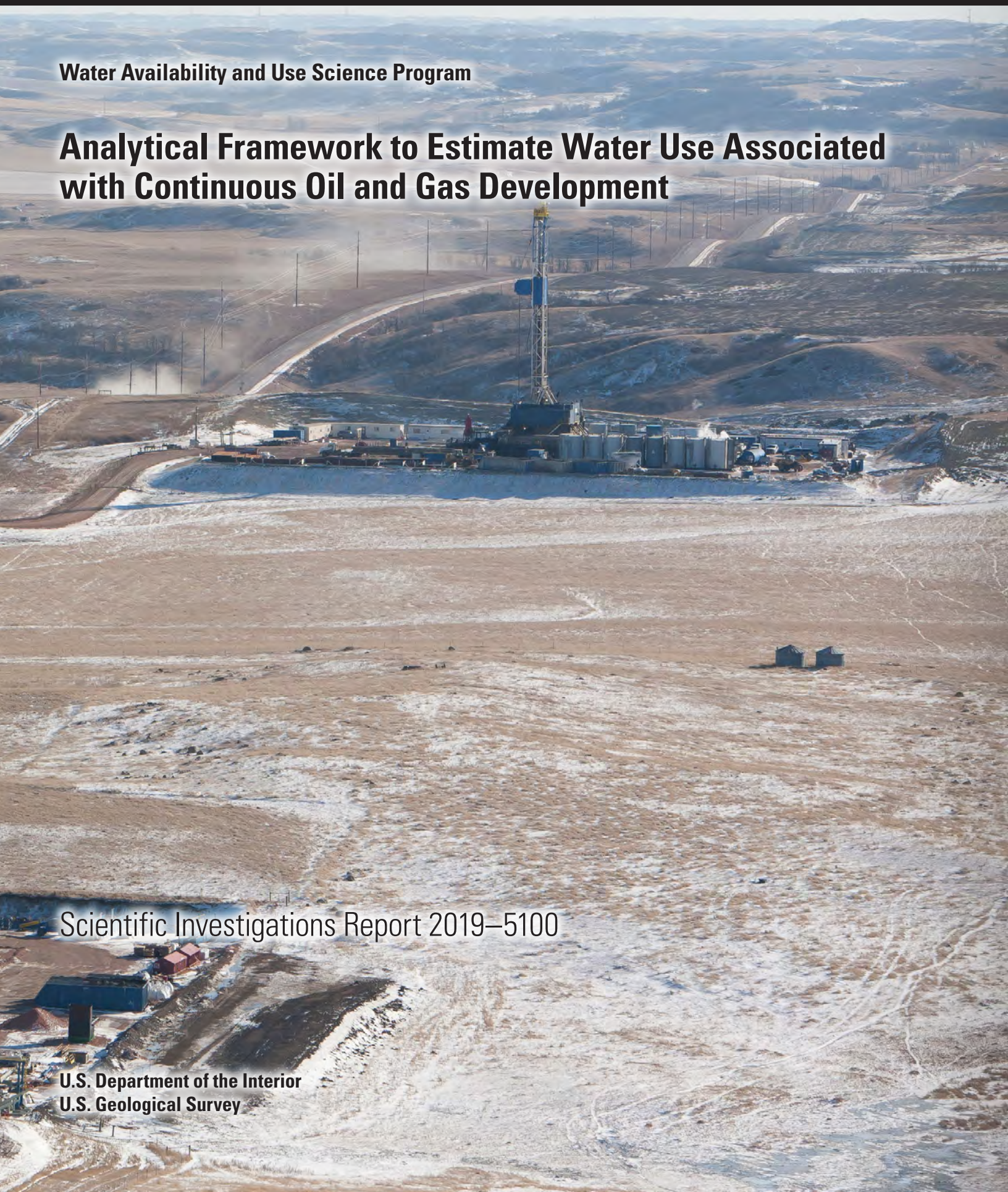
USGS
science for a changing world

Water Availability and Use Science Program

Analytical Framework to Estimate Water Use Associated with Continuous Oil and Gas Development



Scientific Investigations Report 2019–5100

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

Cover. Aerial photograph of oil and gas production wells along a service road in North Dakota. Photograph by Vern Whitten Photography, used with permission.

Analytical Framework to Estimate Water Use Associated with Continuous Oil and Gas Development

By Joshua F. Valder, Ryan R. McShane, Theodore B. Barnhart, Spencer L. Wheeling, Janet M. Carter, Kathleen M. Macek-Rowland, Gregory C. Delzer, and Joanna N. Thamke

Water Availability and Use Science Program

Scientific Investigations Report 2019–5100

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: 2019

Suggested citation:
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., https://doi.org/10.3133/sir20195100.

ISSN 2328-0328 (online)

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Volume	
barrel (bbl; petroleum, 1 barrel=42 gal)	0.1590	cubic meter (m^3)
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)
acre-foot (acre-ft)	1,233	cubic meter (m^3)

Abbreviations

COG continuous oil and gas

USGS U.S. Geological Survey

Acknowledgments

The authors thank the North Dakota State Water Commission, the North Dakota Oil and Gas Commission, the Montana Bureau of Mines and Geology, and State officials from North Dakota and Montana for their support of past and ongoing studies that provided valuable information for this study.

This project was funded and supported by the U.S. Geological Survey (USGS) Water Availability and Use Science Program. Several personnel at the USGS assisted with the collection, processing, and analysis of the water-use data. Jessica Garrett, William Eldridge, and Benjamin York (USGS) provided insightful edits and comments for this report. Roy Sando and Seth Haines (USGS) assisted in providing technical guidance and insights supporting the development of predata and postdata processing used in this approach. The authors also thank Molly Maupin and Mindi Dalton (USGS) as valued members of the Continuous Oil and Gas Topical Water-Use Project Management Team for their guidance.

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

Analytical Framework to Estimate Water Use Associated with Continuous Oil and Gas Development

By Joshua F. Valder, Ryan R. McShane, Theodore B. Barnhart, Spencer L. Wheeling, Janet M. Carter, Kathleen M. Macek-Rowland, Gregory C. Delzer, and Joanna N. Thamke

Abstract

An analytical framework was designed to estimate water use associated with continuous oil and gas (COG) development in support of the U.S. Geological Survey Water Availability and Use Science Program. This framework was developed to better understand the relation between the production of COG resources for energy and the amount of water needed to sustain this type of energy development in the United States. The total mean undiscovered, technically recoverable volume of COG has increased, highlighting the continued need to develop approaches to better characterize water use associated with COG development.

