Water Availability and Use Science Program

Estimates of Water Use Associated with Continuous Oil and Gas Development in the Williston Basin, North Dakota and Montana, 2007–17

Scientific Investigations Report 2020–5012

U.S. Department of the Interior
U.S. Geological Survey

Cover. Aerial photograph of oil and gas wells under development in North Dakota. Photograph by Vern Whitten Photography, used with permission.

Estimates of Water Use Associated with Continuous Oil and Gas Development in the Williston Basin, North Dakota and Montana, 2007–17

By Ryan R. McShane, Theodore B. Barnhart, Joshua F. Valder, Seth S. Haines, Kathleen M. Macek-Rowland, Janet M. Carter, Greg C. Delzer, and Joanna N. Thamke

Water Availability and Use Science Program

Scientific Investigations Report 2020–5012

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DAVID BERNHARDT, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2020

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit https://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit https://store.usgs.gov/.

Suggested citation:
McShane, R.R., Barnhart, T.B., Valder, J.F., Haines, S.S., Macek-Rowland, K.M., Carter, J.M., Delzer, G.C., and Thamke, J.N., 2020, Estimates of water use associated with continuous oil and gas development in the Williston Basin, North Dakota and Montana, 2007–17: U.S. Geological Survey Scientific Investigations Report 2020–5012, 26 p., https://doi.org/10.3133/sir20205012.

Associated data for this publication:
Dutton, D.M., Varela, B., Haines, S.S., Barnhart, T.B., McShane, R.R., and Wheeling, S.L., 2019, Water-use data to estimate water use associated with continuous oil and gas development, Williston Basin, United States, 1980–2017: U.S. Geological Survey data release, accessed October 2019 at https://doi.org/10.5066/P9CPKRLW.

ISSN 2328-0328 (online)

Acknowledgments

The authors thank the North Dakota State Water Commission, the North Dakota Industrial Commission, the Montana Department of Natural Resources and Conservation, the Montana Bureau of Mines and Geology, and other officials from North Dakota and Montana for their support of past and ongoing projects that provided valuable information for this study. Hess Corporation and Whiting Petroleum Corporation offered beneficial information on water use by oil and gas operators in North Dakota, and the Southwest Water Authority offered useful information on water use by water permit holders in North Dakota. The authors also thank IHS Markit for approving access to aggregated data on hydraulic fracturing water use in Montana.

This study was supported by the U.S. Geological Survey Water Availability and Use Science Program. The authors thank the following colleagues with the U.S. Geological Survey. DeAnn Dutton, Robert Lundgren, Roy Sando, and Spencer Wheeling assisted with gathering and processing the water-use data. Brian Clark and William Eldridge provided thoughtful edits and comments in reviewing this report, and Molly Maupin and Mindi Dalton provided helpful guidance as members of the project management team.

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Length	
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
acre	4,047	square meter (m^2)
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
barrel (bbl; petroleum, 1 barrel=42 gal)	0.1590	cubic meter (m^3)
gallon (gal)	3.785	liter (L)
gallon (gal)	0.003785	cubic meter (m^3)
million gallons (Mgal)	3,785	cubic meter (m^3)
cubic foot (ft^3)	0.02832	cubic meter (m^3)
	Mass	
pound, avoirdupois (lb)	0.4536	kilogram (kg)

Supplemental Information

Water use is given in million gallons per year (Mgal/yr) or million gallons (Mgal) per well.

Abbreviations

COG	continuous oil and gas
NDIC	North Dakota Industrial Commission
NDSWC	North Dakota State Water Commission
R^2	coefficient of determination
RMSE	root mean square error
RSR	ratio of root mean square error to standard deviation of observations
USGS	U.S. Geological Survey
WAUSP	Water Availability and Use Science Program

Estimates of Water Use Associated with Continuous Oil and Gas Development in the Williston Basin, North Dakota and Montana, 2007–17

By Ryan R. McShane, Theodore B. Barnhart, Joshua F. Valder, Seth S. Haines, Kathleen M. Macek-Rowland, Janet M. Carter, Greg C. Delzer, and Joanna N. Thamke

Abstract

This study of water use associated with development of continuous oil and gas resources in the Williston Basin is intended to provide a preliminary model-based analysis of water use in major regions of production of continuous oil and gas resources in the United States. Direct, indirect, and ancillary water use associated with development of continuous oil and gas resources in the Williston Basin was estimated in North Dakota and Montana from 2007 to 2017. Water-use data were aggregated by county and year, which were the sampling units used in this analysis. Linear and quantile regression models of water use in relation to the number of oil and gas wells developed were fit for the direct, indirect, and ancillary water-use categories for each State. A 95-percent confidence interval for each parameter estimate from the linear regression models was computed as a measure of uncertainty. Additional information on uncertainty can be gained from modeling other distribution parameters, so quantile regression models of the 5th, 50th, and 95th percentiles also were fit. To assess uncertainty in the estimates from the regression models of direct, indirect, and ancillary water use, leave-one-out cross-validation was used. Model performance was evaluated with three goodness-of-fit metrics used to compare the estimates and observations of water use.

Mean annual direct and indirect water use for development of continuous oil and gas resources in North Dakota was estimated at 4,512 million gallons (Mgal) per year (Mgal/yr), with a 95-percent confidence interval of 4,021–5,152 Mgal/yr, and in Montana was estimated at 196 Mgal/yr, with a 95-percent confidence interval of 189–203 Mgal/yr. Ancillary water use (for domestic and public supply) had an estimated annual mean of 2,753 Mgal/yr in North Dakota and 396 Mgal/yr in Montana. The coefficient from the linear regression model of direct water use was 3.86 Mgal per well and hydraulic fracturing water use was 3.70 Mgal per well for North Dakota. The mean estimate of direct water use had a 95-percent confidence interval of 3.48–4.23 Mgal per well. For North Dakota, the coefficient from the linear regression model of indirect water use was 0.453 Mgal per well, with a 95-percent confidence interval of 0.415–0.492 Mgal per well. Direct and indirect water use had a mean estimate of about 4.31 Mgal per well in North Dakota. The mean estimate of ancillary water use (for domestic and public supply) in North Dakota was 2.03 Mgal per well, with a 95-percent confidence interval of 1.76–2.31 Mgal per well. For Montana, the linear regression model of hydraulic fracturing water use had a mean estimate of 2.04 Mgal per well. The 95-percent confidence interval for the mean estimate was 1.80–2.28 Mgal per well. Direct and indirect water use in Montana had a mean estimate of 2.49 Mgal per well. The mean estimate of ancillary water use (for domestic and public supply) in Montana was 2.43 Mgal per well, with a 95-percent confidence interval of 1.76–3.11 Mgal per well.

Introduction

Water is a necessary component for many processes required for developing continuous oil and gas (COG) resources. Improved COG extraction techniques have greatly increased oil and gas production in the United States since the mid-2000s (U.S. Energy Information Administration, 2016). However, the accompanying rapid increase in demand for large volumes of water, often in remote regions, can challenge existing infrastructure and require additional resources to meet water needs. Addressing this water need requires accurate estimates of the volumes of water used to support the various processes common to COG development in the United States in the 21st century.

In 2015, the U.S. Geological Survey (USGS) started a topical study focused on quantifying water use in areas of COG development. The topical study was supported through the USGS Water Availability and Use Science Program (WAUSP), which was authorized by the Science and Engineering to Comprehensively Understand and Responsibly Enhance Water Act (SECURE Water Act) in the Omnibus Public Land Management Act of 2009 (16 U.S.C. 1 note). In the SECURE Water Act, the USGS was tasked with conducting a National Water Census to better quantify water use in the United States, including water supporting COG development. One of the main goals of the WAUSP is to provide accurate estimates of water resources in the United States and to offer methods for determining the quantity and quality of water available for beneficial uses. This topical study to quantify water use related to COG development will help achieve that WAUSP goal.

Previous work supporting this study includes two USGS reports: (1) Valder and others (2018) developed a conceptual model for assessing water use associated with the life cycle of COG development based on the quantity and quality of water-use data available from various sources; and (2) Valder and others (2019) constructed an analytical framework for estimating water use related to COG development, including an uncertainty analysis of the water-use estimates. The analytical framework also included R scripts with the procedures used to estimate water use, which can be adapted for application to areas of COG production throughout the United States.

This study of water use associated with COG development in the Williston Basin is intended to provide a preliminary model-based analysis of water use in major regions of COG production in the United States. The Williston Basin was selected as the pilot region for the following reasons: (1) the estimated large volume of undiscovered oil and natural gas (Gaswirth and others, 2013); (2) the quantity and quality of available water-use data; and (3) a population—with its attendant water use—assumed to change largely in response to oil and gas development.

Description of Study Area

The Williston Basin is a large intracratonic, roughly circular, sedimentary and structural basin that extends from South Dakota, North Dakota, and Montana (Gaswirth and others, 2013; fig. 1) in the United States to Saskatchewan and Manitoba in Canada (not shown), covering more than 100,000 square miles. The geologic basin is deepest at its center, which is located beneath Williston, North Dakota, with the strata becoming shallower and thinner towards its margins (fig. 2; Long and others, 2014). The stratigraphic sequence of the Williston Basin (fig. 2) is described thoroughly in Long and others (2014).

The focal area of this study was the Williston Basin Province in the United States (not in Canada; fig. 1), a region with a boundary that follows the County and State borders most closely matching the boundary of the geologically-defined Williston Basin (Gaswirth and others, 2013). The province contains the Bakken and Three Forks Assessment Units of conventional and COG accumulations (fig. 1; Gaswirth and others, 2013). The primary sources of fresh groundwater in the Williston Basin, which are used in part to support COG development, are the Fort Union, Fox Hills, and Hell Creek Formations (Long and others, 2018) contained in the Lower Tertiary-aged and Upper Cretaceous-aged layers (figs. 2 and 3). Most water used for supporting COG development is sourced from fresh surface waters of lakes or ponds and tributaries of the Missouri River (figs. 1 and 2). The principal COG deposits in the Williston Basin are the Bakken and Three Forks Formations in the Mississippian-aged and Upper Devonian-aged layers (figs. 2 and 3).

History of Oil and Gas Development in the Williston Basin

Since the late 1800s, the type and amount of energy development in the Williston Basin has changed based on many factors including technological developments, production costs, energy prices, and political decisions. Energy resources developed in the Williston Basin have included nonrenewable resources such as coal, oil, and gas, and renewable energies such as biofuels and wind. The Williston Basin has been an important domestic oil- and gas-producing region since the 1950s (Anna and others, 2011; Gleason and Tangen, 2014; Thamke and others, 2014).

Crude oil and natural gas deposits are categorized as conventional or continuous (also referred to as unconventional) based on characteristics of the reservoir (Schmoker, 2005; Schmoker and Klett, 2005; fig. 3). Conventional oil and gas accumulations have discrete deposits with well-defined hydrocarbon-water contacts (where the hydrocarbons are buoyant on a column of water), generally high matrix permeabilities, apparent seals and traps, and relatively high recovery factors (Schmoker and Klett, 2005). COG accumulations are an oil or gas resource, or both, dispersed evenly throughout a geologic formation rather than existing as discrete and localized deposits, such as those in conventional accumulations (Schmoker, 2005). These continuous resources typically require specialized extraction techniques involving directional, horizontal wells, which have been more costly than extraction with vertical wells.

The first recorded energy exploration in the Williston Basin was for natural gas, and the first energy wells completed were exploratory natural gas wells (Bluemle, 2001). Natural gas was first reported in southeastern North Dakota in 1892 in an artesian water well producing from the Dakota Sandstone (Anderson and Eastwood, 1968; Bluemle, 2001). Several years later, shallow deposits of natural gas were discovered in north-central North Dakota and were used to supply several small towns and farms (Anna and others, 2011). The oldest commercial gas production was established in 1913 from shallow Upper Cretaceous units in the northwest part of the Cedar Creek anticline in Montana (Anna and others, 2011). The new discovery was expanded to 11 fields, and in 1932, more than 12 billion cubic feet of gas were sold (Bartram and Erdmann, 1935; Anna and others, 2011).

Before 1910, early oil and gas exploration in the Williston Basin was hampered by primitive technology, and the existing cable-tool rigs, which used a heavy bit on the end of a steel cable to crush the rock by repeatedly dropping the bit into the well borehole, could not drill deeply enough (Bluemle, 2001). Oil was first discovered in the Williston Basin along the

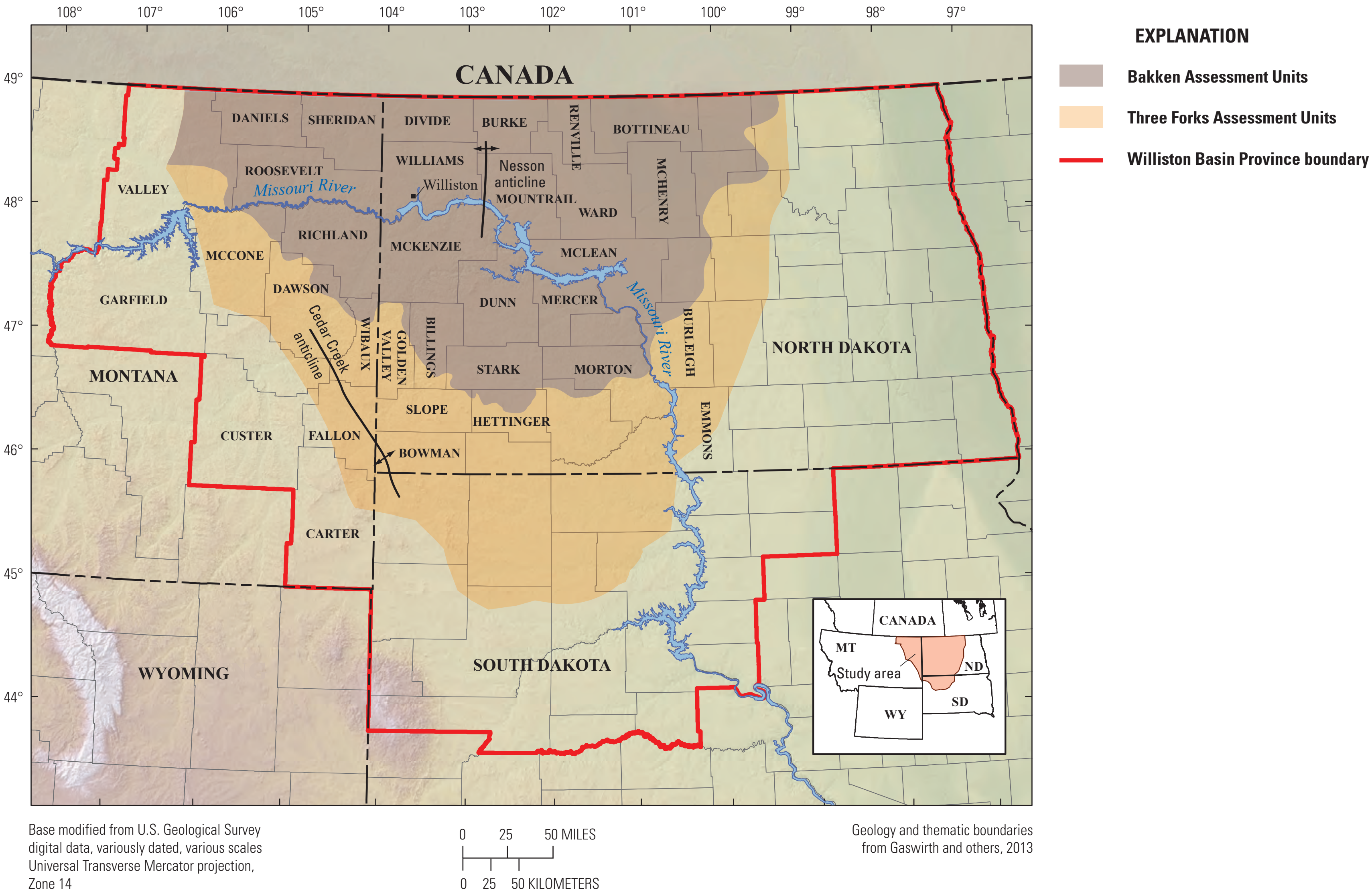

Figure 1. Locations of the Williston Basin Province, Bakken and Three Forks Assessment Units, and other geologic features in the United States (from Gaswirth and others, 2013).

