}$ is the number of housing units lost per unit area. The initial PDW density and $f_{pdw}$ are inferred from the 1990 census results. $N_w$ is calculated from geolocated well drilling records. The quantity $f_{pdw}$ is updated after each incremental calculation is made, allowing for changing spatial patterns of PDW use. Including the loss of housing units accounts in part for the loss of PDWs. Additionally, well records may only indicate wells added. The RW method is applied to states with RW records that date back to at least 1990.

J Am Water Resour Assoc. Author manuscript; available in PMC 2022 October 01.
The second method, NHU, is based only on the net change in housing units:

\[ \rho_{pdw-est} = \rho_{pdw-init} + f_{pdw} \Delta \frac{N_{HU}}{A}, \]

where \( \Delta \frac{N_{HU}}{A} \) is the net change in housing units per unit area (km\(^2\)). The fraction of private well use \( f_{pdw} \) is determined from the 1990 census results. Any estimates which produced negative PDW density for either method were replaced with a value of zero.

The RW method accounts for shifts in PDW usage based on its reliance on actual well drilling logs containing locations and drill dates and therefore allows for a change in the rate of well use. Some well log data are incomplete, for example, some states do not require reporting, underreporting from drillers, poor locational accuracy, and lack of digitized records. We obtained complete datasets from 20 states which have required reporting since at least 1990. A full assessment of available well log data is available in Supporting Information. We use the NHU method, which assumes that the rate of well use is the same as 1990, for all states for the purposes of supplementing the RW method where well log data are not available and as a means of validation for states where well log data are available.

Dasymetric Method

To apply the dasymetric mapping technique, census data from 1990 and 2000 are converted to density values by dividing data values by the calculated area of each census block group. The area was calculated using the `st_area` function (Pebesma 2018) with block groups in an equal area projection and is presented in km\(^2\). Density values were converted into four separate rasters (1990 well density, 1990 housing unit density, 1990 rate of well use, and 2000 housing unit density) with a cell size of 100 m\(^2\). The 2010 census block group boundaries were then overlaid onto each raster, and the cells which fall within each 2010 census boundary were averaged. Block Groups were used because they are the smallest area at which the 1990 water use census data was available. Once the data were converted into 2010 boundaries, the equations for NHU and RW methods were applied. All census data were obtained through the National Historical Geographic Information System (Manson et al. 2018). Datasets used include 1990 source of water from 1990 Summary Tape File 3 (U. S. Census 1990), 1990 housing unit totals from 1990 Summary File 1 (U. S. Census 1990), 2000 housing unit totals from 2000 Summary File 1b (U. S. Census Bureau 2000), and 2010 housing units from 2010 Summary File 1a (U. S. Census 2010). All data were downloaded at the census block group level.

After the estimates were calculated at the census block group level, the block results were calculated by weighting the percent housing units of the parent block group. Block level domestic supply estimates were calculated for the RW states and nationally using the NHU method.

Accounting for Error

Due to the nature of the dasymetric methods used in this research, there are a fraction of block groups that are susceptible to erroneous estimates. In these cases, errors occur where
there was significant redrawing of census boundaries over time. As the calculations are done in percentages and densities, errors can be compounded in select block groups and yield results that are unrealistically high. An effort was made to refine the final estimates for both the RW and NHU methods to account for these errors and to prevent them from skewing the final estimates. To identify these areas, four tests were applied to the results of each block group in the 2010 census:

**Test 1: Housing change from 1990 to 2000.**—Realistic housing unit density change was determined for block groups between 1990 and 2000 by identifying block groups which maintained the exact same boundaries between census collection years and then finding the range of housing unit density changes for these block groups. Due to the fact that the spatial basis for the census boundary files for 1990 and 2000 are slightly different, block groups were determined to be identical if they had <0.1% change in the area between 1990 and 2000.