The analytical framework can be used to estimate water use associated with COG development for three water-use components—direct, indirect, and ancillary water use—that are related to the life cycle of COG development. Direct water use is defined as water used in a wellbore to complete a well, including the water used for drilling, cementing, stimulating, and maintaining the well during production. Indirect water use is the water used at or near the well site, including water used for dust abatement, for cleaning equipment, and for crew and staff use. Ancillary water use is 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 analytical framework includes the data inputs, the processes involved in estimating the water-use coefficients and analyzing their uncertainties, and the outputs. The analytical framework was developed as an R script, which contains the statistical models used to estimate water-use components.

The availability of data across COG reservoirs in the United States is variable and presents challenges for estimating water use for extracting COG from their reservoirs; thus, the R script can be modified for the types of data available within a COG reservoir, the extent and resolution of data available for each water-use component, and the desired output of the water-use assessment. The script was written so that the units of the data in the script were standardized. Water-use estimates were simulated for the mean and 10th, 50th, and 90th percentiles of the data distributions. Uncertainties were quantified with confidence intervals for the estimated coefficients. Uncertainty for estimated or simulated data can be calculated with the R script by providing a range of representative values that are within the appropriate confidence intervals of the mean of the data.

Introduction

Understanding the relation between the production of energy and the water used to produce that energy is a necessary component of any successful long-term (decades) energy strategy within the United States. This relation applies to the entire life cycle of renewable and nonrenewable forms of energy, which includes extraction, production, refinement, delivery, and disposal of waste byproducts. Nonrenewable energy, specifically oil and gas, generally requires large volumes of water for extraction. These nonrenewable forms of energy, such as crude oil, natural gas, and coal, are the primary forms of energy used within the United States. In 2017, crude oil, natural gas, and coal accounted for 36.2, 28.0, and 13.9 percent, respectively, of total energy consumption within the United States (U.S. Energy Information Administration, 2018). According to the U.S. Geological Survey (USGS), fossil fuels consisting of crude oil and natural gas are categorized as conventional (generally vertical drilling) or continuous (generally horizontal drilling) based primarily on their disposition within the geologic strata (U.S. Geological Survey Energy Resources Program, written commun., 2015). Conventional oil and gas accumulations are well-defined hydrocarbon-water contacts and commonly have high matrix permeabilities, which typically will have geologic structural traps with high degrees of recovery (U.S. Geological Survey Energy Resources Program, written commun., 2015). Because of the ease of extraction, conventional oil and gas deposits historically have been the most cost-effective resources to develop (Valder and others, 2018).

As access to conventional oil and gas fields gets scarcer, oil and gas prices increase, and technological advances such as horizontal drilling with hydraulic fracturing evolve; the result

is the advancement of continuous oil and gas (COG) extraction techniques (Valder and others, 2018). The COG deposits are described by the USGS Energy Resources Program (written commun., 2015) as "an oil resource that is dispersed continuously throughout a geologic formation rather than existing as discrete, localized occurrences, such as those in conventional accumulations." The USGS Energy Resources Program leads scientific investigations to quantitatively assess the potential for undiscovered, technically recoverable conventional and COG resources in priority geologic provinces in the United States and around the world (U.S. Geological Survey National Assessment of Oil and Gas Resources Team and Biewick, 2014). The last comprehensive national assessment of U.S. oil and gas resources (1995) used oil and gas reservoirs as the basic level of assessment (Schmoker, 2005). Reservoirs, as defined herein, are established primarily according to similarities of the rocks in which petroleum occurs.

Continuous resources often require special technical drilling and recovery methods. The COG resources are developed using a method that combines directional drilling and hydraulic fracturing techniques. These techniques allow for a larger extraction of oil and gas deposits that previously would have been unrecoverable using conventional drilling techniques (Valder and others, 2018). Although these technological advances have provided the ability to access and extract additional oil and gas resources, they require large volumes of water (Jiang and others, 2014). As such, the USGS began a topical study through the Water Availability and Use Science Program to better understand the amount of water needed for ongoing production of COG resources for energy development in the United States (Valder and others, 2018).

The USGS, as part of a national water-use compilation effort, has compiled water-use data from local, State, and other Federal agencies for each State every 5 years (starting in 1950) and categorizes the water use into 11 categories, including mining, industrial, domestic self-supplied, and public supply. The need for a comprehensive understanding of COG water use in the mining category led to the development of an analytical framework to better estimate the water use associated with energy development, regardless of the geographic location or geologic formation (Carter and others, 2016; Valder and others, 2018).

Purpose and Scope

The purpose of this report is to (1) outline a generalized analytical framework for estimating water use associated with the life cycle of COG development in the United States; (2) specify the data needs for estimating water-use coefficients; and (3) provide methods for analyzing and presenting coefficient uncertainty within the data collected, limitations, and assumptions. The framework is intended to complement other water-use topical studies for the USGS Water Availability and Use Science Program to better understand the relation between the production of COG resources for energy and the amount of water needed for this type of energy development in the United States.

This report documents an analytical approach to quantify water use associated with COG development. The analytical approach described herein is intended to be generic so that for other COG reservoirs, the approach can be adapted to the scale of the COG reservoir, the types of data that are available, the extent and resolution at which data are available, and the desired output of the water-use assessment. The analytical approach includes a three-component workflow: (1) input preprocessing; (2) processing, interpretation, and uncertainty analysis of the water-use coefficients; and (3) output postprocessing for visualization and interpretation. An R script (R Core Team, 2019) composing the statistical models used in the analytical framework is provided in the appendix.

Background

In 2000, the USGS National Oil and Gas Assessment began to use subdivisions of the total petroleum system as the basic level of assessment for water-use requirements because the subdivisions were determined to be more closely associated with the generation and migration of petroleum compared to the previously used reservoirs (Schmoker, 2005). Rather than reassessing all the provinces (also referred to as a play or reservoir) in the United States, the National Oil and Gas Assessment team focused on provinces in the United States that were considered a priority for oil and gas resource development. In 2018, 34 provinces were reassessed for undiscovered oil and gas resources, of which 21 were reassessed for COG resources as highlighted in table 1, which lists the mean assessed volume of water for each of the 21 reservoirs.