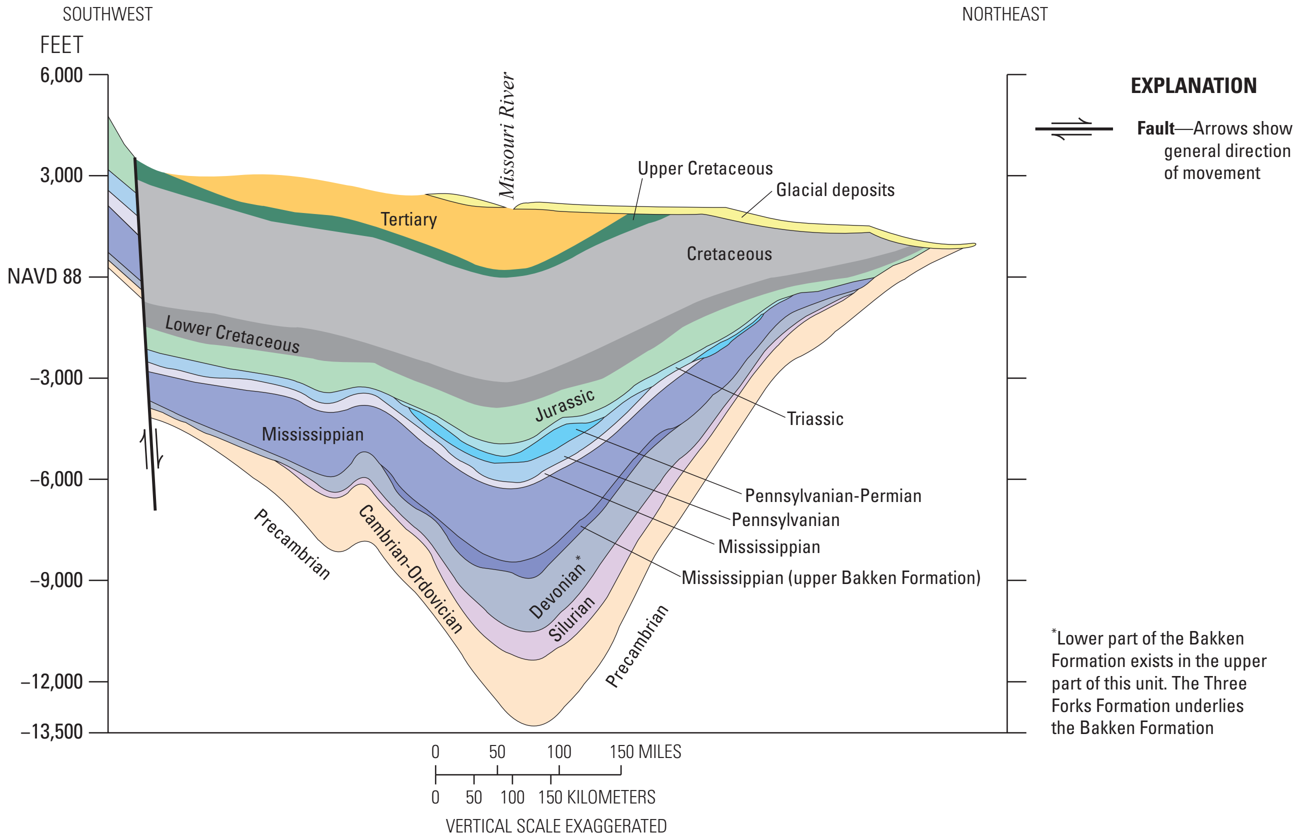

Figure 2. Stratigraphic units of the Williston Basin with geologic periods (from Long and others, 2014).

Cedar Creek anticline in southeastern Montana in the 1920s and 1930s (Bluemle, 2001). Several attempts at drilling the oil reserves about 40 miles east of Williston, N. Dak., were made between 1924 and 1951. In 1938, one of the first modern rotary rigs used in North Dakota drilled a deep test hole (more than 10,000 feet) near Tioga, N. Dak., but missed an oil reservoir at this location by about 1 mile (Bluemle, 2001).

The Clarence Iverson #1 well, which was drilled on the Nesson anticline in northwestern North Dakota, is credited as the first oil-producing well in the Williston Basin, and a granite monument marks the site about 10 miles south of Tioga, N. Dak. (Bluemle, 2001; Anna and others, 2011; American Oil and Gas Historical Society, 2013). On April 4, 1951, the Clarence Iverson #1 well was labeled the discovery well of the Williston Basin. A month later, multiple oil and gas companies and speculators leased more than 30 million acres of mineral rights in North Dakota, unofficially beginning the first of several oil booms in the Williston Basin (American Oil and Gas Historical Society, 2013).

Vertical drilling techniques were productive and profitable in several reservoirs in the Williston Basin but not the Bakken Formation. After 1951, annual oil production increased in North Dakota until 1966 to a high of 26 million barrels and then declined until 1974 to a low of 20 million barrels (Fischer and Bluemle, 1986). Production of 31 million barrels in 1979 surpassed the previous 1966 high, and new highs were recorded each year until 1984, when production rose to 53 million barrels (Fischer and Bluemle, 1986; Bluemle, 2001). Production then began to decline, falling to 28 million barrels in 1994 (Bluemle, 2001).

It was not until the late 1990s, when horizontal drilling, hydraulic fracturing, and other technologies became sufficiently developed, that production from the low-permeability and low-porosity Bakken Formation became economically feasible. Moreover, higher prices for domestic and foreign oil during this time allowed the more costly COG extraction techniques required in the Bakken Formation to become economically viable. Annual oil production in North Dakota alone increased from a nominal amount of 35 million barrels in 2005 to more than 400 million barrels in 2015 (fig. 4).

Increasing oil and gas production resulted in a complementary increase in the number of oil and gas wells and rigs used to drill the wells. The annual average oil and gas rig count in North Dakota began increasing after 2004 and substantially increased beginning in 2010 (Baker Hughes, 2017). In addition, the number of wells producing oil and gas in North Dakota began increasing after 2005 and considerably increased beginning in 2011 (fig. 4). The annual total oil production

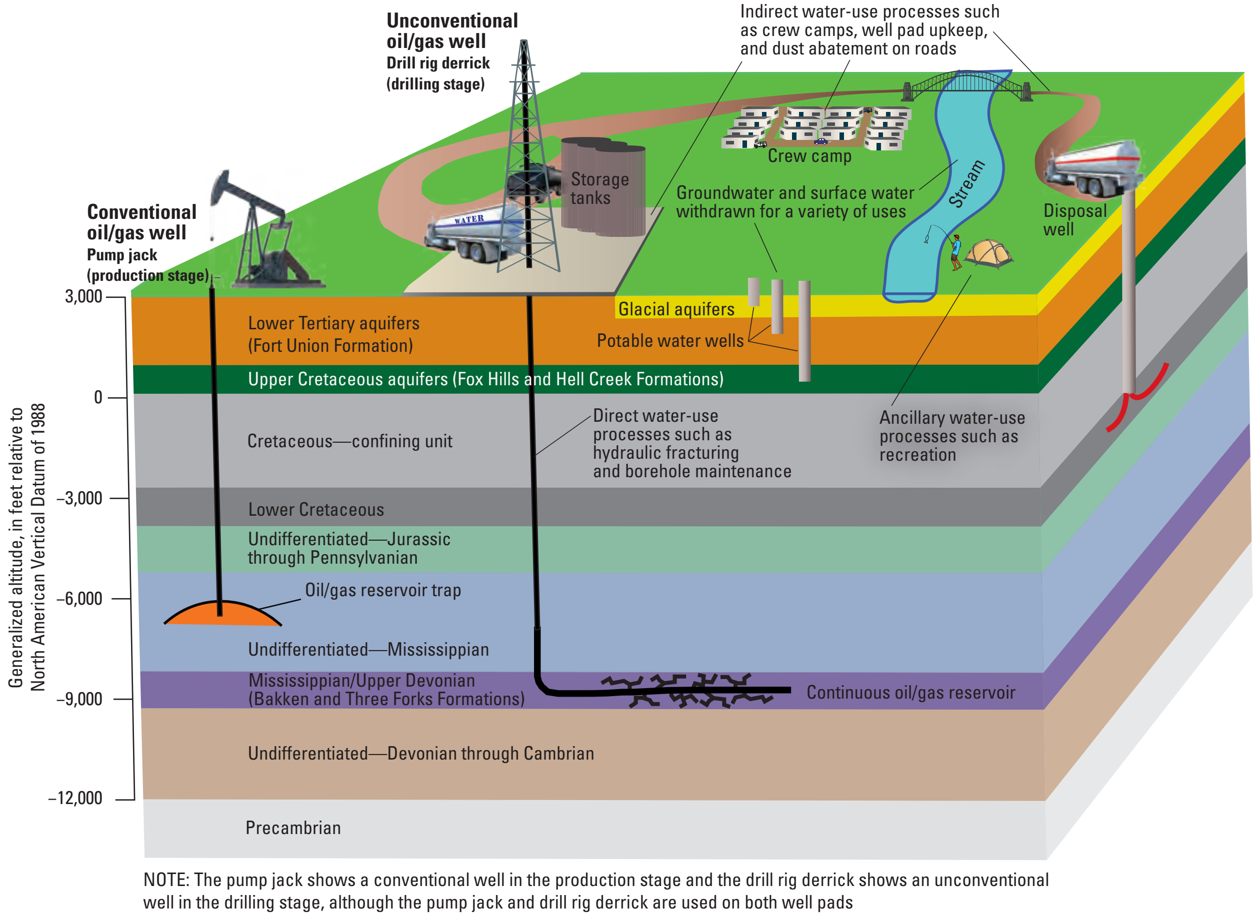

Figure 3. Various water and energy attributes of conventional and continuous (or unconventional) oil and gas development in the Williston Basin (from Carter and others, 2016).

more than doubled between 2010 and 2012 (fig. 4) with the introduction of horizontal drilling and hydraulic fracturing of deposits in the Bakken Formation in the Williston Basin (U.S. Energy Information Administration, 2013). In 2013, the USGS estimated 7.4 billion barrels of technically recoverable oil for the United States part of the Devonian-aged Three Forks Formation and the Devonian- and Mississippian-aged Bakken Formation of the Williston Basin (Gaswirth and others, 2013; figs. 2 and 3). In 2014, North Dakota surpassed oil production of 1 million barrels per day and ranked behind Texas as the second largest oil-producing State (U.S. Energy Information Administration, 2014). In 2019, about 90 percent of drilling in North Dakota targeted the Bakken and Three Forks Formations, and the all-time high number of wells producing from these formations in North Dakota was 15,943 (Helms, 2019).

Purpose and Scope

Direct, indirect, and ancillary water use associated with COG development in the Williston Basin was estimated in North Dakota and Montana from 2007 to 2017. Direct water use is defined as water used in a wellbore to complete a well, which includes water used for drilling, cementing, stimulating, and maintaining the well during production. Indirect water use is defined as water used at or near the well site, including water used for dust abatement, equipment cleaning, materials washing, worker sanitation, and site preparation. Ancillary water use is defined as all other water used during the life cycle of COG development that is not categorized as direct or indirect, such as additional local or regional water use resulting from a change (for example, population) related to COG development.

The water-use estimates presented in this report are compared with other published values for the Williston Basin, and the limitations of the water-use analysis for the Williston Basin are described. The water-use data used in the analysis includes data from 1980 to 2017 for some sources, which are available in a USGS data release (Dutton and others, 2019). Appendix 1 includes scripts coded in R (R Core Team, 2019) for executing the procedures used in the water-use analysis. Appendix 2 includes the estimates of water use by county and year and coefficients of water use per developed oil and gas well for the Williston Basin in North Dakota for 2007–17 and in Montana for 2008–16, which can be reproduced using the R scripts in appendix 1.

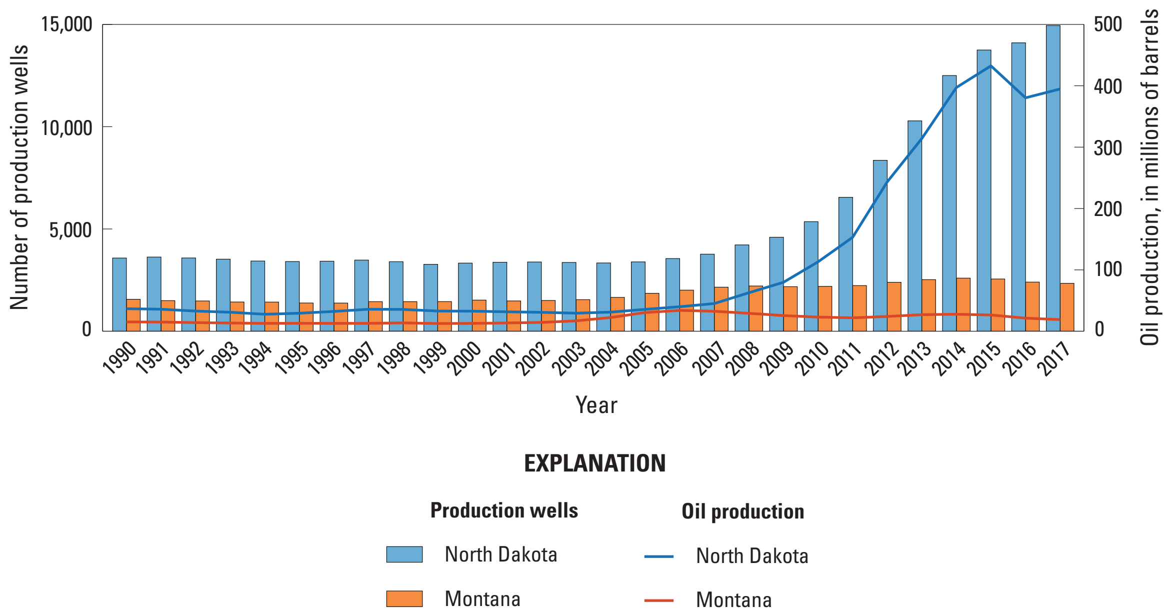

Figure 4. Number of production wells and volume of oil production in the Williston Basin in North Dakota and Montana from 1990 to 2017 (data from North Dakota Oil and Gas Division, 2019; Montana Board of Oil and Gas Conservation, 2019).