**Test 2: Housing change from 2000 to 2010.**—The same method that was applied for 1990–2000 was applied for 2000–2010.

**Test 3: Housing unit density.**—In states where legal setbacks have been adopted, the majority use (at least) a 100-ft (30 m) setback (Supporting Information). This can include setbacks to prevent contamination from potential contaminant sources (such as septic systems) or it can also include setbacks from a property line. This suggests that areas using septic systems would have a maximum housing unit density of 147 housing units/km² without accounting for road networks or public land. Areas without potential septic systems would then have a maximum housing unit density of 278 housing units/km². However, this statistic does not account for developments such as apartment complexes that may skew housing unit density within block groups where larger single-family homes are also present. Therefore, housing unit densities were determined for block groups where well use was reflected in the 1990 census. According to the 1990 census, the median housing unit density of block groups that had at least one respondent declare the use of a drilled or dug well, was 27.05 housing units/km². However, the maximum housing unit density for a block group with a well was 70,000 housing units/km², and the mean value for housing unit density was 188.7 housing units/km². We, therefore, defined realistic housing unit density change by determining the threshold which includes 99.9% of the well use reported in the 1990 census, 1,167 housing units/km².

**Test 4: Predicted well density.**—Analogous to test #3, if the predicted well use density of a block group was >1,167 housing units using wells per km², the results were flagged to indicate that they may be unrealistic.

**Calculating Totals**

After estimates of housing units served by PDWs were calculated for 2010, persons reliant on domestic wells were calculated by dividing 2010 population by 2010 housing units using block group level data from the 2010 decennial census. For each block group, the average population was calculated as total population divided by total housing units, which was then
multiplied by the estimated housing units using wells within that block group to estimate the population served. To determine the relative performance between the RW and NHU methods, an ordinary least squares (OLS) regression was applied to the estimates from the RW and NHU methods on a state-by-state basis (where the RW method was possible). A global $R^2$ was also calculated for the combined data from all 20 states.

**Increasing Resolution to Census Blocks**

Once estimates for housing units served by domestic wells were made at the census block group level, the spatial resolution was further refined to census blocks using housing unit data from the 2010 decennial census. The number of housing units in each block was divided by the number of housing units within the parent block group to determine the percentage of block group housing units within each block. This percentage was then multiplied by the total estimated number of housing units reliant on wells within the parent block group to estimate the total number in each census block.

**RESULTS AND DISCUSSION**

**Tests for Errors**

Estimates were created for each block group in the U.S. using the RW method for those states with well logs and also using the NHU method for all 50 states. The four tests described above were applied to the estimates of each of the 217,221 block groups in the 2010 census. If a block group failed a test, a flag was added to the final estimates.

**Test 1: Housing Change from 1990 to 2000.**—14,586 block groups (7%) were classified as identical based on this method. The range of percent housing unit density change for the block groups that did not change from 1990 and 2000 was: −100% to +1,400%. The distribution for housing unit change was positively skewed (Shapiro–Wilk $p < 0.05$). The mean change was 17.16% (media $N = 10.7\%$) with a standard deviation of 38.5%. While it could be argued that any housing unit change occurring to the maximum observed is reasonable, only two block groups exhibited housing unit change >1,000%, accounting for nine wells from 1990, which is the threshold used. This threshold flagged 1,014 block groups.

**Test 2: Housing Change from 2000 to 2010.**—Using this method, 104,063 block groups were determined to be identical between 2000 and 2010. The range of percent housing unit density change was −100% to +95,100%, that is, the increase would be equivalent to 951 housing units/block group, assuming one existing 1990 well. The mean housing unit density change between 2000 and 2010 for block groups that were not redrawn was 11.7% (media $N = 1.5\%$) with a standard deviation of 4.73. As with housing unit change between 1990 and 2000, outliers were found to positively skew the distribution. Thirty-four block groups had housing unit change >1,000% accounting for 238 housing units using wells from 1990. Again, the threshold was set at 1,000% which flagged 1,390 block groups.

**Test 3: Housing Unit Density.**—Housing unit density >1,167 units per km$^2$ were defined as too dense to be using individual wells. 2,837 block groups (~1.3%) had housing
unit densities >1,167 units per km$^2$ and at least one estimated housing unit using a well. These block groups were flagged as having greater uncertainty.