The 34 reassessed provinces (table 1) represent about 97 percent of the discovered and undiscovered oil and gas resources of the United States (U.S. Geological Survey, 2018). Within the United States, the sum of the mean undiscovered, technically recoverable volume of continuous oil for 21 reassessed provinces is estimated to be 95,838 million barrels, and for 34 reassessed provinces, continuous gas resources are estimated to be 1,405,459 billion cubic feet (U.S. Geological Survey, 2018). When this topical water-use study began (2016), most of the total mean undiscovered oil was within the Williston Basin (fig. 1), but with the recent (2018) reassessment of the Permian Basin and Gulf Coast Basin, the Williston Basin's portion of the total mean undiscovered oil is about 8 percent compared to 74 and 12 percent for the Permian and Gulf Coast Basins, respectively (U.S. Geological Survey, 2018).

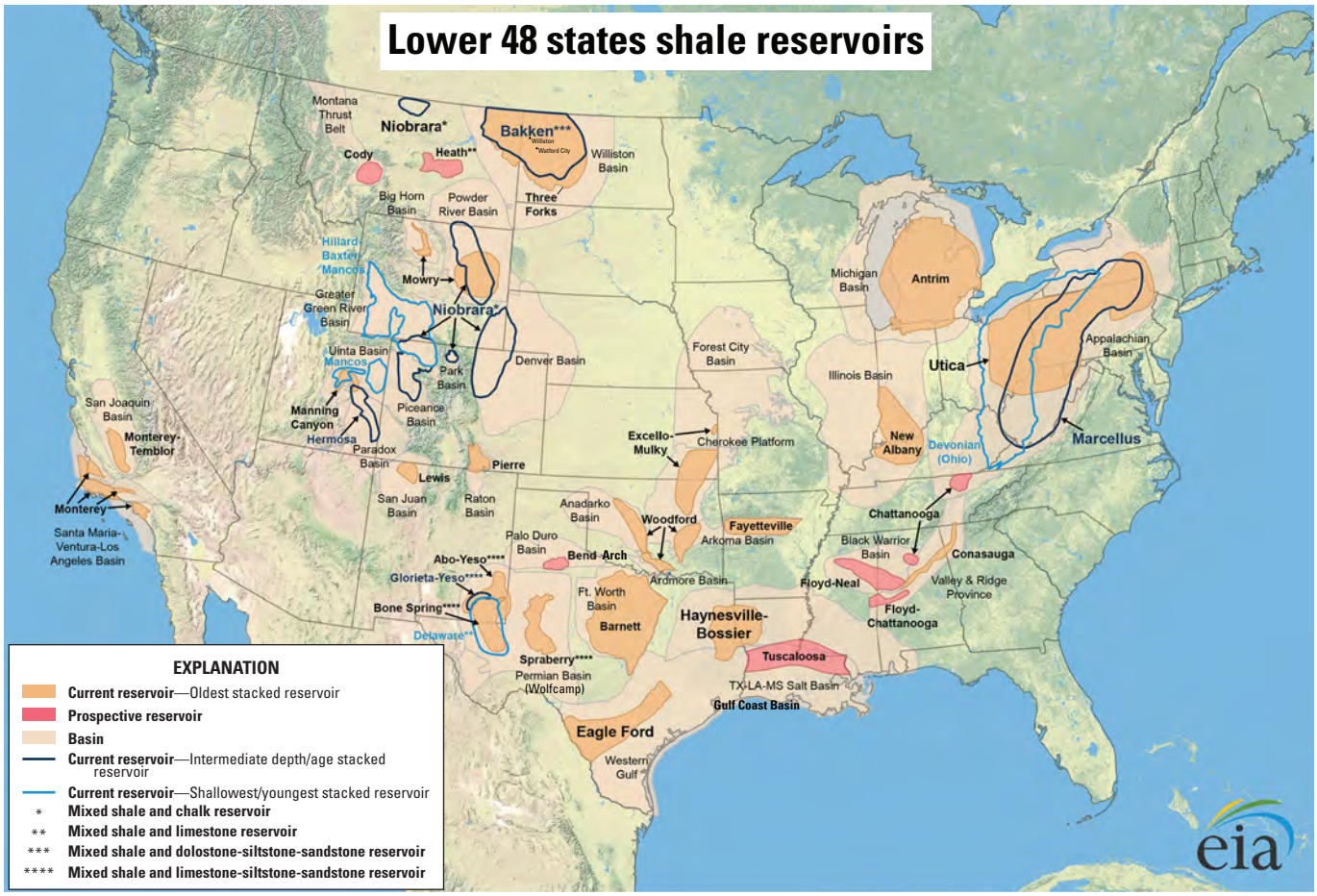
The COG accumulations in the United States can be economically produced using improved techniques such as horizontal drilling and hydraulic fracturing (U.S. Department of Energy, 2019). In general, horizontal drilling exposes larger amounts of thin, horizontal units to the wellbore than do vertical wells, and hydraulic fracturing stimulates the movement of hydrocarbons within units by creating cracks or fractures to allow fluids to flow more freely (Gaswirth and others, 2013).

Table 1. Summary of undiscovered technically recoverable continuous oil and gas resources of 34 provinces reassessed by the U.S. Geological Survey (U.S. Geological Survey Energy Resources Program, 2018).