Methods for Analyzing Water Use

The analysis of water use associated with COG development presented in this report was based on an analytical framework (Valder and others, 2019) derived from a conceptual model (Valder and others, 2018). The conceptual model consisted of five components: (1) input data, (2) processes, (3) decisions, (4) model output, and (5) assessment outcomes. Decisions that could be made in the conceptual model were determined by the quality and quantity of data available for specific regions of COG extraction, which led to different outcomes of how completely the life cycle of water use associated with COG development could be assessed (Valder and others, 2018). It was assumed that an almost complete life cycle of water use associated with COG development in the Williston Basin could be assessed because the Williston Basin has a relatively large quantity and high quality of available water-use data. The analytical framework established a generic approach for estimating the water use—direct, indirect, and ancillary—related to COG development, including a generalized description of the input data; data processing, interpretation, and uncertainty analysis; and model output involved in the water-use analysis (Valder and others, 2019). The next sections describe the procedures for processing the data for analysis and the procedures for modeling the water use and uncertainty for the Williston Basin. The water-use analysis was executed with scripts (appendix 1) coded in R (R Core Team, 2019).

Procedures for Processing Data for Analysis

Several datasets and databases were processed for use in the analysis of direct, indirect, and ancillary water use associated with COG development in the Williston Basin (table 1) and are available in the accompanying data release (Dutton and others, 2019). North Dakota data on hydraulic fracturing treatment water, wellbore cementing (sacks of cement), and wellbore drilling (depth per well) were obtained from the North Dakota Industrial Commission (NDIC, 2018). Montana data on oil and gas well counts and hydraulic fracturing treatment water were provided as 1-square-mile aggregated values (table 1). These data for North Dakota and Montana were used to estimate direct water use. Reported water use by category, such as domestic or industrial, for water permits in North Dakota were acquired from the North Dakota State Water Commission (NDSWC, 2018). By pairing data on direct water use from the NDIC with water-use data from the NDSWC, indirect water use was estimated for North Dakota. Data from the NDSWC were paired with air temperature and precipitation data from the PRISM Climate Group (2018) and population data from the U.S. Census Bureau (2018) to estimate ancillary water use in North Dakota. Ancillary water use in Montana was estimated by pairing water-use data from the U.S. Geological Survey (2019) with data from the PRISM Climate Group and the U.S. Census Bureau (table 1).

Data on hydraulic fracturing water use were used to determine when COG development started the most recent (before 2017) boom in oil production in North Dakota and Montana (fig. 4). For North Dakota and Montana, data on hydraulic fracturing treatment water (table 1) were summed by year, resulting in one value of hydraulic fracturing water use per year. A regression model of the linear relation between hydraulic fracturing water use and year of observation was fit as a segmented, or broken-line, model using the "segmented" package for R (Muggeo, 2018). The broken-line model attempts to estimate a breakpoint (a change in slope) in the linear relation, meaning that data before and after the breakpoint could have different slopes fit by the model. The breakpoint was used to separate the water-use data into observations that preceded the boom in COG development and observations that followed. The observations after the breakpoint were used to estimate water use associated with COG development. The observations before the breakpoint were used to model baseline water use before the boom in COG development. The breakpoints used in the water-use analysis were 2007 for North Dakota and 2008 for Montana. For each State, the data from the breakpoint year and those that followed (2007–17 for North Dakota and 2008–16 for Montana) were used for estimating direct, indirect, and ancillary water use.

Not all data used in the water-use analysis were available on the basis of a given oil or gas well or for a given day or month (table 1), so water-use data were aggregated by county and year, which were the sampling units used in this analysis. However, each county did not necessarily have direct, indirect, or ancillary water use associated with COG development in each year, and each year did not necessarily have water use in each county. For example, a given county had direct and indirect water use in only 5 of 11 years, or a given year had water use in only 5 of 21 counties. North Dakota had 21 counties and 11 years (2007–17) included in the analysis but only 114 sampling units for analyzing direct and indirect water use. Likewise, Montana had 12 counties and 9 years included in the analysis (2008–16) but just 55 sampling units for analyzing direct and indirect water use.

Cementing and drilling water use were estimated from available data for North Dakota on the number of sacks of cement used per well for cementing the casings in the borehole and the measured depth reached per well in drilling the borehole (NDIC, 2018; table 1). This study assumed that cementing water use did not vary greatly by well depth and that water use per sack of cement averaged a ratio of water weight to cement weight of about 0.5. Sacks of cement were multiplied by 5 gallons of water, which is the volume of water with a weight equal to about one-half the weight of a sack of cement (about 47 pounds). This study assumed that drilling water use varied mostly by well depth and that water use per well averaged a ratio of water volume to borehole volume of about 2. Data were deficient for computing borehole volume per well, so it was assumed that wellbores averaged about 0.5 cubic foot of volume per 1 foot of depth. Well depth was multiplied by 7.5 gallons of water (about 1 cubic foot of water), which is the volume of water equal to about twice the average per-foot borehole volume.

Indirect water use (water used at oil and gas well sites for purposes other than drilling, cementing, or hydraulic fracturing wells) was estimated for North Dakota from available water-permit data (NDSWC, 2018; table 1). The water permits had water use reported by year for various categories, such as industrial or irrigation, in accordance with the water-use categories used for the USGS National Water Census (U.S. Geological Survey, 2019). The industrial water-use category included permits for water depots providing water for use at oil and gas well sites and permits to oil and gas companies using water for undefined purposes. Additionally, the

Table 1. Information and data sources used to estimate direct, indirect, and ancillary water use associated with continuous oil and gas development.

[IHS, Information Handling Services; PRISM, Parameter-elevation Regressions on Independent Slopes Model]

Information	Data source
Hydraulic fracturing treatment water per oil and gas well for North Dakota	North Dakota Industrial Commission (2018)
Number of sacks of cement and depth per oil and gas well for North Dakota	
Oil and gas well count per square mile for Montana	IHS Markit (2018) well database
Hydraulic fracturing treatment water per square mile for Montana	
Reported water use per water permit for North Dakota	North Dakota State Water Commission (2018)
Air temperature and precipitation per 4-kilometer grid	PRISM Climate Group (2018)
Number of persons per county	U.S. Census Bureau (2018)
Estimated water use per county for Montana	U.S. Geological Survey (Kenny and others, 2009; Maupin and others, 2014; Dieter and others, 2018)

irrigation water-use category included some permits temporarily repurposed as water depots. This study assumed that water use reported by these permits in the industrial and irrigation water-use categories was associated with COG development (Chris Bader, NDSWC, oral commun., 2018). The reported water use from the applicable permits was summed by county and year and was assumed to represent the total water use at oil and gas well sites in a given county and year. Direct water use (water used for hydraulic fracturing, cementing, or drilling wells) in a given county and year was subtracted from total water use, and the remainder was assumed to represent indirect water use. For this study, direct and indirect water use in combination represent the part of the USGS mining water-use category related to COG development and is herein referred to as COG mining water use.

Water-permit data (NDSWC, 2018; table 1) also were used to estimate ancillary water use in North Dakota. Ancillary water use is other water used during the life cycle of COG development that is not categorized as direct or indirect, such as additional local or regional water use resulting from a change related to COG development. By default, ancillary water use is not combined with direct and indirect water use as a part of the USGS mining water-use category; instead, ancillary water use is represented separately as parts of other nonmining water-use categories. This study assumed that the domestic and public-supply water-use categories were most directly affected by COG development because of attendant population growth, and these categories are herein referred to as population-based ancillary water use. Additionally, this study assumed that industrial, mining other than COG (herein referred to as non-COG mining), and thermoelectric power water use might be indirectly affected by COG development because of potential accompanying economic growth, and these categories are herein referred to as other potential ancillary water use. However, the relation of the other potential ancillary water-use categories to COG development was more uncertain than for the population-based categories (domestic and public supply).

The breakpoint analysis was used to account for preexisting trends in the ancillary water-use categories (domestic, public supply, industrial, non-COG mining, and thermoelectric power) before the boom in COG development. A linear regression model of water use for these categories in relation to a combination of year and county of observation, total annual precipitation, mean annual temperature, and population was fit for the years preceding the breakpoint (2007 for North Dakota and 2008 for Montana). The antecedent trend in water use was extrapolated forward into the years after the breakpoint (2007–17 for North Dakota and 2008–16 for Montana). The estimates of water use in the years after the breakpoint were subtracted from the reported water use from the NDSWC water permits for North Dakota and the USGS water-use data for Montana, which removed the effect of population and climate on the ancillary water-use categories before the boom in COG development. The remaining water use was attributed to COG development based on the assumption that no factors other than COG development affected the ancillary water-use categories. This assumption may be more applicable for some categories than others. For example, thermoelectric power, an ancillary water-use category, is affected by energy demands in other States connected to the same power transmission grid serviced by power plants in Montana and North Dakota.

Ancillary water use was more difficult to model than direct or indirect water use because data on direct and indirect water use represented water use at oil and gas wells developed in a given county and year, whereas data on ancillary water use represented water withdrawals for any number of water-use categories (for example, domestic or industrial) in a given county and year irrespective of development of oil and gas wells. In other words, in a given county and year, ancillary water-use data did not necessarily coincide with direct and indirect water-use data. Regression models were used to estimate water use in counties with oil and gas wells developed in a given year, which required the assignment of ancillary water-use data from the counties with observed water withdrawals to the counties with developed wells. For each State, observed water use by category, and the number of oil and gas wells, were summed by year. Observed water use then was assigned proportionally to the counties with developed oil and gas wells by using the ratio of the number of wells in each county to the number of developed wells in all counties. Therefore, observed water use was restricted to the sampling units (county and year) used in the regression models for estimating water use. Estimated water use in a given year then was reassigned proportionally to the counties with observed water withdrawals by using the ratio of the volume of observed water withdrawals in each county to the volume of observed water withdrawals in all counties. In other words, ancillary water use in a given year was modeled in the counties with developed oil and gas wells, and water-use estimates for each category then were reallocated to the counties with observed water withdrawals for each respective category.

Data on indirect water use and cementing and drilling water use were not available for Montana, so coefficients of water use per well (in million gallons [Mgal]) derived from the analysis for North Dakota for the cementing, drilling, and indirect water-use categories were applied to the water-use analysis for Montana. For example, if the coefficient for indirect water use in North Dakota was 0.5 Mgal per well and a sampling unit in Montana had 10 oil and gas wells, then the indirect water use for the sampling unit was estimated at 5 Mgal (10 wells multiplied by 0.5 Mgal per well). Additionally, annual ancillary water-use data were not available for Montana; therefore, water-use data from the USGS National Water Census, compiled every 5 years for water-use categories such as domestic, public supply, industrial, mining, and thermoelectric power (U.S. Geological Survey, 2019), were used (table 1). Annual values were estimated by linear interpolation between every fifth year.

Procedures for Modeling Water Use and Uncertainty

Linear and quantile regression models of water use in relation to the number of oil and gas wells developed were fit for the direct, indirect, and ancillary water-use categories for each State. A 95-percent confidence interval for each parameter estimate (coefficient) from the linear regression models was computed as a measure of uncertainty. However, a 95-percent confidence interval alone may not adequately capture the model uncertainty because linear regression estimates the mean of the sampling distribution, which may not sufficiently represent the data variance. Additional information on uncertainty can be gained from modeling other distribution parameters, so quantile regression models of the 5th, 50th (median), and 95th percentiles also were fit using the "quantreg" package for R (Koenker, 2018). Quantile regression can provide a more detailed understanding of estimated water use at the extremes of the sampling distribution or for a skewed sampling distribution (Koenker, 2005). Water-use coefficients from the linear and quantile regression models are in million gallons per well, whereas water-use estimates from the regression models are in million gallons per sampling unit (county and year). Therefore, water use was estimated for each sampling unit that had observed water use.

The regression models were not fit with other explanatory variables specific to the Williston Basin; instead, they were intentionally generalized to apply to other regions of COG extraction around the United States, where data availability for explanatory variables likely differs. Water use associated with COG development likely depends on many factors that vary from region to region and that would be prohibitive to quantify nationally with any certainty, such as social or economic factors that drive the price of oil and gas and the cost of water and labor, or geologic factors that affect the development and implementation of new or improved extraction techniques. Additionally, for the purpose of developing water-use coefficients applicable to other regions, the county and year of the water-use data in the Williston Basin were not explicitly factored into the regression models for several reasons. A spatial relation of water use to county in the Williston Basin would not be analogous to counties with COG development in another region, and a temporal relation of water use to year in the Williston Basin might not be concurrent with years of COG development in another region. Likewise, there is not necessarily a clear trend toward more or less water use because various hydraulic fracturing methods may be implemented at the same time but in different places.

To assess uncertainty in the estimates from the regression models of direct, indirect, and ancillary water use, leave-one-out cross-validation was used (Hastie and others, 2009). Leave-one-out cross-validation is an iterative process that leaves out a sampling unit and fits a model to all the remaining sampling units. The fitted model then was used to predict the observation for the left-out sampling unit. This process was repeated for each sampling unit. For example, for direct water use in North Dakota, there were 114 sampling units. Sampling unit 1 was left out and a model was fit to sampling units 2–114. The fitted model then was used to predict the observed water use of sampling unit 1. Then, sampling unit 2 was left out and a model was fit to sampling units 1 and 3–114. The fitted model then was used to predict the observed water use of sampling unit 2. This process proceeded until all 114 sampling units had been left out once, resulting in 114 estimates of water use. These estimates were then validated against the 114 observations of water use for the sampling units.