**Test 4: Predicted Well Density.**—Where the predicted density of housing units using wells within a block group was >1,167 wells per km$^2$ a flag was also added. Only three block groups (0.0001%) had estimated densities of housing units using wells of >1,167 wells per km$^2$. Similarly, these block groups were flagged as having some uncertainty.

In addition, there were 3,313 block groups where calculations failed and returned “not applicable” (NA) values. Each block group that returned NA values for estimated housing units reliant on wells was confirmed to have failed based on one of the following three issues: There were no housing units within the block group in 1990 ($N = 76$), there were no housing units in the block group in 2000 ($N = 3,275$), or there were no housing units in the block group in 2010 ($N = 140$).

**RW and NHU Methods**

Twenty states were determined to have required well drilling reports to be filed for all new domestic water wells since 1990 for the application of the RW method. In total, these 20 states were composed of 63,082 census block groups in the 2010 census. Of these block groups, 37,678 had at least one estimated housing unit using a private well, and 36,070 of those (96%) passed all four tests. For the 20 states with well records, between 1990 and 2010, wells increased from 5,058,405 to 6,268,112: an increase of 1,209,707 wells, or 24%. Figure 1 provides a visual summary incorporating the use of the 1990 census data and the results of the RW method applied to the 20 states. The change in well use from 1990 to 2010 for these 20 states is depicted in Figure 2.

The NHU method was applied to all 50 states and Washington, District of Columbia, a total of 217,221 census block groups. After applying the previously described tests to the estimates from the NHU method, 208,816 block groups passed all four tests with no flags.

The RW method used recorded well driller logs. These empirical data show that the domestic ratio, that is, the density of housing units relying on wells within a block group divided by the total density of housing units can change over time for some block groups. As the RW method uses state well logs it can provide a good indication of the domestic ratio over time (1990–2010). Using the 20 states that have both the NHU and RW models we can determine if the assumption that the NHU model variability in the domestic ratio is relatively similar to the RW method, which measures domestic ratio variability over time.

We test these assumptions by binning our block groups by housing unit density change between 1990 and 2010 (x-axis in Figure 3). Our results show that the strength of the correlation between the NHU and RW methods is dependent on the housing unit density change bins ranging from a correlation of 0.98 to 0.24. The NHU method replicates the RW method more accurately as housing unit density change approaches zero. The three bins with the most block groups, consisting of 68% of the block groups, have correlations ranging from 0.88 to 0.98 between the NHU and RW methods.
The correlations remain relatively high until the threshold of 400 housing unit density change. This 400 housing unit density change suggests a tipping point where incentives may exist for the expansion of public water utilities. After 400 housing unit density change, which is only 7% of the block groups as shown in light gray, the relationship between the NHU and RW becomes less pronounced. Ninety-three percent of block groups exhibit relatively strong relationships between the NHU and RW methods.

Johnson et al. (2019) suggest a variable ratio in which block groups with >14.2 housing units/km² experience a decrease in well use as household density rises over time. By comparing the domestic ratios from the RW method (variable) and the NHU method (static) we find that the domestic ratio remained relatively constant in the vast majority of block groups between 1990 and 2010 (Figure 4). Fitting a linear regression model between the NHU and RW methods demonstrates an $R^2$ value of 0.89. There is a trend in housing unit density change moving from top-left (housing unit loss represented in red) to bottom-right (housing unit increase represented in green), shown in Figure 4a. This trend represents a bias where, for example, the NHU method tends to underestimate housing units using private wells where block group housing unit density decreased over time. Conversely, the NHU method tends to overestimate housing units using wells where block group housing unit density increased. While this demonstrates that there is a density effect within the relationship, there is no linear trend between housing density and well use over time. For example, we found that within RW states, 8,254 of 55,519 (15%) of block groups with housing unit densities >14.2/km² (Johnson et al. threshold) saw an increased domestic ratio between 1990 and 2010. A linear regression between the change in domestic ratio from 1990 and 2010 and 2010 housing unit density for these block groups yields an $r^2 = 0.014$, confirming a lack of a trend. In block groups that skew strongly towards the NHU method (dark green) it is likely that public water infrastructure expansion occurred between 1990 and 2010. These are block groups that experienced substantial housing unit gains with few well drilling records. Excluding the extreme examples of housing unit density changes, there is a close relationship between the NHU and RW method. For context, Figure 4b illustrates the 95% nearest block groups relative to the regression line. By comparing the domestic ratios from the RW method (variable) and the NHU method (static) we find that the domestic ratio remained relatively constant in the vast majority of block groups between 1990 and 2010 (Figure 4), in contrast to Johnson et al. (2019).