[MMbbl, million barrels; Bcf, billion cubic feet; NA, not assessed]

Province	Year of most recent oil or gas reassessment	Mean of assessed volume of oil (MMbbl)	Mean of assessed volume of gas (Bcf)
Anadarko Basin	2010	393	24,785
Appalachian Basin	2017	1,404	192,228
Arkoma Basin	2010	NA	36,972
Atlantic Coastal Plain[1]	2011	NA	3,860
Bend Arch-Fort Worth Basin	2015	172	52,985
Big Horn Basin	2008	5	550
Black Warrior Basin	2002	NA	7,056
Burgos Basin[1]	2007	NA	123
Cherokee Platform	2015	460	11,104
Denver Basin	2001	40	2,408
Eastern Oregon-Washington[1]	2006	NA	2,122
Gulf Coast Basin	2018	11,785	393,215
Hannah, Laramie, Shirley Basins[1]	2005	38	19
Illinois Basin	2007	NA	4,235
Los Angeles Basin	2016	13	22
Michigan Basin	2004	NA	7,475
Montana Thrust Belt	2002	28	111
North-Central Montana[1]	2018	637	661
Northern Alaska1	2012	940	60,062
Paradox Basin	2011	471	11,867
Permian Basin (Wolf Camp)	2018	70,459	339,831
Powder River Basin	2002	424	15,475
Raton Basin-Sierra Grande Uplift[1]	2004	NA	1591
San Joaquin Basin	2015	21	27
San Juan Basin[1]	2002	NA	50,419
Southern Alaska[1]	2011	NA	5312
Southwestern Wyoming[1]	2002	104	82,169
Uinta-Piceance Basin	2015	290	81,241
Ventura Basin	2017	4	7
Western Gulf	2007	NA	3,936
Western Oregon-Washington[1]	2009	NA	1,489
Williston Basin	2013	7,622	7,635
Wind River Basin[1]	2018	528	3,908
Wyoming Thrust Belt[1]	2017	NA	559

[1]Not shown on figure 1.



Modified from U.S. Energy Information Administration, June 2016

Figure 1. Prospective continuous oil and gas reservoirs in the conterminous United States (U.S. Energy Information Administration, 2016).

These recent technological advances have rapidly expanded the production of COG reservoirs, most notably in the Bakken and Three Forks Formations in the Williston Basin. Oil production in North Dakota alone increased from a mean of 97,740 barrels of oil per day in 2005 to 1,184,346 barrels of oil per day in 2015 (North Dakota Department of Mineral Resources, 2019). Similar patterns in production are assumed to occur in other COG reservoirs around the United States, such as the Barnett (Fort Worth Basin, Texas), Eagle Ford (Gulf Coast Basin, Tex.), Wolfcamp (Permian Basin), and Marcellus (Appalachian Basin, Pennsylvania; fig. 1) Formations. As mentioned previously, with the recent (2018) reassessment of additional provinces, the total mean undiscovered, technically recoverable volume of COG has increased, highlighting the continued need to develop approaches to better characterize water use associated with COG development.

The conceptual model for assessing total COG water use divides the estimate into three categories: (1) direct, (2) indirect, and (3) ancillary (Valder and others, 2018). Direct water use is defined as water used in a wellbore to complete a well, including the water used for drilling, cementing, stimulating, and maintaining the well during production (Valder and others, 2018). Indirect water use is the water used at or near the well site, including water used for dust abatement, for cleaning equipment, and for crew and staff use. Ancillary water use is 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.

Estimating ancillary water use requires calculating increased public supply water use from population increases because of COG operations. The U.S. Environmental Protection Agency defines public supply water use as water withdrawn by public and private water suppliers that provide water to at least 25 people or have a minimum of 15 connections (U.S. Environmental Protection Agency, 2017; Dieter and others, 2018). It can be assumed that the increased population in areas of energy development may be attributed to the number of COG workers that moved to the area. An example that highlights the relation between increased energy development and population growth is seen in two North Dakota cities within the Williston Basin, Williston and Watford City. The permanent populations of both cities increased from 2010 through 2017, as oil and gas extraction increased in the

basin; the population of Williston increased from 15,940 in 2010 to 25,586 in 2017 (60.5 percent), and the population of Watford City increased from 1,790 in 2010 to 6,523 in 2017 (264.4 percent; U.S. Census Bureau, 2017). As a result of rapid increases in population, municipal utility infrastructures, such as public water supplies, can have an increase in the water demands from the energy industry. Additional water supplies were required for the increased populations, and municipal water permits for the city of Williston increased from 2,545.9 acre-feet in 2010 to 6,249.7 acre-feet in 2017, or a nearly 41 percent increase. Not as easy to document in the population counts by the Census Bureau are the transient workers, hereafter referred to as "temporary workers," or the increase in temporary housing built by oil and gas companies to accommodate workers in the areas of COG development. These temporary workers commonly move into and out of areas of energy development in between official census counts, which may not reflect the actual population; however, these workers place demands on local water supplies while in residence (Jiang and others, 2014).

In rural areas of the country, the increase in population by temporary workers can place new demands on the existing water supply (Horner and others, 2016); for example, the increase in temporary workers can be estimated by the number of active or permitted drilling rigs in an area, which can be a direct cause of the demands on water supply. In rural North Dakota, the North Dakota Industrial Commission documented a mean of 189 active drilling rigs per month from January 2011 through December 2014, which was during the time of rapid increases in oil development (North Dakota Industrial Commission, 2019). Operation of a single drilling rig requires an estimated 50 workers, of which 25 workers are specific to a drilling rig whereas the remaining workers tend to work on several rigs at any given time (Anthony Sarnoski, Luff Exploration Co., oral commun., 2019). Additionally, it is estimated that about 40 workers are needed for a single crew to hydraulically fracture a well and another 10 workers are needed to provide supporting services, such as trucking water or proppant (Lutey, 2017). According to the North Dakota Industrial Commission, the peak number of hydraulic fracturing crews operating in the Williston Basin in North Dakota was 45 in 2014 (North Dakota Department of Mineral Resources, 2019). Even with those 45 crews operating full time, there was still a shortage of workers as estimations of hundreds of wells remained in "drilled but uncompleted" status (Lutey, 2017).

Analytical Framework

The analytical framework presented in this report was based on a conceptual model (Valder and others, 2018) that included various components and definitions necessary to quantify water use associated with COG production. The conceptual model consisted of five elements: (1) input data, (2) processes, (3) decisions, (4) output data, and (5) outcomes. The potential outcomes of the conceptual model were determined by the quality and quantity of the data available for the COG reservoir. An analytical framework is presented in the following sections for the three water-use components—direct, indirect, and ancillary—that are related to the life cycle of COG development as described in Valder and others (2018) and that include input data; data processing, interpretation, and uncertainty analysis; and output data (fig. 2).