Model performance was evaluated with three goodness-of-fit metrics used to compare the estimates and observations of water use. The first metric was the coefficient of determination (R^2), which is the square of the Pearson product-moment correlation coefficient (Legates and McCabe, 1999). An R^2 value greater than 0.5 may be considered good, particularly with the understanding that for this study relating water use to the number of developed oil and gas wells disregards meaningful spatial and temporal variability in the data. The second metric was root mean square error (RMSE), which is the square root of the mean square error of the regression model (Legates and McCabe, 1999). This metric also can be defined as the standard deviation of the residuals (unexplained variance) of the regression model (Moriasi and others, 2007). A good value for RMSE is difficult to determine because the values are specific to the modeled data. For example, an RMSE of 1 Mgal per well for data with a mean for the observations of 10 Mgal per well (10 percent of the mean) would be much better than an RMSE of 0.5 Mgal per well for data with a mean for the observations of 1 Mgal per well (50 percent of the mean). Therefore, this study used RMSE in comparison to the parameter estimate from the regression model as a measure of goodness-of-fit. The third metric was RMSE relative to the standard deviation of the observations (RSR), which is the relation of the standard deviation of the unexplained variance in the regression model to the standard deviation of the observations (Moriasi and others, 2007). For example, if the RMSE was 1 Mgal per well, whereas the standard deviation of the observations was 2 Mgal per well, then the RSR would be 0.5. A RSR value less than 0.5 is considered good (Moriasi and others, 2007).

Population was modeled to provide context for analysis of ancillary water use for the domestic and public-supply categories, which are based on population, by demonstrating whether population growth could be explained by COG development. Population in North Dakota and Montana was estimated with a linear regression model of the number of persons in relation to the number of oil and gas wells developed in the year of the observed population and in the 1, 2, and 3 years before the observed population in case population growth had a lagged response to COG development. In other words, population in a given year might be better explained by prior-year COG development.

Results of Water-Use Analysis

The water-use analysis for the Williston Basin contains three elements: (1) estimates of water use, in million gallons, by county and year; (2) coefficients of water use from the regression models, in million gallons per oil and gas well developed; and (3) performance (based on goodness-of-fit metrics) of the regression models in estimating (predicting) the observed water use. Results are presented in the following sections for North Dakota and Montana, including a discussion about the differences between the States. Comparisons with estimates of water use associated with COG development from other studies and limitations of the water-use analysis used for this study are also presented in the following sections. Hereafter, unless specified otherwise, ancillary water use refers only to the domestic and public-supply (population-based) categories.

Estimates of Water Use by County and Year

Estimates (predictions) from the linear and quantile regression models of direct, indirect, and ancillary water use in the Williston Basin were summed across all counties and averaged over all years in the analysis for each State (table 2) to facilitate comparison of the water-use categories between the States. These mean annual estimates apply to 2007–17 for North Dakota and 2008–16 for Montana, and they are specific to each State because North Dakota and Montana had different data sources for water use associated with COG development.

Water use for COG development (direct and indirect, or COG mining; table 2) in the Williston Basin was not distributed uniformly between the States—more than 95 percent of direct and indirect water use per year was in North Dakota. Mean annual direct and indirect water use for COG development in North Dakota was estimated at 4,512 Mgal per year (Mgal/yr), with a 95-percent confidence interval of 4,021–5,152 Mgal/yr, and in Montana was estimated at 196 Mgal/yr, with a 95-percent confidence interval of 189–203 Mgal/yr. Ancillary water use (for domestic and public supply) had an estimated annual mean of 2,753 Mgal/yr in North Dakota and 396 Mgal/yr in Montana (table 2).

Direct and indirect water use in North Dakota and Montana varied considerably by year (fig. 5). Direct water use (for hydraulic fracturing, cementing, and drilling wells) had an upward trend that peaked in 2014 at 9,239 Mgal in North Dakota (fig. 5*A*) and 385 Mgal in Montana (fig. 5*B*). Mean annual direct water use in North Dakota was 3,945 Mgal/yr and in Montana was 155 Mgal/yr (table 2), which is about 25 times more in North Dakota than in Montana. Both States declined in direct water use after the peak in 2014; however, direct water use in North Dakota expanded to 7,450 Mgal in 2017 (fig. 5*A*), whereas Montana's part of direct water use in the Williston Basin continued to contract (fig. 5*B*). The variability in direct water use reflects differences in the number of oil and gas wells developed per year and the volume of water used per well among years. Mean annual indirect water use was 567 Mgal/yr in North Dakota and 41.2 Mgal/yr in Montana (table 2). Indirect water use varied less by year than did direct water use. The standard deviation relative to the mean (coefficient of variation) for direct and indirect water use was about 0.8 and 0.55, respectively, in North Dakota and about 0.95 and 0.75, respectively, in Montana, indicating less variation by year in indirect water use than in direct water use for hydraulic fracturing in both States. Additionally, the ratio of indirect to direct water use was larger in Montana (about 0.25) than in North Dakota (about 0.15). The difference in ratios probably results from applying the indirect water-use coefficient from the North Dakota analysis to Montana, which lacked data to analyze indirect water use.

Ancillary water use (for domestic and public supply) varied by year in North Dakota and Montana (fig. 5) but trended upward over the years in both States. For North Dakota, the ratio of ancillary water use to direct and indirect water use was about 0.6, but it was about 2 for Montana. The difference in ratios most likely results from the difference in data sources between the States. The source for Montana was the USGS water-use data (U.S. Geological Survey, 2019), which compiles water use by county, whereas the source for North Dakota was the NDSWC water permits (North Dakota State Water Commission, 2018), which reports water use by individual water permit holders. Ancillary water use continued to increase after the peak in direct water use in North Dakota and Montana in 2014 (fig. 5). This trend in water use demonstrates that ancillary water use did not correspond to the years when COG mining (direct and indirect) water use was highest; the effects of population on ancillary water use associated with COG development may occur in later years than when oil and gas wells are developed.

Direct and indirect water use in North Dakota and Montana also varied substantially by county (fig. 6). Direct water use was greatest in McKenzie, Williams, Mountrail, and Dunn Counties in North Dakota, ranging from 1,438 to 611 Mgal/yr (fig. 6*A*), and greatest in Richland and Roosevelt Counties in Montana at 106 and 48.0 Mgal/yr, respectively (fig. 6*B*). A larger proportion of counties in the Williston Basin in North Dakota (13 of 21 counties) had at least 1 Mgal/yr of direct water use than in Montana (4 of 13 counties). The county with the most direct water use in North Dakota used about 15 times more than the county with the most direct water use in Montana. The variation in direct water use reflects differences in both the number of oil and gas wells developed per county and the volume of water used per well among counties (for example, because of differences in the geologic formations containing COG reservoirs). In North Dakota, McKenzie, Mountrail, Williams, and Dunn Counties had the most indirect water use (water used at oil and gas well sites for purposes other than drilling, cementing, or hydraulic fracturing wells), ranging from 179 to 98.7 Mgal/yr (fig. 6*A*), and in Montana, Richland and Roosevelt Counties had the most at 25.5 and 8.33 Mgal/yr, respectively (fig. 6*B*). Indirect water use varied less by county than did direct water use. The coefficient of variation for direct and indirect water use was

Table 2. Mean annual estimates from linear and quantile regression models of direct, indirect, and ancillary water use associated with continuous oil and gas development in the Williston Basin in North Dakota (2007–17) and Montana (2008–16).

[Values in million gallons per year. COG, continuous oil and gas]

Water use	Mean	95-percent confidence interval limits		Percentiles		
		Lower	Upper	5th	50th	95th
North Dakota						
Direct and indirect water use						
Direct—All	3,945	3,510	4,515	2,381	3,360	6,473
Direct—Hydraulic fracturing	3,776	3,411	4,236	2,227	3,204	6,294
Direct—Cementing	15.4	13.6	17.7	14.7	15.3	16.1
Direct—Drilling	153	136	177	146	152	164
Indirect	567	512	637	401	596	780
Direct and indirect—COG mining	4,512	4,021	5,152	2,781	3,956	7,253
Ancillary (population-based) water use						
All	2,753	2,576	2,961	1,473	2,544	4,530
Domestic	946	721	1,428	742	891	1,287
Public supply	1,808	1,706	1,923	1,350	1,801	2,223
Other potential ancillary water use						
Industrial	619	499	825	331	540	1,129
Non-COG mining	1,869	1,182	4,619	1,690	1,905	2,074
Thermoelectric power	443	343	629	247	348	740
Montana						
Direct and indirect water use						
Direct—All	155	151	158	99.9	156	259
Direct—Hydraulic fracturing	140	110	197	86.9	142	245
Direct—Cementing	1.29	1.20	1.41	1.29	1.29	1.29
Direct—Drilling	12.8	11.9	14.0	12.8	12.8	12.8
Indirect	41.2	38.1	44.8	41.2	41.2	41.2
Direct and indirect—COG mining	196	189	203	141	197	300
Ancillary (population-based) water use						
All	396	240	551	247	403	526
Domestic	50.4	12.7	88.1	46.2	54.1	54.7
Public supply	345	211	480	242	331	457
Other potential ancillary water use						
Industrial	42.3	6.16	78.7	33.0	38.6	51.0
Non-COG mining	1,851	1,261	2,440	1,309	1,894	2,330
Thermoelectric power	9,195	6,815	11,576	4,687	9,105	11,210

about 2 and 1.8, respectively, in North Dakota and about 2.4 and 1.9, respectively, in Montana, which indicates less variation by county in indirect water use than in direct water use for hydraulic fracturing in both States, most likely because hydraulic fracturing methods vary greatly among oil and gas operators throughout the region. However, comparisons between the relative measures of variability by county and variability by year suggest that indirect and direct water use vary less temporally than spatially.

Ancillary water use (for domestic and public supply) also varied by county in North Dakota and Montana (fig. 6) and was dispersed unevenly across the counties in each State. A much larger proportion of counties in Montana had ancillary water use that exceeded direct and indirect water use (9 of 12 counties) in comparison to North Dakota (10 of 22 counties). The difference between the States suggests that ancillary water use in Montana was more dispersed than in North Dakota. However, in North Dakota, more than 50 percent of

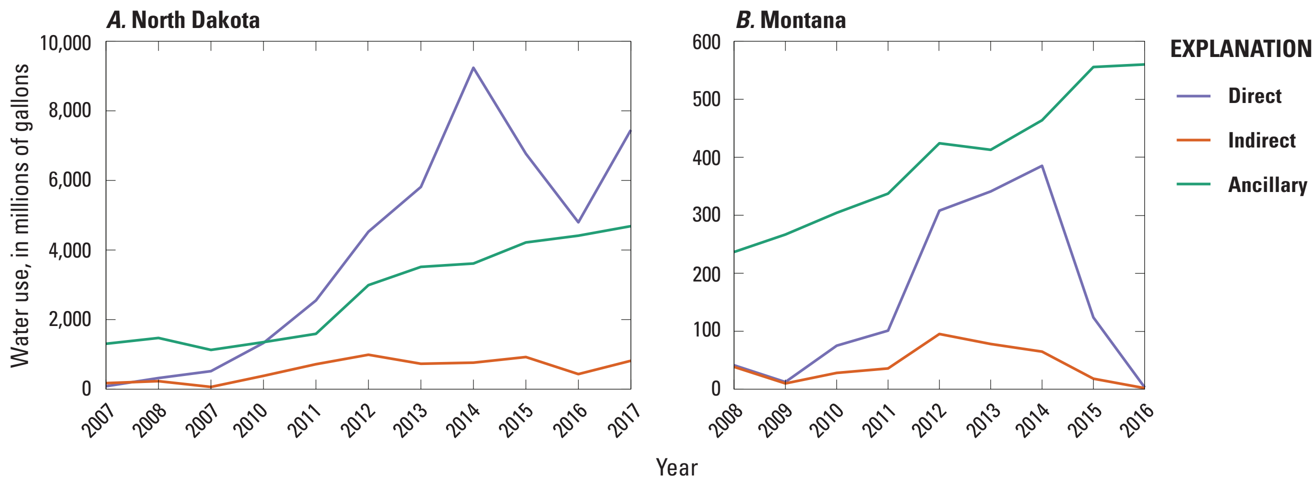

Figure 5. Annual estimates of direct, indirect, and ancillary water use in the Williston Basin. *A*, North Dakota from 2007 to 2017. *B*, Montana from 2008 to 2016.

ancillary water use was in Mercer and Williams Counties; in Montana, Roosevelt County alone had about 45 percent of ancillary water use. In addition, McKenzie County in North Dakota and Richland County in Montana (the counties with the most direct water use in their respective State) had ancillary water use that was about 10 and 25 percent of their respective direct water use. This pattern of water use also demonstrates that ancillary water use did not correspond to the counties where direct and indirect water use (COG mining) was largest; the populations affecting ancillary water use associated with COG development may occur in counties distant from where oil and gas wells are developed.

Analysis comparing ancillary water use to direct water use indicated that ancillary water use did not change immediately with change in direct water use but instead changed after a time lag. In the 21 North Dakota counties in the Williston Basin with at least 1 developed oil and gas well, population increased from 232,000 in 2007 to 299,000 in 2015 and then decreased to 293,000 in 2017 (fig. 7*A*; U.S. Census Bureau, 2018). In the 12 Montana counties in the Williston Basin with at least 1 developed oil and gas well, population increased between 2008 and 2015 from 60,000 to 66,000 and then decreased in 2016 to 65,000 (fig. 7*B*; U.S. Census Bureau, 2018). This trend in population growth was best explained with a linear regression model of the relation between the current-year population and the number of oil and gas wells developed 2 years previously (a 2-year lag); for example, population in 2015 was best explained by the number of wells developed in 2013. The mean estimates from the linear regression models for North Dakota (R^2=0.92) and Montana

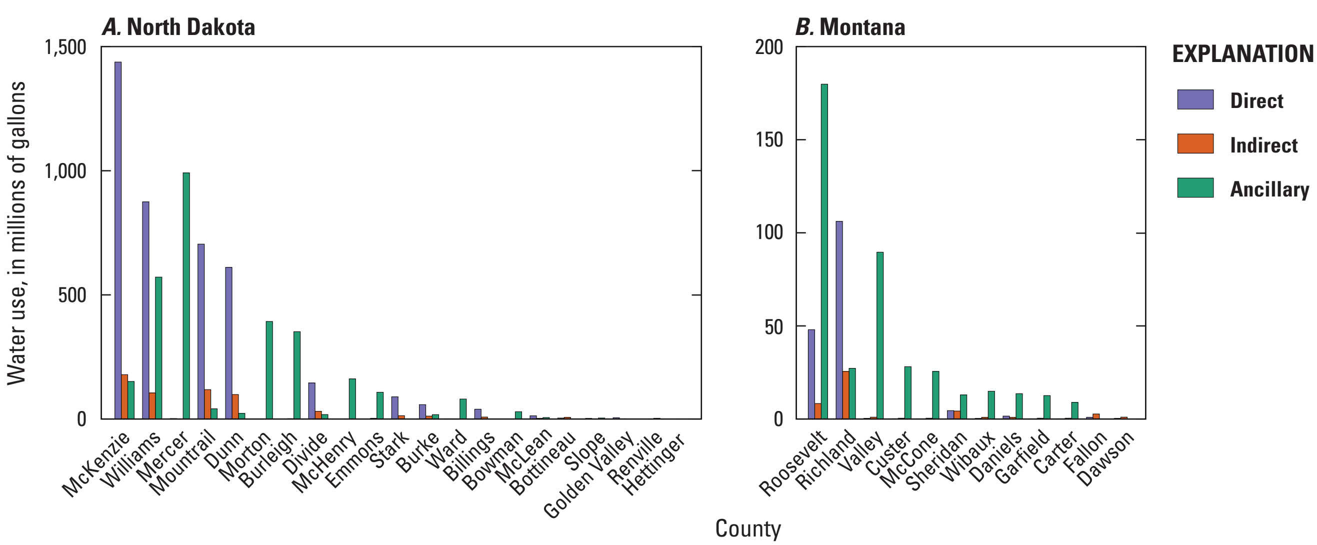

Figure 6. Mean annual estimates of direct, indirect, and ancillary water use by county in the Williston Basin. *A*, North Dakota from 2007 to 2017. *B*, Montana from 2008 to 2016.