In Figure 4, note the block groups that exhibit a domestic ratio >1. While this is empirically unlikely, this phenomenon is likely due to the dasymetric mapping approach. Block groups that have an RW domestic ratio >1 have a mean density of 27 housing units/km² whereas block groups with an RW domestic ratio below one have a mean density of 599 housing units/km². Dasymetric mapping can cause the reallocation of housing units from one census block group to another. These rural areas are likely close to 100% reliant on wells. By mis-aggregating a few housing units between block groups, an RW domestic ratio of >1 may occur. However, this applies to only a small number of block groups (3.4%) and estimated wells.
Overall, in comparing the RW and NHU methods for the 20 states with drill logs, OLS regressions for all 20 states returned p-values of zero and $R^2$ values ranged from 0.15 to 0.95 with a global $R^2$ of 0.78 (Table 1; Figure 5). By demonstrating that the RW and NHU methods yield similar estimates, we established the NHU method as an alternative method for estimating domestic well use nationwide.

Using the NHU method, there is an estimated 22.9 million housing units in the U.S. that are reliant on privately owned wells (Figure 6). The 1990 census data showed 15.1 million housing units reliant on private wells. This constitutes a 50% increase in housing units relying on wells nationally from 1990 to 2010. We look at areas that are near 100% reliant on domestic well supply. These areas are especially susceptible to impacts from groundwater contamination with reduced access to alternative water supplies. In 1990, there were 9.5 million people living in 11,617 block groups that reported >95% reliance on wells. In 2010, using the NHU method, it is estimated that 10.5 million people in 7,470 block groups had >95% reliance on wells. The well use change nationally from 1990 to 2010 is shown in Figure 7. Not surprisingly, declines are seen in the High Plains and increases in western states and coastal areas. To estimate population served, mean population per housing unit was calculated for each block group and multiplied by estimated housing units reliant on wells from the NHU method, leading to our estimate that 52.6 million people, or 17% of the population relied on private wells in 2010. An open-source application (https://gispub.epa.gov/wellmap) has been developed that provides the NHU and RW results at two scales — census block group and census block.

Increasing the spatial resolution from census block groups to census blocks (Figure 8) resulted in a significant increase in spatial certainty of well locations by removing areas known to be uninhabited. The increase was evidenced by 4.8 million (43.7%) census blocks, spanning 4.1 Mkm$^2$ (44.4% of the total area of the U.S.), which had zero population in 2010. Note that census blocks are drawn based on a variety of spatial boundaries such as roads, streams, and transmission lines. Given the density of these features is typically greater in urban and suburban areas, these blocks tend to be much smaller than rural blocks. The reader should note this dynamic in Figure 8—larger, more rural blocks dominate the map. However, many smaller suburban blocks account for a larger number of housing units using wells, which can be more difficult to distinguish at the map’s national scale.

In general, the NHU method estimates more households using wells than the RW method (Figure 9). This result can be observed through the regression plots or the coefficients of the OLS Regression statistics (15 out of 20 coefficients are >1). This may be partly explained by results by Weaver et al., showing undercounts (16%–96%) of RW in selected areas of Oklahoma (Weaver et al. 2017, table 6). This perhaps leads to the RW method undercounting housing units using wells because of incomplete reporting.