The analytical framework was developed as an R script (R Core Team, 2019), which contains the statistical methods used to estimate water-use components. The R script can be modified according to the types of data available within a COG reservoir, the extent and resolution of data available for each component, and the desired output of the water-use assessment. The script requires data in specified units of measure: water variables are in million gallons and oil and gas variables are in million barrels and thousand cubic feet, respectively. The script was written to produce estimates at a per-well scale that can be scaled to the county, State, or region. Water-use estimates include the mean and 10th, 50th, and 90th percentiles of the distribution of the data. Uncertainties are quantified with confidence intervals around the estimated coefficients. Uncertainty for estimated or simulated data can be calculated with the R script by providing a range of representative values that are within the appropriate confidence intervals of the mean of the data.

Overview of Data Requirements and Domain

The availability of data and the level of detail in which the data are collected are important factors in the analytical approach used for a COG reservoir. For example, in the Williston Basin (fig. 1), data used to estimate water use associated with hydraulic fracturing were collected from a variety of sources, including State agencies, Federal agencies, and private organizations, with data available publicly and privately (table 2). Statewide databases that provide, for example, data pertaining to COG production, well counts, and hydraulic fracturing treatments, were available for the Williston Basin from the North Dakota Industrial Commission (2019). Additionally, private databases such as the IHS Markit™ well database (IHS Markit™, 2018) also were available, which allows additional comparisons between datasets on various scales. Hydraulic fracturing treatments also were available for the Williston Basin from a public database, FracFocus (FracFocus, 2018), which can be used to further facilitate comparisons of datasets at various scales. The availability of statewide databases, which may contain data such as COG production and water-use permitting data, can strengthen an assessment of water use if detailed records on permitting are collected and made accessible; for example, water-use permitting data can be used to identify water withdrawals or allocations for specific purposes across a State. In addition to water use and COG well information, nationwide datasets describing population changes, provided by the U.S. Census Bureau (2017), and descriptors of

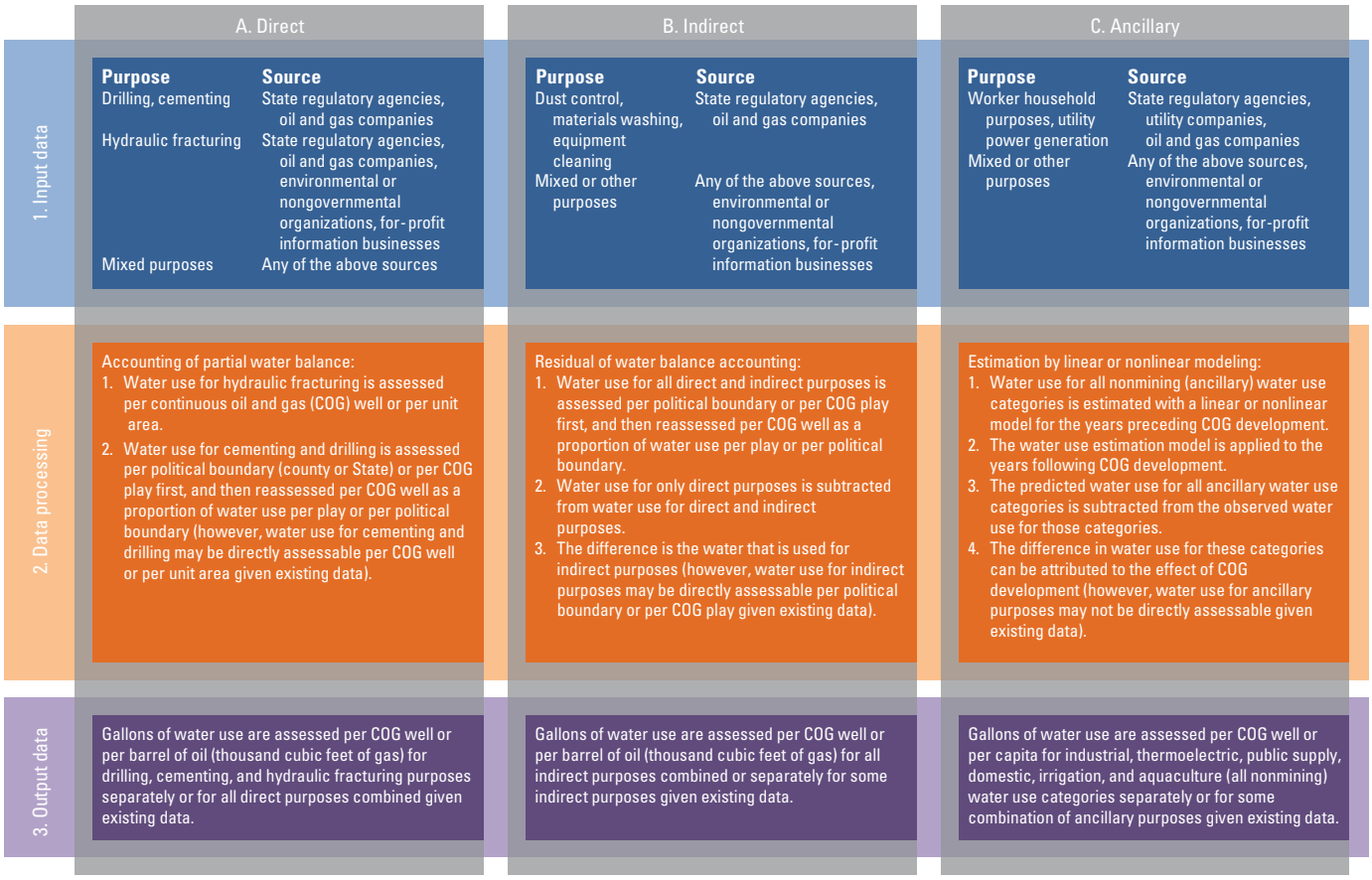


Figure 2. Schematic diagram of the analytical framework and desirable data used for estimating water use associated with continuous oil and gas development.

Table 2. Summary of potential data sources and information categorized by availability and scale for estimating continuous oil and gas water use (from Dutton and others, 2019).