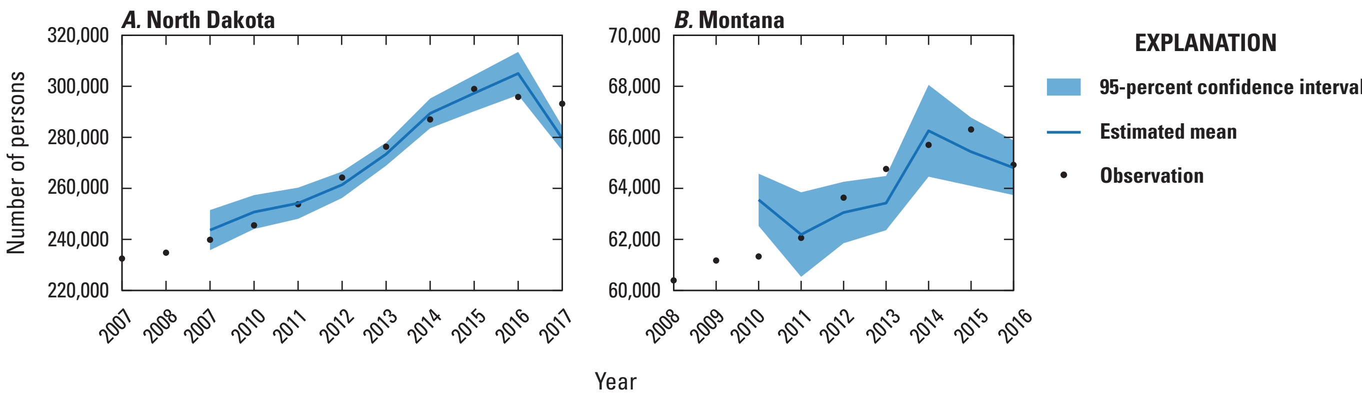

Figure 7. Linear regression models of population as explained by number of oil and gas wells developed per year in the Williston Basin. *A*, North Dakota from 2007 to 2017. *B*, Montana from 2008 to 2016.

(R^2=0.60) were about 27 and 22 additional persons per oil and gas well developed, respectively. Even though direct water use declined at least temporarily after the peak in 2014, ancillary water use continued to increase because population growth lagged well development.

Estimates of water use by county and year used in the analyses for North Dakota and Montana are available in appendix 2. These estimates include direct and indirect water use (COG mining) and ancillary water use for the domestic, public supply, industrial, non-COG mining, and thermoelectric power categories. Ancillary water use for the domestic and public-supply categories (population-based) is directly related to COG development, whereas the other potential ancillary water-use categories (industrial, non-COG mining, or thermoelectric power) are indirectly associated with COG development. The values for the ancillary water-use categories (appendix 2) represent the part of those categories attributable to COG development as determined by the water-use analysis for the Williston Basin. Total water use for the ancillary categories, which also includes water use not related to COG development, can be found in the USGS water-use data (U.S. Geological Survey, 2019).

Coefficients of Water Use per Oil and Gas Well

Coefficients from the linear and quantile regression models of direct, indirect, and ancillary water use in the Williston Basin (parameter estimates of water use per oil and gas well developed) were produced in the analysis for each State (table 3). The coefficients are specific to each State because data sources for water use associated with COG development were different for North Dakota and Montana. However, the coefficients of direct water use for cementing and drilling and of indirect water use were the same for each State because the coefficients from the North Dakota water-use analysis were applied to the analysis for Montana, which lacked data on these water-use categories.

The coefficient (mean estimate) from the linear regression model of direct water use was 3.86 Mgal per well and hydraulic fracturing water use was 3.70 Mgal per well for North Dakota (table 3). The mean estimate of cementing and drilling water use was 0.014 and 0.142 Mgal per well, respectively, and totaled about 0.156 Mgal per well, which was smaller than 5 percent of the mean estimate of hydraulic fracturing water use. More than 95 percent of direct water use was for hydraulic fracturing. The mean estimate of direct water use had a 95-percent confidence interval of 3.48–4.23 Mgal per well. From the quantile regression model, the 5th, 50th, and 95th percentile estimates of direct water use were 1.28, 3.24, and 7.66 Mgal per well, respectively. In other words, 95 percent of the oil and gas wells had at least 1.28 Mgal of direct water use, 50 percent had at least 3.24 Mgal, and 5 percent had at least 7.66 Mgal. Direct water use at the lower and higher percentile estimates were outside the confidence interval for the mean estimate, which means that spatial, temporal, or both variations are characteristic of direct water use in North Dakota.

For North Dakota, the coefficient from the linear regression model of indirect water use was 0.453 Mgal per well, with a 95-percent confidence interval of 0.415–0.492 Mgal per well (table 3). The 5th, 50th, and 95th percentile estimates from the quantile regression model were 0.327, 0.555, and 0.866 Mgal per well, respectively.

Direct and indirect (COG mining) water use had a mean estimate of about 4.31 Mgal per well in North Dakota. The 5th percentile estimate of COG mining water use was about 1.61 Mgal per well, or less than 50 percent of the mean estimate, whereas the 95th percentile estimate was 8.52 Mgal per well, or almost 100 percent higher than the mean estimate. The mean estimate of COG mining water use was about 10 percent greater than the mean estimate of direct water use, meaning that estimates of only direct water use could underestimate total water use for COG development, which includes water used for completing a well and other water used at the well site, by about 10 percent.

Table 3. Coefficients from linear and quantile regression models of direct, indirect, and ancillary water use per oil and gas well developed in the Williston Basin in North Dakota and Montana.

[Values in million gallons per well. COG, continuous oil and gas]

Water use	Mean	95-percent confidence interval limits		Percentiles		
		Lower	Upper	5th	50th	95th
North Dakota						
Direct and indirect water use						
Direct—All	3.86	3.48	4.23	1.28	3.24	7.66
Direct—Hydraulic fracturing	3.70	3.33	4.08	1.12	3.09	7.49
Direct—Cementing	0.014	0.014	0.014	0.014	0.014	0.015
Direct—Drilling	0.142	0.140	0.143	0.134	0.139	0.164
Indirect	0.453	0.415	0.492	0.327	0.555	0.866
Direct and indirect—COG mining	4.31	3.90	4.72	1.61	3.79	8.52
Ancillary (population-based) water use						
All	2.03	1.76	2.31	1.32	1.69	5.33
Domestic	0.755	0.679	0.830	0.540	0.639	2.02
Public supply	1.28	1.08	1.48	0.782	1.05	3.64
Other potential ancillary water use						
Industrial	0.354	0.255	0.452	0.131	0.416	1.70
Non-COG mining	1.13	0.934	1.33	0.730	0.906	3.23
Thermoelectric power	0.168	0.077	0.259	0.017	0.178	0.486
Montana						
Direct and indirect water use						
Direct—All	2.04	1.80	2.28	0.703	1.72	3.15
Direct—Hydraulic fracturing	1.88	1.65	2.12	0.547	1.56	2.99
Direct—Cementing	0.014	0.014	0.014	0.014	0.014	0.014
Direct—Drilling	0.142	0.142	0.142	0.142	0.142	0.142
Indirect	0.453	0.453	0.453	0.453	0.453	0.453
Direct and indirect—COG mining	2.49	2.26	2.73	1.16	2.17	3.60
Ancillary (population-based) water use						
All	2.43	1.76	3.11	2.02	2.35	2.35
Domestic	0.306	0.205	0.406	0.188	0.309	0.309
Public supply	2.13	1.54	2.71	1.83	2.04	2.04
Other potential ancillary water use						
Industrial	0.260	0.193	0.326	0.186	0.242	0.242
Non-COG mining	10.8	7.66	13.9	8.87	10.2	14.8
Thermoelectric power	62.0	48.3	75.7	50.5	64.9	64.9

The mean estimate of ancillary water use (for domestic and public supply) in North Dakota was 2.03 Mgal per well, with a 95-percent confidence interval of 1.76–2.31 Mgal per well (table 3). Public supply water use was almost 75 percent higher than domestic water use. Other potential ancillary water use included 0.354, 1.13, and 0.168 Mgal per well (mean estimates) for industrial use, non-COG mining use, and thermoelectric power use, respectively. The 5th, 50th, and 95th percentile estimates for other potential ancillary water use ranged from 0.017 Mgal per well for thermoelectric power to 3.23 Mgal per well for non-COG mining.

For Montana, the linear regression model of hydraulic fracturing water use had a mean estimate of 2.04 Mgal per well (table 3), which was almost 50 percent lower than the estimate for North Dakota. The 95-percent confidence interval for the mean estimate was 1.80–2.28 Mgal per well.

The ratio of the range of the 95-percent confidence interval to the mean estimate was higher for Montana (about 0.25) than for North Dakota (about 0.2). The greater uncertainty of the mean estimate in Montana compared to North Dakota could be explained by the difference in data sources between the States; variability in the hydraulic fracturing methods used in each State; or differences in the geologic formations containing COG reservoirs in Montana and North Dakota that affect the extraction of oil and gas. The 95th percentile estimate from the quantile regression model of direct water use for hydraulic fracturing was 2.99 Mgal, or less than 50 percent of the 95th percentile estimate for North Dakota. However, if the data on hydraulic fracturing water use in North Dakota and Montana were combined in the analysis, then estimates for the Williston Basin in its entirety would more closely match the estimates for North Dakota. This closer match likely is because of the difference in the number of oil and gas wells developed between the States—North Dakota had about 15 times as many wells developed as Montana.

The mean estimates from the linear regression models of cementing and drilling water use in Montana were the same as the estimates in North Dakota (table 3) because the coefficients from the North Dakota water-use analysis were applied to the analysis for Montana, which lacked data on cementing and drilling. Moreover, because the North Dakota coefficients were applied uniformly to oil and gas wells in Montana, estimates per well did not vary, so the limits of the confidence interval did not differ from the mean estimate.

The mean estimate from the linear regression model of indirect water use in Montana also was not different from the estimate in North Dakota (table 3). The coefficient from the North Dakota analysis was applied to the water-use analysis for Montana because Montana had no data from water permits. The limits of the confidence interval were the same as the mean estimate of indirect water use because there was no variation in the per-well estimates.

Direct and indirect water use in Montana had a mean estimate of 2.49 Mgal per well, although about 25 percent of this value is based on the coefficients of cementing and drilling water use and indirect water use applied from the North Dakota analysis. The 95th percentile estimate of direct and indirect water use was about 3.60 Mgal per well.

The mean estimate of ancillary water use (for domestic and public supply) in Montana was 2.43 Mgal per well, with a 95-percent confidence interval of 1.76–3.11 Mgal per well (table 3). Domestic water use was less than 25 percent of public supply water use. Other potential ancillary water use included mean estimates of 0.260, 10.8, and 62.0 Mgal per well for industrial, non-COG mining, and thermoelectric power, respectively. The mean and percentile estimates of some ancillary water-use categories for Montana ranged from double to an order of magnitude larger than for North Dakota most likely because of the difference in data sources between the States.

Performance of Regression Models of Water Use

Model performance was evaluated with goodness-of-fit metrics (R^2, RMSE, and RSR) computed from leave-one-out cross-validation for the linear and quantile regression models of direct, indirect, and ancillary water use (table 4); performance metrics of just the linear regression models are presented in table 4. The values of R^2 differed less among water-use categories within a State than between the States (table 4). The differences in values of R^2 between the States may result from differences in sample sizes. For example, North Dakota had 114 sampling units for direct water use compared to 55 for Montana. In addition, goodness-of-fit metrics were not computed for the model fit of cementing and drilling water use and indirect water use in Montana (table 4) because the mean estimates for these water-use categories from the North Dakota analysis were applied to the water-use analysis for Montana.

The model performance for direct water use in both States was good (table 4), with an R^2 of 0.91 for North Dakota and 0.87 for Montana. The minor difference in performance may reflect a qualitative difference of the data source for Montana (data were available as 1-square-mile aggregated values for this study) compared to North Dakota (data originated from the individual oil and gas well operators), or it also could result from the difference in sample sizes between the States. However, the model fits of direct water use to number of oil and gas wells developed were reasonable for both North Dakota (fig. 8*A*) and Montana (fig. 8*B*), and the differences between estimated and observed values of water use were not qualitatively greater for Montana (fig. 9*B*) than for North Dakota (fig. 9*A*). For North Dakota, values of RMSE for the direct water-use categories ranged from 0.008 Mgal per well for cementing water use to 1.69 Mgal per well for direct water use, ranging from about 45 to 60 percent of their respective mean coefficients (also in million gallons per well; table 3). Values of RSR for the direct water-use categories ranged from 0.10 to 0.40, meaning that the standard deviation of the residuals from a linear regression model of a water-use category ranged from 10 to 40 percent of the standard deviation of the observations. For Montana, the RSR value for hydraulic fracturing water use was 0.10, and the RMSE value was 0.373 Mgal per well, or about 20 percent of its respective mean coefficient (table 3).