Additionally, there are sample-based errors incorporated within the 1990 census which led to an overestimation of wells in some U.S. block groups. For example, the source of water was a sample-based collection, imputed to housing units. Imputation errors led to 3,959 block groups being identified as having more wells than housing units which represents 1.7% of...
all block groups or 4% of all block groups with at least 1 well. Relatedly, it is estimated that roughly 1%–2% of housing units using wells in 1990 were vacant but still counted. As an example, a vacancy includes a temporarily unoccupied rental unit. Given the underlying reliance on the 1990 census data, this phenomenon was embedded through this analysis for both the RW and NHU methods and resulted in some block groups having higher estimates of wells than housing units.

There are some limited areas where the RW method estimated more wells than the NHU method. A definitive reason is not known, but it could be related to anomalies in the methods or specific local factors. For example, Weaver et al. (2017) found drilling of private wells for lawn watering during drought water-use restrictions in Oklahoma. Lawn-watering wells would increase estimates from the RW method, while the number of housing units remains the same. Replacement of older wells could create the same result.

A limitation of both the RW and NHU method stems from the absence of a well data source analogous to the 1990 census. Therefore, projections forward in time are subject to potentially higher error than would be the case for a complete dataset. As such, the results should be viewed as approximations of well counts and density by census block. Additionally, the estimates do not directly take into account the expansion or construction of public water utilities and resulting conversion of water supply from private well to public supply. As noted earlier, this may be especially relevant where housing unit density change exceeds 400. Considerable obstacles exist in converting from PDWs to public water due to the cost of connection and payment of a water bill, so that pockets of PDW use persist within areas served by public water (Weaver et al. 2017). OLS regressions support our hypothesis that, overall, areas reliant on PDWs in 1990 were still reliant on them in 2010. However, there can be localized conditions, for example, water quality impairment, where public water supplies may be considered.

It should also be noted that because the way in which domestic wells are defined varies by state, estimates can be affected. Some states include small farming operations in the domestic water category depending on the pump rate of the water well they are using. Arizona, for example, defines groundwater wells as either exempt or nonexempt. Exempt wells are defined by the Arizona Department of Water Resources as “having a maximum pump capacity of 35 gallons per minute. Typical uses include nonirrigation purposes, noncommercial irrigation of <2 acres of land, and watering stock. Most exempt wells are used for residences and are more than adequate for household use” (Arizona Department of Water Resources, “A Practical Guide to Drilling a Well in Arizona). This definition leaves room for many of the exempt wells in the Arizona database to be used for reasons other than supplying domestic water to a private residence. This would also explain why the RW method estimated fewer domestic wells than existed within the well log database in Arizona in 2010. In fact, Arizona had 79,513 exempt wells in its database from earlier than 1990 which is more than the 77,229 recorded in the 1990 census. Arizona began requiring the reporting of wells in 1984 so most wells constructed prior to 1984 would not appear in the database. It is, therefore, reasonable to suggest that if Arizona recorded more specific use types in their well log records, we may see an increased $R^2$ value for the relationship between the hybrid method and the NHU method.

J Am Water Resour Assoc. Author manuscript; available in PMC 2022 October 01.
DATA AVAILABILITY STATEMENT

All code necessary for the reproduction of the analyses presented in this paper is available online at: https://github.com/USEPA/PDW_Paper_2020. A web mapping application providing the well data, by census block group and census block, is available at: https://gispub.epa.gov/wellmap.

CONCLUSIONS

The two geospatial datasets created contain 2010 estimates of the number of housing units using domestic wells within census block groups, and census blocks. One set is for 20 states using the RW logs method and the second is for the entire U.S. using the NHU method. Census block groups are a significant improvement in spatial resolution relative to previously available county level estimates; however, they are not granular enough for localized studies such as most site-specific analysis or local planning. By projecting the well use information from census block group to the census block it further increases the resolution by a factor of 50. The estimates presented in this paper represent the most comprehensive picture of PDWs through 2010 with improved spatial resolution incorporating both state well logs and housing units. While these methods were used to estimate well use for 2010, we will build upon these approaches with the anticipated release of the 2020 decennial census and the continued expansion of the availability of well drilling logs across the U.S. It must be noted, however, that these are estimates and should not be mistaken for exact numbers of domestic wells within the given geographic areas. That being said, there is a significant potential for applying these estimates to address drinking water vulnerability from contaminants, water resource planning, land use planning, agricultural needs, and emergency planning. In emergency response operations, having private well geospatial data can assist emergency response efforts in identifying and restoring drinking water supplies for impacted communities — this will be especially important under flooding conditions. Knowing the geospatial density by census block will enable emergency responders to better plan for replacement water needs and the States and communities can work with well owners to communicate needed public health information and assist in the recovery of contaminated wells. These estimates will be especially useful for the 60% of the states which do not have detailed well logs dating to 1990. Additionally, these improved estimates will allow states as well as tribes and local communities to better protect wells from contamination, assist in the recovery of wells and use this information for future resource planning. For states with well logs, the NHU method can potentially assist these states in refining their methods, for example, where well logs have not been reported.