Potential sources for data acquisition	Potential information contained in data	Availability	Scale
North Dakota State Water Commission (2015)	Water permits and reported water use	Public	State
North Dakota Industrial Commission (2019)	Oil and gas well count Oil, gas, and water production Oil and gas hydraulic fracturing treatments Oil and gas well cementing Produced water injection and disposal	Public	State
IHS Markit™ (IHS Markit™, 2018)	Oil, gas, and water production Oil and gas well count Oil and gas hydraulic fracturing treatments	Private	National
FracFocus (FracFocus, 2018)	Oil and gas hydraulic fracturing treatments	Public	National
U.S. Census Bureau (2017)	Population	Federal	National
PRISM Climate Group (2019)	Air temperature and precipitation	Public	National

the regional climate, such as air temperature and precipitation data (PRISM Climate Group, 2019) for a study area, are useful to normalize resource extraction or water use by population in the study area and to account for climate variability in the water-use analysis, respectively. For example, periods of warmer temperatures generally increase evaporation of stored water, which potentially requires increased water use for COG extraction. Conversely, lower temperatures generally decrease evaporation of stored water and could decrease the water required for COG operations.

Data collection and reporting can be variable when comparing datasets. This variability in data is even more apparent when grouping similar datasets together from multiple States or agencies. The variability in the data from various sources highlights the importance of utilizing all available data. Examples that illustrate the potential range of data sources and datasets that could be used in an assessment of a COG reservoir to quantify water use are available for the Williston Basin as a USGS data release by Dutton and others (2019).

The availability of multiple datasets allows for comparisons between datasets and for calculating uncertainty; for example, water-use permit types may need to be consolidated or filtered to identify permits pertaining to hydraulic fracturing. It also may be necessary to acquire water use, hydraulic fracturing treatment, and COG production data from multiple States; for example, in the Williston Basin, water-use data were collected from multiple States because of the extent of the basin (fig. 1). Water permitting and reporting requirements may differ depending on the State agency standards for data collection, permitting, and reporting. Inconsistency across State boundaries in data collection or reporting could increase the uncertainty and variability of water-use estimates.

In selecting a domain to assess water use related to COG development, the spatial extent of the COG reservoir is an important factor for COG water-use estimations. In addition to the geologic boundaries of the COG reservoir, the boundaries of the governing agencies in and adjacent to the reservoir, such as counties, States, or Canadian Provinces, also are important because those agencies generally collect and maintain the data necessary for analysis. Hydrologic features, such as watersheds and aquifers, also could be considered because they may affect water availability and permitting. In addition to spatial extents, temporal limitations also require consideration. The temporal extent is particularly important if analyses comparing conventional oil and gas development with COG development are of interest. These analyses in general benefit from acquiring long-term (years), time-relevant data records to establish baseline conditions.

Direct Water Use

Direct water use is defined as the water used directly in developing the well itself and in maintaining the well during production (Valder and others, 2018). This section provides an overview of the sources of data that could be used in estimating direct water use and a detailed description of the analytical approach used to estimate direct water use using the R script (appendix). A synopsis of the direct water-use analysis is shown in figure 2.

Data Sources

Direct water-use data are available from multiple sources that include private databases, national databases, or State agency databases (table 2). The quality and availability of these data are subject to reporting requirements by those agencies or organizations collecting or compiling the data. If multiple estimates of direct water use are available, all data can be used to estimate the central tendency of direct water use and the variability of the data. Temporally, datasets providing the most use in this assessment would be reported for each year or a period of years by well, county, or other spatial or temporal boundaries. The methodology of how the various datasets or similar datasets can be used to estimate direct water use is described in the following sections.

Analytical Approach

The analytical approach to estimating direct water use for a specific COG reservoir is provided as an adaptable tool, an R script, that can be modified as needed (see appendix). One section of the script processes and standardizes the data input, another section analyzes water-use data and uncertainty in the parameter estimates, and a third section processes the output. A synopsis of the direct water-use analysis is shown in figure 2.

Input Data Standardization

The inputs to the direct water-use analysis can include volume of water use, year of observation, and county of observation (or another spatial unit, such as a square-mile-grid cell) in columns and the corresponding spatial or temporal observations by row. The data are input as multiple objects, one for each direct purpose: drilling, cementing, and hydraulic fracturing. Ideally, these data would be observations of water use for each direct purpose on a per-well basis; however, data on water use for hydraulic fracturing purposes are most likely available for an entire COG reservoir and not at the per-well level of detail.

Water-use estimations for well completion, including cementing and drilling, require several assumptions. New data on cementing may become available, at least in part, but typically the data for cementing is generally reported by sacks of cement used, which requires assuming the water required per bag of cement; for example, an experienced driller familiar with the reservoir and hydraulic fracturing technique may estimate that a mean of 7 gallons of water is used per sack of cement. Data for the water required for drilling also are not reported for every well; therefore, assumptions are necessary, such as assuming that the drilling process for a COG well

requires 50 percent of the water volume required to cement the same well. Additionally, the water required for drilling depends on the geologic material and the depth and length of the borehole. These data are not necessary to run the R script; however, the final estimate could underestimate the direct water use for the well if the data are excluded or unavailable.

Data Processing, Interpretation, and Uncertainty Analysis

Direct water-use analysis begins with estimating a temporal water-use breakpoint using a segmented linear regression method. Breakpoint analysis identifies the year in which the temporal trend of water use changes. A hypothetical example of the breakpoint analysis that identifies the breakpoint year is shown in figure 3. For this example, the change in the trend is attributed to the initial period of COG development. Another example would be to apply this similar approach for water-use estimates based on the number of wells; a linear regression is fit to the direct water-use data with the volume of water use as the response and the year of observation as the predictor. To apply this regression, the direct water-use data are summarized to single values per year of volume of water use; that is, all observations of water use per year are summed. The fitted linear regression is then updated by including a piecewise linear relation using the "segmented" package in R (Muggeo, 2018). This segmented linear regression determines the year the water use changed. The identified year is the estimated breakpoint between years before and after the beginning of COG development. All analyses for direct, indirect, and ancillary water use depend on the breakpoint analysis for identifying the baseline data associated with water use. If a breakpoint cannot be estimated, or if multiple breakpoints are estimated for the segmented linear regression, then additional information from other published sources, such as well drilling reports, local municipal water-use reports, and treatment reports, could be used to select an appropriate year representing the start of COG development.