For North Dakota, the model performance for indirect water use also was good (table 4), with an R^2 of 0.86. The model fit of indirect water use in North Dakota also was satisfactory (fig. 8*C*), as were the differences between water-use estimates and observations (fig. 9*C*). The indirect water-use estimates in Montana necessarily equaled their respective observed water use as shown in figures 8*D* and 9*D*, nullifying computations of R^2, RMSE, or RSR for the indirect water-use category for Montana (table 4) because the coefficients of indirect water use from the North Dakota analysis were applied

Table 4. Goodness-of-fit metrics from leave-one-out cross-validation for linear regression models of direct, indirect, and ancillary water use per oil and gas well developed in the Williston Basin in North Dakota and Montana.

[R^2, coefficient of determination; RMSE, root mean square error; Mgal, million gallons; RSR, ratio of root mean square error to standard deviation of observations; COG, continuous oil and gas; --, none]

Water use	R^2	RMSE (Mgal per well)	RSR
North Dakota			
Direct and indirect water use			
Direct—All	0.91	1.69	0.27
Direct—Hydraulic fracturing	0.91	1.62	0.26
Direct—Cementing	0.78	0.008	0.40
Direct—Drilling	0.78	0.084	0.40
Indirect	0.86	0.275	0.38
Ancillary (population-based) water use			
All	0.85	1.23	0.34
Domestic	0.83	0.479	0.38
Public supply	0.86	0.832	0.34
Other potential ancillary water use			
Industrial	0.84	0.259	0.28
Non-COG mining	0.82	0.967	0.43
Thermoelectric power	0.89	0.173	0.23
Montana			
Direct and indirect water use			
Direct—All	0.87	0.485	0.12
Direct—Hydraulic fracturing	0.92	0.373	0.10
Direct—Cementing	--	--	--
Direct—Drilling	--	--	--
Indirect	--	--	--
Ancillary (population-based) water use			
All	0.67	1.40	0.23
Domestic	0.57	0.186	0.22
Public supply	0.68	1.23	0.23
Other potential ancillary water use			
Industrial	0.79	0.159	0.25
Non-COG mining	0.60	6.31	0.23
Thermoelectric power	0.52	34.4	0.25

to the water-use analysis for Montana. The linear regression model of indirect water use in Montana could not be compared to the linear regression model for North Dakota. The model fit of indirect water use in North Dakota had an RMSE of 0.275 Mgal per well (table 4), or about 60 percent of its respective mean coefficient (table 3), and the RSR value was 0.38, which means RMSE less than 40 percent of the standard deviation of the observations.

The model performance for ancillary water use was more variable for Montana than for North Dakota (table 4). Values of R^2 for the ancillary water-use categories ranged from 0.82 to 0.89 for North Dakota and 0.52 to 0.79 for Montana. For most categories, values of R^2 were about 20 to 25 percent higher for North Dakota than for Montana. For example, the model performance for domestic water use had an R^2 of 0.83 for North Dakota but 0.57 for Montana. Furthermore, values of RMSE for the ancillary water-use categories were larger for Montana, except for domestic and industrial water use (table 4). However, values of RSR were smaller for Montana than for North Dakota for the ancillary water-use categories, except for thermoelectric power water use. For Montana, the high values of RMSE relative to the low values of RSR

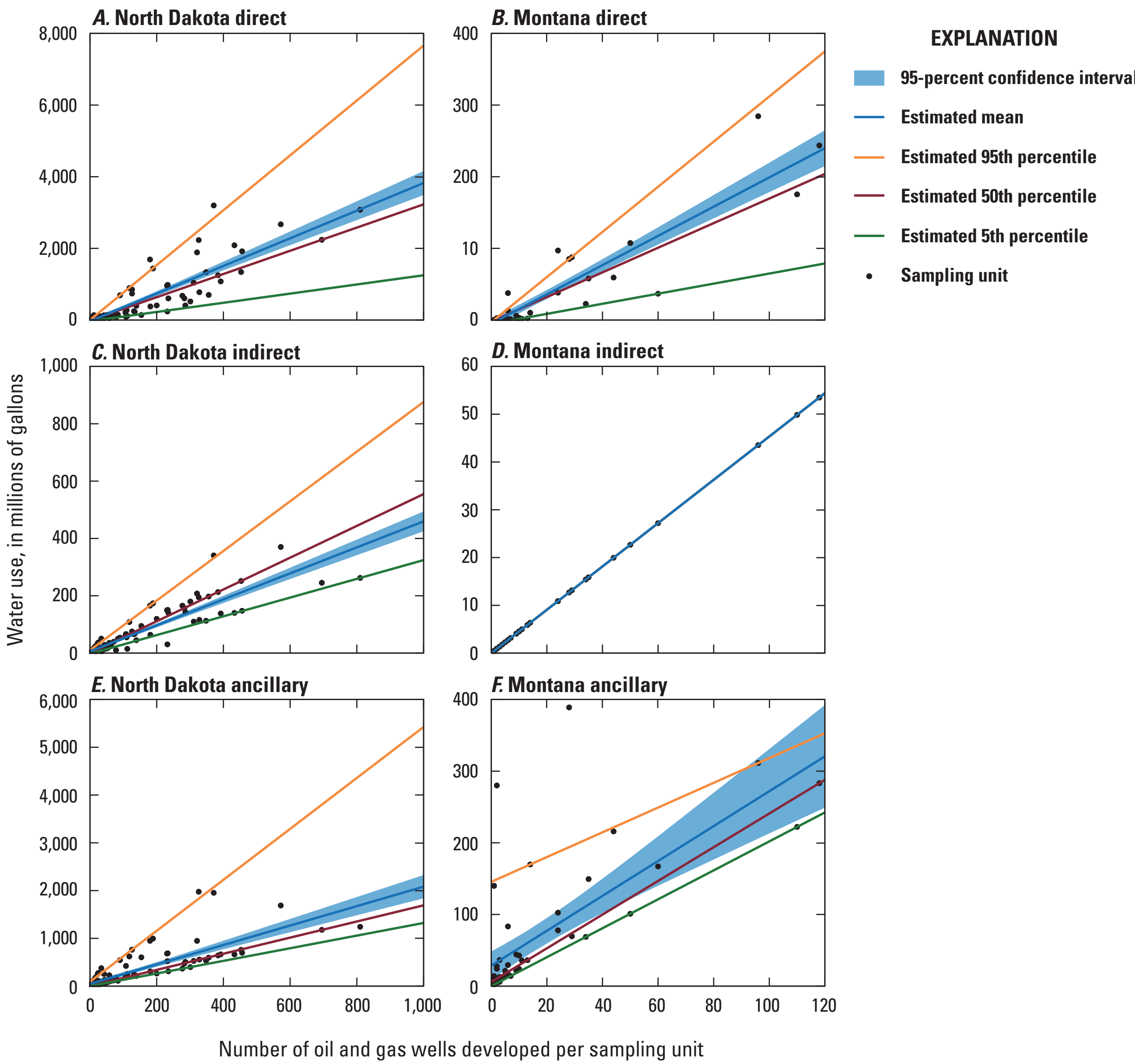

Figure 8. Linear and quantile regression models of direct, indirect, and ancillary water use as explained by number of oil and gas wells developed per sampling unit for North Dakota and Montana. *A*, North Dakota direct. *B*, Montana direct. *C*, North Dakota indirect. *D*, Montana indirect. *E*, North Dakota ancillary. *F*, Montana ancillary.

suggests greater variability in water-use observations, which may reflect the difference in data sources, differences in sample sizes, or a difference in volumes of ancillary water use not accounted for by the number of oil and gas wells developed for Montana as for North Dakota. Furthermore, the model fit of ancillary water use (for domestic and public supply) was qualitatively better for North Dakota (fig. 8*E*) than for Montana (fig. 8*F*). However, the differences between observed and estimated water use (figs. 9*E* and 9*F*) were nearly equivalent between the States.

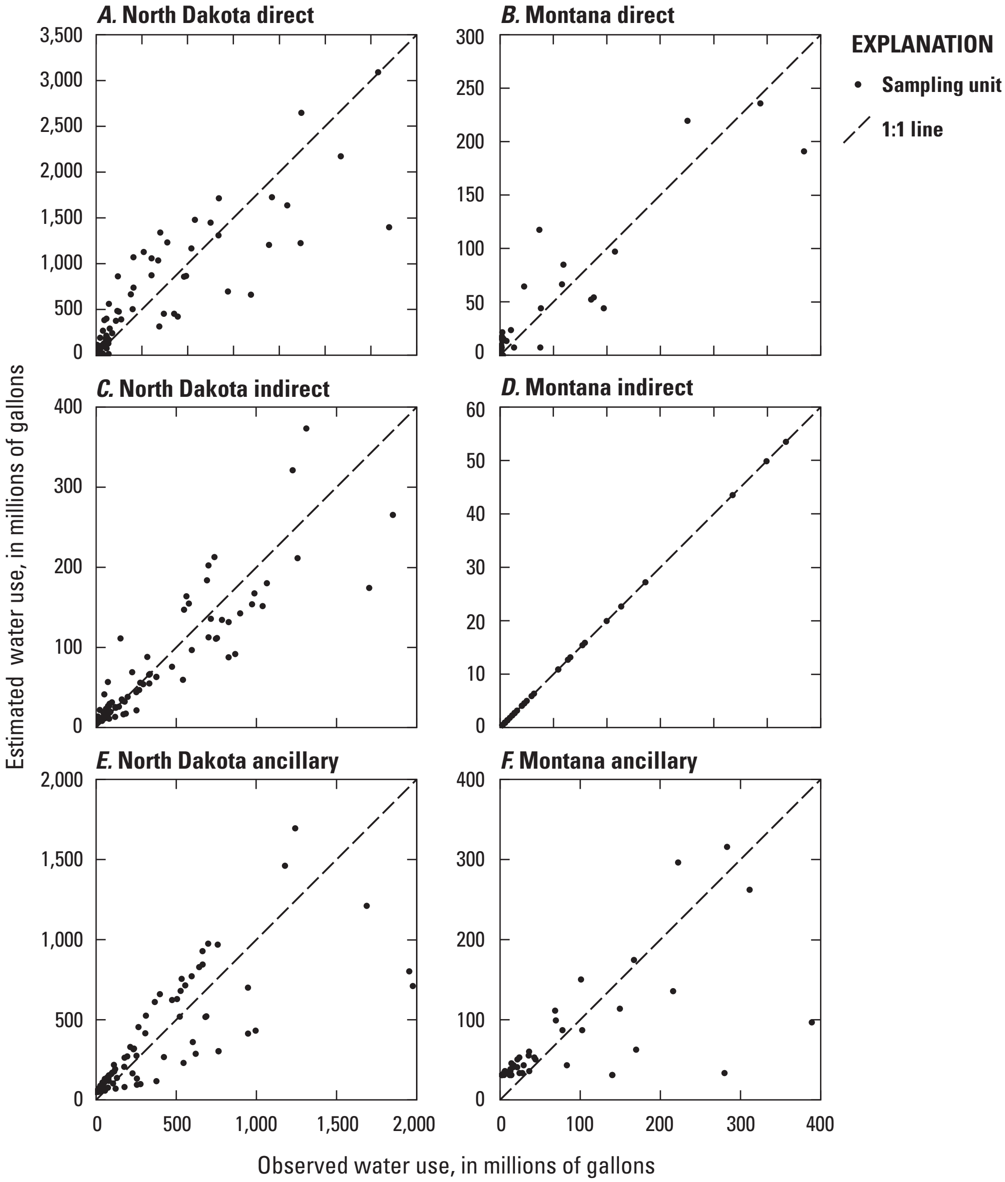

Figure 9. Relation between observed and estimated direct, indirect, and ancillary water use per sampling unit as mean value from linear regression models for North Dakota and Montana. *A*, North Dakota direct. *B*, Montana direct. *C*, North Dakota indirect. *D*, Montana indirect. *E*, North Dakota ancillary. *F*, Montana ancillary.

Comparisons to Water-Use Estimates from Other Studies

Estimates from the water-use analysis in this study (table 3) were compared to water-use estimates from other studies (table 5) of per-well water use for direct water use (for drilling, cementing, and hydraulic fracturing) and ancillary water use (for domestic and public supply). Estimates of indirect water use related to COG development were not found in other studies. The similarity between water-use estimates in this study and from other studies provide some validation of the methods used to analyze water use in this study.

Drilling and cementing water use was estimated in this study at 0.148–0.179 and 0.156 Mgal per well for North Dakota and Montana, respectively (table 3), and estimates of these water-use categories from other studies ranged from 0.150 to 0.236 Mgal per well (Haines and others, 2017; Scanlon and others 2014; table 5). The similarity between estimates of drilling and cementing water use in

Table 5. Estimates of direct, indirect, and ancillary water use associated with continuous oil and gas development in the Williston Basin from other studies.

[--, none]

Water use	Volume per well (million gallons)	Source	Comments
Direct water use			
Drilling, cementing	0.236	Scanlon and others, 2014, table S14a	Drilling use estimated as 6 times borehole volume.
	0.150	Haines and others, 2017	Mean value of range estimated for possible future use.
Drilling, maintenance	0.461	Lin and others, 2018, table S3	Included maintenance water and possibly other uses.
Maintenance	570–850[a]	Horner and others, 2016, section 3.2	--
	465–685[a]	Lin and others, 2018, table S3	--
Hydraulic fracturing	1.82	Gallegos and others, 2015, table S1	Based on data from 2011 to 2014.
	1.97	Kondash and Vengosh, 2015, table 1	Median value.
	3.25	Chen and Carter, 2016, table S5	Mean value in 2014 for Montana and North Dakota.
	2.14	Horner and others, 2016, table 2, section 2.1	Median value of data available from FracFocus at time of study for North Dakota.
	3.80	Scanlon and others, 2016, section 4.4.1	Mean value of potential future range.
	4.17	Haines and others, 2017	Mean value of range estimated for possible future use.
	5.58	Kondash and others, 2018, table S2	Median value in 2016.
	3.92	Lin and others, 2018, table S2	Mean value in 2014.
Indirect water use			
--	--	--	--
Ancillary water use			
Domestic	1.22–2.44	Horner and others, 2016	Use by service personnel in 2012 for North Dakota.

[a]Units in gallons per day.

this study and from other studies may reflect the limited data sources available for estimating drilling and cementing water use.

Estimates of hydraulic fracturing water use in this study were 1.12–7.49 and 0.547–2.99 Mgal per well for North Dakota and Montana, respectively (table 3), which are comparable to estimates from other studies that ranged from 1.82 to 5.58 Mgal per well (Gallegos and others, 2015; Kondash and others, 2018; table 5). The published range of values is less than the range in this study. The lowest published estimate was based on data from 2011 to 2014, the early years of COG development in the Williston Basin when hydraulic fracturing treatments used less water. Similarly, the highest published estimate was based on data from 2016, and water use for hydraulic fracturing treatments have increased with time. In this study, the use of linear and quantile regression models provided a range of parameter estimates (including the mean, the median, and the 5th and 95th percentiles) of per-well hydraulic fracturing water use based on data from 2007 to 2017.