Lastly, given the reliance in the U.S. on PDWs and their potential vulnerability, the U.S. American Community Survey/Housing Survey could have a key role in collecting data on the sources of drinking water in order to protect this resource. It is incumbent upon national, state, and local entities to work toward protecting this vital drinking water resource. It is clear that our understanding of the locations of privately owned domestic water wells is lacking and outdated. It is equally clear that because drinking water can act as a delivery system for a range of contaminants and that private wells lack robust testing, the more we can do to understand where they are in use, the more we can do to prevent contaminants
from entering the wells to protect public health moving forward. Contaminant transport occurs irrespective of geographical boundaries. An integrated approach at the local, tribal, state, and national level is required to determine PDW locations in context of contaminant sources. This approach to decision making will enable us, collectively, to protect existing and future water supplies.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**ACKNOWLEDGMENTS**

The authors of this paper have no conflicts of interest to declare. The views expressed in this paper are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. All regression statistics and spatial analyses were completed using R version 3.6.0. All census boundary and table data were obtained through the National Historical Geographic Information System: Minnesota Population Center. *National Historical Geographic Information System: Version 2.0. Minneapolis, MN: University of Minnesota* (http://www.nhgis.org).

**LITERATURE CITED**

American Academy of Pediatrics. 2009. Drinking Water From Private Wells and Risks to Children. *PEDIATRICS* 123 (6): 1599–1605. 10.1542/peds.2009-0751. [PubMed: 19482772]

American Water Works Association. 2018. State of the Water Industry Report 2018. Denver, CO: American Water Works Association. [https://www.awwa.org/Portals/0/AWWA/Development/Managers/2018_SOTWI_Report_Final_v3.pdf](https://www.awwa.org/Portals/0/AWWA/Development/Managers/2018_SOTWI_Report_Final_v3.pdf).

Arizona Department of Water Resources. “A Practical Guide to Drilling a Well in Arizona.” [http://infoshare.azwater.gov/docushare/dsweb/Get/WellRegDoc-371981/A%20Practical%20Guide%20to%20Well%20Drilling%20in%20Arizona.pdf](http://infoshare.azwater.gov/docushare/dsweb/Get/WellRegDoc-371981/A%20Practical%20Guide%20to%20Well%20Drilling%20in%20Arizona.pdf).

Burchfield M, Overman HG, Puga D, Turner MA 2006. Causes of Sprawl: A Portrait from Space. *The Quarterly Journal of Economics* 121 (2): 587–633. 10.1162/qjec.2006.121.2.587.

Dai D, Rhoads WJ, Katner A, Strom L, Edwards MA, Pruden A, and Pieper KJ 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: 1464–77. 10.1039/C9EW00109C.

Eicher CL, and Brewer CA 2001. “Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation.” *Cartography and Geographic Information Science* 28 (2): 125–38. 10.1559/152304001782173727.

Flanagan SV, Marvinney RG, Johnston RA, Yang Q, and Zheng Y. 2015. “Dissemination of Well Water Arsenic Results to Homeowners in Central Maine: Influences on Mitigation Behavior and Continued Risks for Exposure.” *Science of the Total Environment* 505: 1282–90. 10.1016/j.scitotenv.2014.03.079.