Direct water use was estimated with the data that follow the breakpoint by fitting a simple linear regression with the volume of water use as the response and number of wells developed as the predictor. To fit this regression analysis, the direct water-use data are aggregated into single values of volume of water use and number of wells developed per county (or some other spatial unit) per year. These aggregated data are the sampling units used in all analyses for direct, indirect, and ancillary water use for initially fitting the regression analysis and then validating the fitted linear regression analysis, which is explained in the following section. Linear regression analyzes the mean of the distribution of the data, which may not provide enough detailed information at a given temporal or spatial scale. To provide additional information about potential water use, quantile regression is used to fit models of the 10th, 50th, and 90th percentiles of the sampling distribution using the "quantreg" package in R (Koenker, 2018). Quantile regression is used to provide additional understanding of direct, indirect, or ancillary water use at the extremes of the sampling distribution, or for a skewed sampling distribution (Koenker, 2005). For each parameter, a 95-percent confidence interval around the estimate is used for assessing uncertainty. The R script includes functions with built-in arguments that can be user defined to estimate coefficients for any percentile of the data distributions with levels of confidence around the parameter estimate (for example, 90-percent or 99-percent confidence intervals instead of the default of 95 percent).

To validate the fitted linear and quantile regression analyses, leave-one-out cross validation (Hastie and others, 2009) is used on the sampling units of the data; that is, the aggregated values of volume of water use and number of wells developed per county per year. A linear or quantile regression is fit to the data for all sampled units with the exception of a single unit that is intentionally omitted from the analysis. The fitted regression is used to predict the water use of the unit that is omitted. This analysis is repeated for every unit. Once a unit is removed and after the cross validation is completed, the analysis adds the unit that was previously omitted. All predictions of the omitted sampling units can be compared against their observed values using goodness-of-fit metrics, such as root mean square error, mean absolute error, or coefficient of determination (Hastie and others, 2009), which can be used to evaluate the performance of the regressions in estimating the volume of direct water use against the number of wells developed.

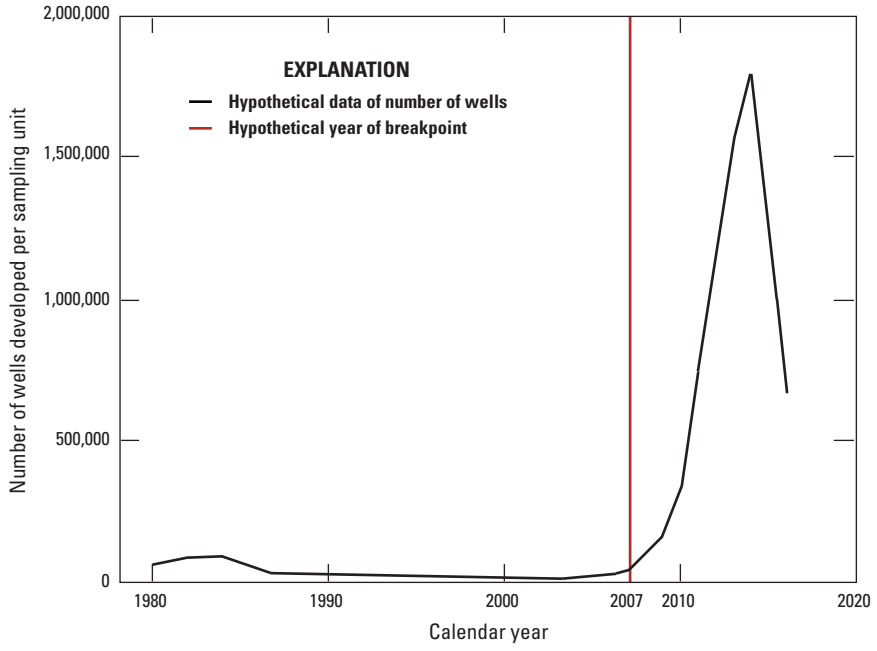


Figure 3. Hypothetical breakpoint analysis output showing the change in the number of wells drilled per year.

Output Data Visualization

The resulting output from the direct water-use analysis can be interpreted or visualized in tabular and graphical forms with estimated values from the analyses as water use in million gallons per well. By modifying the R script, output water-use estimates can be changed to units of water use per barrel of oil, or water use per cubic foot of natural gas. The R script produces a table that includes (1) columns of estimated values for the mean and 10th, 50th, and 90th percentiles of the distributions of the data and 95-percent confidence intervals around these parameters and (2) rows for drilling, cementing, and hydraulic fracturing purposes individually and collectively for all direct purposes (table 3). The generated figures plot the volume of water use compared to the number of wells. A hypothetical example of the resulting output plot is shown in figure 4. One figure is produced for each parameter of the sampling distributions (that is, the mean and percentiles) for each direct purpose, including total direct water use. The figures include points for the observed or simulated values of water-use values for each direct use subcategory, a trend line for the linear (or quantile) regression model of the mean (or percentiles), and a ribbon for the confidence interval around the modeled parameter.

Indirect Water Use

Indirect water use is defined by Valder and others (2018) as the water used at or near a well pad but not for direct purposes. Examples of indirect water use include dust abatement, equipment cleaning, materials washing, worker sanitation, and site preparation (Valder and others, 2018). This section provides an overview of the sources of data that could be used to estimate indirect water use and a detailed description of the analytical approach to estimate indirect water use using the R script (appendix). A synopsis of the indirect water-use analysis is shown in figure 2.

Data Sources

Indirect water use is estimated by comparing direct water use with reported and permitted water-use volumes supporting COG development. This process involves acquiring water-use permit data from local sources, such as a State water agency, and identifying permits supporting COG development. Then, the data are aggregated into spatial and temporal units of interest. An example of how these data may be used for the indirect water-use estimation would be to assess water permits annually for the individual water depots. Water permits for water depots generally are classified as industrial water permits, temporary permits, or agricultural permits. Permitted water withdrawals are of higher priority interest than temporary permits because these data would establish an upper bound on the total water use to account for possible reporting inconsistencies in actual water withdrawal information. The total reported and permitted water withdrawals are an initial starting point for constructing a water budget to estimate indirect water use. Aggregating all the data available in an area and classifying data appropriately are iterative processes because data collection and data types may vary between COG production locations. Example data for estimating indirect water use are available as a USGS data release (Dutton and others, 2019) for the Williston Basin. The methodology of how the various datasets or similar datasets can be used to estimate indirect water use is described in the following sections.