This study estimated ancillary water use (for domestic and public supply) in North Dakota and Montana at 1.19–5.24 and 1.77–3.13 Mgal per well, respectively (table 3), which is much greater than the range (1.22–2.44 Mgal per well) reported for domestic water use in North Dakota in 2012 (Horner and others, 2016; table 5). Horner and others (2016) estimated domestic water use using population estimates near Williston, N. Dak., and a range of per-capita water-use coefficients. The estimates from Horner and others (2016) were based on a ratio of temporary field personnel to permanent workers, which changes with time and could lead to inconsistent estimates. The estimates in this study were based on the reported water use from the NDSWC water permits for North Dakota and the USGS water-use data for Montana using a linear regression model that reduced climate and population effects in the modeled data.

Limitations of Water-Use Analysis for the Williston Basin

The water-use analysis for the Williston Basin had several limitations that potentially affected how accurately water use associated with COG development in the Williston Basin could be estimated. These limitations mostly concern the availability of water-use data and constraints of the modeling approach. The analytical framework from Valder and others (2019), derived from a conceptual model (Valder and others 2018), was the basis for this study, and these publications provided additional context on some limitations that will be discussed in this section.

The coefficients from the regression models of cementing and drilling water use were less certain than those for hydraulic fracturing because several assumptions were necessary because North Dakota and Montana do not require reporting water use for drilling and cementing. Instead, for North Dakota, data on the number of sacks of cement used per well and the measured depth reached per well were used to estimate water use for cementing and drilling. Cementing was assumed to use 5 gallons of water per sack of cement, but it is likely that this value would vary among wells and within a well from the upper to lower depths of cementing the casings because of changes in pressure, temperature, and moisture and because of variability in the geologic formation, including layers containing aquifers. Additionally, drilling was assumed to use 7.5 gallons of water per 1 foot of depth, but this value also would likely vary among wells, depending on geology of the formations, operations of the oil and gas companies, and prices of the water and other materials required for drilling. Moreover, drilling through the surface geologic layers that contain aquifers (fig. 3) uses only water, but the remaining depths can be drilled using other fluids in addition to water. No data were available on the use of water compared to other fluids, so it was assumed that drilling fluids were predominantly water-based. However, the effect of these assumptions is probably small because drilling and cementing water use contribute less than 5 percent to the direct water-use estimate; more than 95 percent of the direct water-use estimate consists of hydraulic fracturing water use.

The regression models of the other potential ancillary water-use categories had an implied assumption that changes in industrial, non-COG mining, and thermoelectric power water use were caused by COG development. However, coefficients from the regression models merely suggest water use for those categories that might be correlated to the presence of COG development; in other words, if COG development were absent, water use for those categories might have been reduced from what was observed by the estimated value. Therefore, the estimates for the other potential ancillary water-use categories were treated separately from the estimates for the population-based categories (domestic and public supply), which had a more-certain relation to COG development.

The confidence interval for the coefficient from the linear regression model had potential limitations to its capacity to adequately capture the data variance. For all water-use categories for each State (except for public supply and thermoelectric power water use for Montana), the 5th, 95th, or both percentile estimates from the quantile regression model were outside the 95-percent confidence interval from the linear regression model, suggesting that the mean estimate for these water-use categories may be less certain than indicated by the linear regression model. Instead, the 95th-percentile estimate provides a more-conservative estimate of water use, which might be more appropriate to use in management decisions about water available for beneficial uses where regions of COG development are in arid climates or when drought is resulting in water scarcity. Additionally, a mean estimate does not provide information on the spatial and temporal variation in water use, so the 95th-percentile estimate might be more robust to changes in water use, particularly with the knowledge that hydraulic fracturing water use per well has been trending upward nationally (Gallegos and Varela, 2015).

A key assumption affecting the estimates of direct, indirect, and ancillary water use was the estimation of a breakpoint in the data on hydraulic fracturing water use. Breakpoints were estimated at 2007 and 2008 for North Dakota and Montana, respectively; however, other years could have been chosen from the breakpoint analysis, and selecting a different year would have affected the water-use estimates. Different years of data used to model water use after the boom in COG development would affect how much direct water use is estimated because the coefficients of the regression models would likely differ. Likewise, the trend in direct water use in the years preceding the boom in COG development is sensitive to the years of data fit by the regression model. Because this trend is extrapolated forward into the years after the breakpoint, a different breakpoint would affect how much ancillary water use is attributed to COG development and how much of the COG mining water use is contributed by indirect water use.

The evaluation of uncertainty in the water-use estimates and performance of the linear and quantile regression models was limited by the goodness-of-fit metrics and validation procedures used for this study. Although this study used leave-one-out cross-validation, other procedures, including bootstrap, jackknife, or k-fold cross-validation (Hastie and others, 2009), also may have been appropriate. However, it is unlikely the use of other validation procedures would have greatly affected model performance. Goodness-of-fit metrics other than R^2, RMSE, and RSR, which this study used, also can be informative. However, these three metrics are commonly used to evaluate the performance of hydrologic models (Moriasi and others, 2007), and they provided a reasonable assessment of model performance in this study.

The simple linear and quantile regression models used for this study were potentially limited in their ability to explain water use associated with COG development. Linear regression models tend to have low variance but high bias, underfitting the data, whereas more sophisticated nonlinear or multivariate statistical approaches or machine-learning algorithms tend to have lower bias but higher variance, overfitting the data. A more complex model than linear or quantile regression might have explained more variance in water use in the Williston Basin. However, a more complex model needs more explanatory variables likely unavailable for all regions of COG production, so the best-fitting model for the Williston Basin might not be applicable to another region. The regression models used for this study do not require data on explanatory variables potentially unavailable in other regions, so these models are more readily adapted to existing water-use data for estimating water use associated with COG development nationally for the USGS National Water Census and Water Budget Estimation and Evaluation Project (supported by the USGS Water Availability and Use Science Program). Additionally, the regionally based coefficients from the regression models (including the mean, the median, and the 5th and 95th percentiles) deliver a probable range of water use.

Summary

Water is a necessary component for many processes required for developing continuous oil and gas (COG) resources. Improved COG extraction techniques have greatly increased oil and gas production in the United States since the mid-2000s. However, the accompanying rapid increase in demand for large volumes of water, often in remote regions, can challenge existing infrastructure and require additional resources to meet water needs. Addressing this water need requires accurate estimates of the volumes of water used to support the various processes common to COG development in the United States in the 21st century.

This study of water use associated with COG development in the Williston Basin is intended to provide a preliminary model-based analysis of water use in major regions of COG production in the United States. Direct, indirect, and ancillary water use associated with COG development in the Williston Basin was estimated in North Dakota and Montana from 2007 to 2017. The water-use estimates presented in this report are compared with other published values for the Williston Basin, and the limitations of the water-use analysis for the Williston Basin are described.

The water-use analysis was executed with scripts coded in R. Water-use data were aggregated by county and year, which were the sampling units used in the analysis. Linear and quantile regression models of water use in relation to the number of oil and gas wells developed were fit for the direct, indirect, and ancillary water-use categories for each State. A 95-percent confidence interval for each parameter estimate from the linear regression models was computed as a measure of uncertainty. Additional information on uncertainty can be gained from modeling other distribution parameters, so quantile regression models of the 5th, 50th, and 95th percentiles also were fit. Water-use coefficients from the linear and quantile regression models are in million gallons per well, whereas water-use estimates from the regression models are in million gallons per sampling unit. To assess uncertainty in the estimates from the regression models of direct, indirect, and ancillary water use, leave-one-out cross-validation was used. Model performance was evaluated with three goodness-of-fit metrics used to compare the estimates and observations of water use.

Mean annual direct and indirect water use for COG development in North Dakota was estimated at 4,512 million gallons (Mgal) per year (Mgal/yr), with a 95-percent confidence interval of 4,021–5,152 Mgal/yr, and in Montana was estimated at 196 Mgal/yr, with a 95-percent confidence interval of 189–203 Mgal/yr. Ancillary water use (for domestic and public supply) had an estimated annual mean of 2,753 Mgal/yr in North Dakota and 396 Mgal/yr in Montana.

The coefficient from the linear regression model of direct water use was 3.86 Mgal per well and hydraulic fracturing water use was 3.70 Mgal per well for North Dakota. The mean estimate of direct water use had a 95-percent confidence interval of 3.48–4.23 Mgal per well. For North Dakota, the coefficient from the linear regression model of indirect water use was 0.453 Mgal per well, with a 95-percent confidence interval of 0.415–0.492 Mgal per well. Direct and indirect water use had a mean estimate of about 4.31 Mgal per well in North Dakota. The mean estimate of ancillary water use (for domestic and public supply) in North Dakota was 2.03 Mgal per well, with a 95-percent confidence interval of 1.76–2.31 Mgal per well. For Montana, the linear regression model of hydraulic fracturing water use had a mean estimate of 2.04 Mgal per well. The 95-percent confidence interval for the mean estimate was 1.80–2.28 Mgal per well. Direct and indirect water use in Montana had a mean estimate of 2.49 Mgal per well. The mean estimate of ancillary water use (for domestic and public supply) in Montana was 2.43 Mgal per well, with a 95-percent confidence interval of 1.76–3.11 Mgal per well.

The model performance for direct water use in both States was good, with a coefficient of determination (R^2) of 0.91 for North Dakota and 0.87 for Montana. For North Dakota, the model performance for indirect water use also was good, with an R^2 of 0.86. The model performance for ancillary water use was more variable for Montana than for North Dakota. Values of R^2 for the ancillary water-use categories ranged from 0.82 to 0.89 for North Dakota and 0.52 to 0.79 for Montana. For most categories, values of R^2 were about 20 to 25 percent higher for North Dakota than for Montana.

The similarity between water-use estimates in this study and from other studies provide some validation of the methods used to analyze water use in this study. Drilling and cementing water use was estimated in this study at 0.148–0.179 and 0.156 Mgal per well for North Dakota and Montana, respectively, and estimates of these water-use categories from other studies ranged from 0.150 to 0.236 Mgal per well. Estimates of hydraulic fracturing water use in this study were 1.12–7.49 and 0.547–2.99 Mgal per well for North Dakota and Montana, respectively, which are comparable to estimates from other studies that ranged from 1.82 to 5.58 Mgal per well. This study estimated ancillary water use (for domestic and public supply) in North Dakota and Montana at 1.19–5.24 and 1.77–3.13 Mgal per well, respectively, which is much greater than the range (1.22–2.44 Mgal per well) reported for domestic water use in North Dakota in 2012.

The water-use analysis for the Williston Basin had several limitations that potentially affected how accurately water use associated with COG development in the Williston Basin could be estimated. These limitations mostly concern the availability of water-use data and constraints of the modeling approach. A more complex model than linear or quantile regression might have explained more variance in water use in the Williston Basin. However, a more complex model needs more explanatory variables likely unavailable for all regions of COG production, so the best-fitting model for the Williston Basin might not be applicable to another region. The regression models used for this study do not require data on explanatory variables potentially unavailable in other regions, so these models are more readily adapted to existing water-use data for estimating water use associated with COG development nationally.

References Cited

American Oil and Gas Historical Society, 2013, First North Dakota oil well: American Oil and Gas Historical Society, accessed February 10, 2017, at https://aoghs.org/states/north-dakota-williston-basin.

Anderson, S.B., and Eastwood, W.P., 1968, Natural gas in North Dakota, *in* Natural gases in rocks of Paleozoic age: American Association of Petroleum Geologists Memoir no. 9., part 3, p. 1304–1326.

Anna, L.O., Pollastro, R., and Gaswirth, S.B., 2011, Williston Basin Province—Stratigraphic and structural framework to a geologic assessment of undiscovered oil and gas resources, chap. 2 *of* U.S. Geological Survey Williston Basin Province Assessment Team, Assessment of undiscovered oil and gas resources of the Williston Basin Province of North Dakota, Montana, and South Dakota, 2010: U.S. Geological Survey Digital Data Series 69–W, 17 p. [Also available at https://doi.org/10.3133/ds69W.]

Baker Hughes, 2017, North American rig count: Baker Hughes, digital data, accessed February 17, 2017, at https://bakerhughesrigcount.gcs-web.com/na-rig-count.

Bartram, J.G., and Erdmann, C.E., 1935, Natural gas in Montana, *in* Geology of natural gas: American Association of Petroleum Geology special volume 7, p. 245–254.

Bluemle, J.P., 2001, The 50th anniversary of the discovery of oil in North Dakota: North Dakota Geological Survey, Miscellaneous Series no. 89, 57 p., accessed February 3, 2017, at https://www.dmr.nd.gov/ndgs/documents/Publication_List/pdf/MiscSeries/MS-89.pdf.

Carter, J.M., Macek-Rowland, K.M., Thamke, J.N., and Delzer, G.C., 2016, Estimating national water use associated with unconventional oil and gas development: U.S. Geological Survey Fact Sheet 2016–3032, 6 p. [Also available at https://doi.org/10.3133/fs20163032.]

Chen, H., and Carter, K.E., 2016, Water usage for natural gas production through hydraulic fracturing in the United States from 2008 to 2014: Journal of Environmental Management, v. 170, p. 152–159. [Also available at https://doi.org/10.1016/j.jenvman.2016.01.023.]

Dieter, C.A., Maupin, M.A., Caldwell, R.R., Harris, M.A., Ivahnenko, T.I., Lovelace, J.K., Barber, N.L., and Linsey, K.S., 2018, Estimated use of water in the United States in 2015: U.S. Geological Survey Circular 1441, 65 p. [Also available at https://doi.org/10.3133/cir1441.]

Dutton, D.M., Varela, B., Haines, S.S., Barnhart, T.B., McShane, R.R., and Wheeling, S.L., 2019, Water-use data to estimate water use associated with continuous oil and gas development, Williston Basin, United States, 1980–2017: U.S. Geological Survey data release, accessed October 2019 at https://doi.org/10.5066/P9CPKRLW.

Fischer, D.W., and Bluemle, J.P., 1986, Oil exploration and development in the North Dakota Williston Basin—1984–1985 update: North Dakota Geological Survey Miscellaneous Series 67, 40 p.

Gallegos, T.J., and Varela, B.A., 2015, Trends in hydraulic fracturing distributions and treatment fluids, additives, proppants, and water volumes applied to wells drilled in the United States from 1947 through 2010—Data analysis and comparison to the literature: U.S. Geological Survey Scientific Investigations Report 2014–5131, 15 p. [Also available at https://doi.org/10.3133/sir20145131.]

Gallegos, T.J., Varela, B.A., Haines, S.S., and Engle, M.A., 2015, Hydraulic fracturing water use variability in the United States and potential environmental implications: Water Resources Research, v. 51, no. 7, p. 5839–5845. [Also available at https://doi.org/10.1002/2015WR017278.]