Flanagan SV, Marvinney RG, and Zheng Y. 2015. “Influences on Domestic Well Water Testing Behavior in a Central Maine Area with Frequent Groundwater Arsenic Occurrence.” *Science of the Total Environment* 505: 1274–81. 10.1016/j.scitotenv.2014.05.017.

Flanagan SV, Spayd SE, Procopio NA, Chilrud SN, Braman S, and Zheng Y. 2016. “Arsenic in Private Well Water Part 1 of 3: Impact of the New Jersey Private Well Testing Act on Household Testing and Mitigation Behavior.” *Science of the Total Environment* 562: 999–1009. 10.1016/j.scitotenv.2016.03.196.

Flanagan SV, and Zheng Y. 2018. “Comparative Case Study of Legislative Attempts to Require Private Well Testing in New Jersey and Maine.” *Environmental Science and Policy* 85: 40–46. 10.1016/j.envsci.2018.03.022.
Fox MA, Nachman KE, Anderson B, Lam J, and Resnick B. 2016. “Meeting the Public Health Challenge of Protecting Private Wells: Proceedings and Recommendations from an Expert Panel Workshop.” Science of the Total Environment 554–555: 113–18. 10.1016/j.scitotenv.2016.02.128.

GAO (Government Accountability Office). 2020. “Abandoned Hardrock Mines: Information on Number of Mines, Expenditures, and Factors That Limit Efforts to Address Hazards.” GAO-20238, March 2020. https://www.gao.gov/assets/710/705146.pdf.

Goodchild MF, and Lam NSN 1980. “Areal Interpolation: A Variant of the Traditional Spatial Problem.” Geo-Processing 1: 297–312. https://asu.pure.elsevier.com/en/publications/areal-interpolation-a-variant-of-the-traditional-spatial-problem.

Holt JB, Lo CP, and Hodler TW 2004. “Dasymetric Estimation of Population Density and Areal Interpolation of Census Data.” Cartography and Geographic Information Science 31 (2): 103–21. 10.1559/1523040041649407.

Hohover J, Gonzales M, Shuey C, Barney Y, Lewis J. 2017. Elevated Arsenic and Uranium Concentrations in Unregulated Water Sources on the Navajo Nation, USA. Exposure and Health 9 (2): 113–124. 10.1007/s12403-016-0226-6. [PubMed: 28553666]

Hu XC, Lindstrom AB, Bruton TA, Schaider LA, Grandjean P, Lohmann R. et al. 2016. “Detection of Poly- and Perfluoroalkyl Substances (PFASs) in U.S. Drinking Water Linked to Industrial Sites, Military Fire Training Areas, and Wastewater Treatment Plants.” Environmental Science & Technology Letters 3 (10): 344–50. 10.1021/acs.estlett.6b00260. [PubMed: 27752509]

Humphreys E, van der Kerk A, and Fonseca C. 2018. “Public Finance for Water Infrastructure Development and Its Practical Challenges for Small Towns.” Water Policy 20 (Suppl 1): 100–11. 10.2166/wp.2018.007.

Hutson SS 2007. “Guidelines for Preparation of State Water-Use Estimates for 2005.” U.S. Geological Survey Techniques and Methods Book 4, Chap. E1, 36 pp. https://pubs.er.usgs.gov/publication/tm4E1.

Irvin V, Rohlman D, Vaughan A, Amantia R, Berlin C, and Kile M. 2019. “Development and Validation of an Environmental Health Literacy Assessment Screening Tool for Domestic Well Owners: The Water Environmental Literacy Level Scale (WELLS).” Int. J. Environ. Res. Public Health, 16 (5), 881; 10.3390/ijerph16050881.

Johnson TD, and Belitz K. 2015. “Identifying the Location and Population Served by Domestic Wells in California.” Journal of Hydrology: Regional Studies 3: 31–86. 10.1016/j.ejrh.2014.09.002.

Johnson TD, and Belitz K. 2017. “Domestic Well Locations and Populations Served in the Contiguous US: 1990.” Science of the Total Environment 607: 658–68. 10.1016/j.scitotenv.2017.07.018.