Gaswirth, S.B., Marra, K.R., Cook, T.A., Charpentier, R.R., Gautier, D.L., Higley, D.K., Klett, T.R., Lewan, M.D., Lillis, P.G., Schenk, C.J., Tennyson, M.E., and Whidden, K.J., 2013, Assessment of undiscovered oil resources in the Bakken and Three Forks Formations, Williston Basin Province, Montana, North Dakota, and South Dakota, 2013: U.S. Geological Survey Fact Sheet 2013–3013, 4 p. [Also available at https://doi.org/10.3133/fs20133013.]

Gleason, R.A., and Tangen, B.A., eds., 2014, v. 2014–5017. Brine contamination to aquatic resources from oil and gas development in the Williston Basin, United States, U.S. Geological Survey Scientific Investigations Report, 127 p. [Also available at https://doi.org/10.3133/sir20145017.]

Haines, S.S., Varela, B.A., Hawkins, S.J., Gianoutsos, N.J., Thamke, J.N., Engle, M.A., Tennyson, M.E., Schenk, C.J., Gaswirth, S.B., Marra, K.R., Kinney, S.A., Mercier, T.J., and Martinez, C.D., 2017, Assessment of water and proppant quantities associated with petroleum production from the Bakken and Three Forks Formations, Williston Basin Province, Montana and North Dakota, 2016: U.S. Geological Survey Fact Sheet 2017–3044, 4 p. [Also available at https://doi.org/10.3133/fs20173044.]

Hastie, T., Tibshirani, R., and Friedman, J., 2009, The elements of statistical learning—Data mining, inference, and prediction 2nd ed.: New York, Springer, 745 p.

Helms, L., 2019, Director's cut—July 2019 production: North Dakota Industrial Commission, accessed September 20, 2019, at https://www.dmr.nd.gov/oilgas/directorscut/directorscut-2019-09-16.pdf.

Horner, R.M., Harto, C.B., Jackson, R.B., Lowry, E.R., Brandt, A.R., Yeskoo, T.W., Murphy, D.J., and Clark, C.E., 2016, Water use and management in the Bakken shale oil play in North Dakota: Environmental Science & Technology, v. 50, no. 6, p. 3275–3282. [Also available at https://doi.org/10.1021/acs.est.5b04079.]

IHS Markit, 2018, U.S. well database: Englewood, Colo., IHS Markit, accessed April 27, 2018, at https://ihsmarkit.com/industry/oil-gas.html.

Kenny, J.F., Barber, N.L., Hutson, S.S., Linsey, K.S., Lovelace, J.K., and Maupin, M.A., 2009, Estimated use of water in the United States in 2005: U.S. Geological Survey Circular 1344, 52 p. [Also available at https://pubs.usgs.gov/circ/1344.]

Koenker, R., 2005, Quantile regression: Cambridge, United Kingdom, Cambridge University Press, 349 p. [Also available at https://doi.org/10.1017/CBO9780511754098.]

Koenker, R., 2018, quantreg—Quantile regression: R package, ver. 5.38, accessed July 2018 at https://cran.r-project.org/web/packages/quantreg/.

Kondash, A., and Vengosh, A., 2015, Water footprint of hydraulic fracturing: Environmental Science & Technology Letters, v. 2, no. 10, p. 276–280. [Also available at https://doi.org/10.1021/acs.estlett.5b00211.]

Kondash, A.J., Lauer, N.E., and Vengosh, A., 2018, The intensification of the water footprint of hydraulic fracturing: Science Advances, v. 4, no. 8, eaar5982. [Also available at https://doi.org/10.1126/sciadv.aar5982.]

Legates, D.R., and McCabe, G.J., Jr., 1999, Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation: Water Resources Research, v. 35, no. 1, p. 233–241. [Also available at https://doi.org/10.1029/1998WR900018.]

Lin, Z., Lin, T., Lim, S.H., Hove, M.H., and Schuh, W.M., 2018, Impacts of Bakken shale oil development on regional water uses and supply: Journal of the American Water Resources Association, v. 54, no. 1, p. 225–239. [Also available at https://doi.org/10.1111/1752-1688.12605.]

Long, A.J., Aurand, K.R., Bednar, J.M., Davis, K.W., Mckaskey, J.D.R.G., and Thamke, J.N., 2014, Conceptual model of the uppermost principal aquifer systems in the Williston and Powder River structural basins, United States and Canada: U.S. Geological Survey Scientific Investigations Report 2014–5055, 41 p. [Also available at https://doi.org/10.3133/sir20145055.]

Long, A.J., Thamke, J.N., Davis, K.W., and Bartos, T.T., 2018, Groundwater availability of the Williston Basin, United States and Canada: U.S. Geological Survey Professional Paper 1841, 62 p. [Also available at https://doi.org/10.3133/pp1841.]

Maupin, M.A., Kenny, J.F., Hutson, S.S., Lovelace, J.K., Barber, N.L., and Linsey, K.S., 2014, Estimated use of water in the United States in 2010: U.S. Geological Survey Circular 1405, 56 p. [Also available at https://doi.org/10.3133/cir1405.]

Montana Board of Oil and Gas Conservation, 2019, Montana well production data: Montana Department of Natural Resources and Conservation, accessed September 13, 2019, at http://www.bogc.dnrc.mt.gov/WebApps/DataMiner/Default.aspx.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., and Veith, T.L., 2007, Model evaluation guidelines for systematic quantification of accuracy in watershed simulations: Transactions of the ASABE, v. 50, no. 3, p. 885–900. [Also available at https://doi.org/10.13031/2013.23153.]

Muggeo, V.M.R., 2018, segmented—Regression models with break-points/change-points estimation: R package, ver. 0.5–3.0, accessed July 2018 at https://cran.r-project.org/web/packages/segmented/.

North Dakota Industrial Commission, 2018, Information for oil and gas wells permitted in North Dakota: North Dakota Industrial Commission, accessed April 13, 2018, at https://www.dmr.nd.gov/oilgas.

North Dakota Oil and Gas Division, 2019, North Dakota general statistics on oil and gas production: North Dakota Industrial Commission, accessed March 14, 2016, at https://www.dmr.nd.gov/oilgas/stats/statisticsvw.asp.

North Dakota State Water Commission, 2018, Information for water permits and temporary water permits issued in North Dakota: North Dakota State Water Commission, accessed April 6, 2018, at https://www.swc.nd.gov/info_edu/map_data_resources.

PRISM Climate Group, 2018, PRISM climate data: Oregon State University, accessed July 6, 2018, at http://prism.oregonstate.edu.

R Core Team, 2019, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, ver. 3.5.1., accessed July 2018 at https://www.R-project.org.

Scanlon, B.R., Reedy, R.C., and Nicot, J.P., 2014, Comparison of water use for hydraulic fracturing for unconventional oil and gas versus conventional oil: Environmental Science & Technology, v. 48, no. 20, p. 12386–12393. [Also available at https://doi.org/10.1021/es502506v.]

Scanlon, B.R., Reedy, R.C., Male, F., and Hove, M., 2016, Managing the increasing water footprint of hydraulic fracturing in the Bakken Play, United States: Environmental Science & Technology, v. 50, no. 18, p. 10273–10281. [Also available at https://doi.org/10.1021/acs.est.6b01375.]

Schmoker, J.W., 2005, U.S. Geological Survey assessment concepts for continuous petroleum accumulations, chap. 13 *of* USGS Southwestern Wyoming Province Assessment Team, Petroleum systems and geologic assessment of oil and gas in the Southwestern Wyoming Province, Wyoming, Colorado, and Utah: U.S. Geological Survey Digital Data Series DDS–69–D, 7 p. [Also available at https://pubs.usgs.gov/dds/dds-069/dds-069-d/REPORTS/69_D_CH_13.pdf.]

Schmoker, J.W., and Klett, T.R., 2005, U.S. Geological Survey assessment concepts for conventional petroleum accumulations, chap. 19 *of* USGS Southwestern Wyoming Province Assessment Team, Petroleum systems and geologic assessment of oil and gas in the Southwestern Wyoming Province, Wyoming, Colorado, and Utah: U.S. Geological Survey Digital Data Series DDS–69–D, 6 p. [Also available at https://pubs.usgs.gov/dds/dds-069/dds-069-d/REPORTS/69_D_CH_19.pdf.]

Thamke, J.N., LeCain, G.D., Ryter, D.W., Sando, R., and Long, A.J., 2014, Hydrogeologic framework of the upper principal aquifer systems in the Williston and Powder River structural basins, United States and Canada: U.S. Geological Survey Scientific Investigations Report 2014–5047, 38 p. [Also available at https://doi.org/10.3133/sir20145047.]

U.S. Census Bureau, 2018, Intercensal state and county population estimates: U.S. Census Bureau, accessed April 6, 2018, at https://www2.census.gov/programs-surveys/popest/tables.

U.S. Energy Information Administration, 2013, North Dakota oil production reaches new high in 2012, transported by trucks and railroads: U.S. Energy Information Administration, accessed February 24, 2017, at https://www.eia.gov/todayinenergy/detail.php?id=10411.

U.S. Energy Information Administration, 2014, North Dakota and Texas now provide nearly half of U.S. crude oil production: U.S. Energy Information Administration, accessed February 24, 2017, at https://www.eia.gov/todayinenergy/detail.php?id=16931.

U.S. Energy Information Administration, 2016, Hydraulic fracturing accounts for about half of current U.S. crude oil production: U.S. Energy Information Administration, accessed September 20, 2019, at https://www.eia.gov/todayinenergy/detail.php?id=25372.

U.S. Geological Survey, 2019, Water-use data: U.S. Geological Survey, digital data, accessed August 30, 2019, at https://water.usgs.gov/watuse/data.

Valder, J.F., McShane, R.R., Barnhart, T.B., Sando, R., Carter, J.M., and Lundgren, R.F., 2018, Conceptual model to assess water use associated with the life cycle of unconventional oil and gas development: U.S. Geological Survey Scientific Investigations Report 2018–5027, 22 p. [Also available at https://doi.org/10.3133/sir20185027.]

Valder, J.F., McShane, R.R., Barnhart, T.B., Wheeling, S.L., Carter, J.M., Macek-Rowland, K.M., Delzer, G.C., and Thamke, J.N., 2019, Analytical framework to estimate water use associated with continuous oil and gas development: U.S. Geological Survey Scientific Investigations Report 2019–5100, 19 p. [Also available at https://doi.org/10.3133/sir20195100.]

Appendix 1. R Scripts

The R scripts for the Williston Basin water-use analysis are available at https://doi.org/10.3133/sir20205012 in a zipped archive, sir20205012_appendix1.zip, which contains the following files and folders:

- README.txt—A "how-to" document for running the scripts;
- munge_data_release.R—A script used for preparing data from the accompanying data release;
- run_analysis.R—The script used for running the water-use analysis;
- wrangle.R—A script with functions for preparing input data to the water-use analysis;
- model.R—A script with functions for processing the water-use analysis;
- visualize.R—A script with functions for preparing model output from the water-use analysis as tables and figures;
- Raw—A folder for downloading data from the prior data release;
- Functions—A folder with the scripts (wrangle.R, model.R, and visualize.R) used for processing the water-use analysis;
- Data—A folder for saving prepared input data;
- Analyses—A folder for saving model output prepared as tables; and
- Plots—A folder for saving model output prepared as figures.

Appendix 2. Water-Use Estimates and Coefficients

The Williston Basin water-use estimates and coefficients are available at https://doi.org/10.3133/sir20205012 in a zipped folder, sir20205012_appendix2.zip, with the following delimited files:

- WaterUseEstimates.csv—A comma-separated values file of estimates from the regression models used in the water-use analysis; and
- WaterUseCoefficients.csv—A comma-separated values file of coefficients from the regression models used in the water-use analysis.

The WaterUseEstimates.csv file has 10 columns and 635 rows. The first column lists the State and contains "North Dakota" or "Montana." The second column lists the County and contains one of the following: "Billings," "Bottineau," "Bowman," "Burke," "Burleigh," "Carter," "Custer," "Daniels," "Dawson," "Divide," "Dunn," "Emmons," "Fallon," "Garfield," "Golden Valley," "Hettinger," "McCone," "McHenry," "McKenzie," "McLean," "Mercer," "Morton," "Mountrail," "Oliver," "Renville," "Richland," "Roosevelt," "Sheridan," "Slope," "Stark," "Valley," "Ward," "Wibaux," or "Williams." The third column contains the year and ranges from 2007 to 2017. The fourth column lists the water-use category and contains one of the following: "COG mining," "Domestic," "Industrial," "Non-COG mining," "Public supply," or "Thermoelectric power." Descriptions of the water-use categories are in the report. The abbreviation COG means continuous oil and gas. The fifth column contains the mean coefficients from the linear regression models of water use and ranges from 0.028 to 10,701.604. The sixth and seventh columns contain the lower and upper confidence limits of the 95-percent confidence intervals for the mean coefficients and range from 0 to 7,519.259 and from 0.037 to 15,019.646, respectively. The eighth, ninth, and tenth columns contain the 5th, 50th, and 95th percentile coefficients from the quantile regression models of water use and range from 0 to 10,614.283, from 0 to 10,614.283, and from 0.033 to 11,456.377, respectively. Values are in million gallons per year.

The WaterUseCoeffients.csv file has 8 columns and 24 rows. The first column lists the State and contains "North Dakota" or "Montana." The second column lists the water-use category and contains one of the following: "Hydraulic fracturing," "Cementing," "Drilling," "Direct," "Indirect," "COG mining," "Domestic," "Public supply," "Population-based ancillary," "Industrial," "Non-COG mining," or "Thermoelectric power." Descriptions of the water-use categories are in the report. The third column contains the mean coefficients from the linear regression models of water use and ranges from 0.014 to 61.991. The fourth and fifth columns contain the lower and upper confidence limits of the 95-percent confidence intervals for the mean coefficients and range from 0.014 to 48.309 and from 0.014 to 75.673, respectively. The sixth, seventh, and eighth columns contain the 5th, 50th, and 95th percentile coefficients from the quantile regression models of water use and range from 0.014 to 50.544, from 0.014 to 64.918, and from 0.014 to 64.918, respectively. Values are in million gallons per oil and gas well developed.

For more information about this publication, contact:
Director, USGS Wyoming-Montana Water Science Center
3162 Bozeman Avenue
Helena, MT 59601
406–457–5900

For additional information, visit: https://www.usgs.gov/centers/wy-mt-water/

Publishing support provided by the
Rolla Publishing Service Center

McShane and others—**Estimates of Water Use Associated with Continuous Oil and Gas Development in the Williston Basin**—SIR 2020–5012

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20205012