Johnson TD, Belitz K, and Lombard M. 2019. “Domestic Well Locations and Populations Served in the Contiguous U.S. for 2000 and 2010.” Science of the Total Environment 687 (2019): 1261–73. 10.1016/j.scitotenv.2019.06.036.

Kane J, and Tomer A. 2019. “Shifting Into an Era of Repair: US infrastructure Spending Trends.” Brookings Institution. https://www.brookings.edu/research/shifting-into-an-era-of-repair-us-infrastructure-spending-trends/.

Latchmore T, Hynds P, Stephen Brown R, Schuster-Wallace C, Dickson-Anderson S, McDermott K, and Majury A. 2020. “Analysis of a Large Spatiotemporal Groundwater Quality Dataset, Ontario 2010–2017: Informing Human Health Risk Assessment and Testing Guidance for Private Drinking Water Wells.” Science of The Total Environment 738: 2010–17. 10.1016/j.scitotenv.2020.140382.

Lombard MA, Daniel J, Jedyd Z, Hay LE, and Ayotte JD 2021. “Assessing the Impact of Drought on Arsenic Exposure from Private Domestic Wells in the Conterminous United States.” Environmental Science & Technology 55 (3): 1822–31. 10.1021/acs.est.9b05835. [PubMed: 33439623]

Malecki KMC, Schultz AA, Severtson DJ, Anderson HA, and VanDerslice JA 2017. “Private-Well Stewardship among a General Population Based Sample of Private Well-Owners.” Science of the Total Environment 601–602: 1533–43. 10.1016/j.scitotenv.2017.05.284.

Manson S, Schroeder J, Van Riper D, and Ruggles S. 2018. IPUMS National Historical Geographic Information System: Version 13.0 Database. Minneapolis: University of Minnesota. 10.18128/D050.V13.0.
FIGURE 1.
Map showing 2010 estimated housing units using wells by the reported wells (RW) method by block group.
FIGURE 2.
Map showing 2010 estimated housing units using wells by the RW methods minus the RW in the 1990 census by block group.
FIGURE 3.
Pearson $r$ correlations between net housing unit (NHU) and RW methods, binned by housing unit density change between 1990 and 2010.
FIGURE 4.
Linear regression between the NHU and domestic ratio (rate of housing units using wells), colored by housing unit density change. 95% of the data fall within the light gray area.
FIGURE 5.
Linear regressions between the NHU and RW methods by state.
FIGURE 6.  
Map showing 2010 estimated housing units using wells by the NHU method by block group.
FIGURE 7.
Map showing 2010 estimated housing units using wells by the NHU method minus the RW in the 1990 census by block group.
FIGURE 8.
Map showing 2010 estimated housing units using wells by the NHU methods, after being spatially refined from census block groups to census blocks.
FIGURE 9.
Bar graph showing estimates of housing units using wells in 2010 broken down by state and method.
TABLE 1.

$R^2$ and Y-int for regressions between the RW and NHU methods by state.

| State     | $R^2$ | Y-int |
|-----------|-------|-------|
| Arizona   | 0.37  | 5.3   |
| Arkansas  | 0.86  | 4.7   |
| Colorado  | 0.66  | 3.7   |
| Idaho     | 0.53  | 63.3  |
| Kansas    | 0.8   | 5.6   |
| Kentucky  | 0.95  | −1.4  |
| Louisiana | 0.88  | 4.9   |
| Maine     | 0.95  | 12.4  |
| Maryland  | 0.82  | 3.2   |
| Michigan  | 0.9   | 2.8   |
| Minnesota | 0.8   | 9.8   |
| Missouri  | 0.88  | 6.3   |
| Montana   | 0.78  | 26.6  |
| Nevada    | 0.15  | 11.5  |
| New Jersey| 0.83  | 2.2   |
| New Mexico| 0.69  | 12.3  |
| Ohio      | 0.82  | 8.2   |
| Oklahoma  | 0.91  | 4     |
| Vermont   | 0.95  | 17.2  |
| Wyoming   | 0.89  | 2.1   |