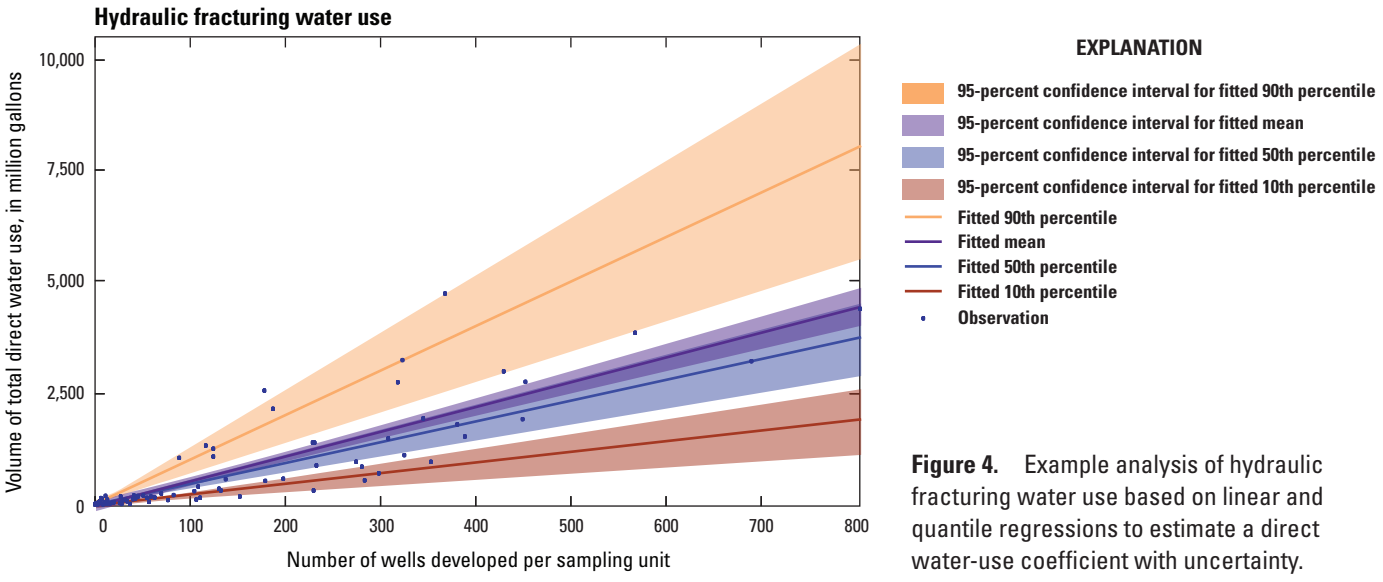


Figure 4. Example analysis of hydraulic fracturing water use based on linear and quantile regressions to estimate a direct water-use coefficient with uncertainty.

Analytical Approach

The approach in the R script (appendix) to analyzing indirect water use requires that direct water use be estimated as an initial step in computing indirect water use. Without the direct water-use analysis, indirect water use cannot be estimated because of the high likelihood that the data for direct and indirect water-use purposes are aggregated into one dataset. A synopsis of the indirect water-use analysis is shown in figure 2.

Input Data Standardization

The inputs for the indirect water-use analysis require preprocessing to an appropriate data structure with rows for observations, columns for year and county of observation, and estimates of volume of water use. The data are input as a single object for total indirect calculations. Unlike the direct water-use data, which could be observations of water use on a per-well basis, data on water use for indirect purposes at a per-well scale likely are not available. These data may not be attributable to each well and instead may be determined only from water-use reporting from permits held by oil and gas companies or by other organizations appropriating water for COG development; however, it is important to note that this reported water use is a total accounting of all water use for direct and indirect purposes of developing wells. Therefore, without data for direct water use, it would be impossible to estimate indirect water use. If data are missing for indirect purposes, the R script can be run to calculate water use for direct and ancillary uses.

Data Processing, Interpretation, and Uncertainty Analysis

Estimating water use for total indirect purposes is similar to estimating water use for direct purposes. Data processing for indirect water-use estimates depends on the calculated breakpoint for splitting data into two water-use time domains: pre- and post-COG development (fig. 3). Data on water use for indirect purposes are derived from the reporting of total direct and indirect water use by permits approved for developing COG wells; however, some of the reported water use may be unrelated to COG development. Total water-use data reported by the permits are analyzed by fitting a simple linear regression with the volume of water use as the response and the year of observation as the predictor for years preceding the breakpoint. Data need to be summarized as single values of volume of water use per year for this method. A simple linear regression is used to predict the total water use for years after the breakpoint. The remaining water use (that is, the difference between the observed values and the predicted values for each year) is that which is related to COG development, although this remainder is still inclusive of the total direct and indirect water use. Direct water use is subtracted from the data; however, the total water-use data are in units of volume per year, whereas the direct water-use data are in units of volume per year per county. The total water use per year can be distributed per county based on the ratio of the number of wells developed for any county to the total number of wells developed for all counties. The direct water use is removed from the total water use, which results in the indirect water-use data in a volume per year per county.

The linear regression is applied to predict the total water use and is based on a simulated mean of the distribution of the data. The upper and lower limits of a 95-percent confidence interval around the parameter estimate also can be used as input values for the fitted regression, introducing uncertainty in the indirect water-use data. There may be additional uncertainty related to the permits that support COG development, so that uncertainty related to COG development can be assigned to the permits before running the R script.

Linear regression and quantile regression are applied to fit models of the mean and the 10th, 50th, and 90th percentiles of the sampling distribution (Koenker, 2018), and 95-percent confidence intervals around the parameter estimates are applied to evaluate uncertainty. Leave-one-out cross validation (Hastie and others, 2009) is applied to validate the fitted linear and quantile regression models. The performance of the regressions in estimating the volume of indirect water use against the number of wells developed is assessed using goodness-of-fit metrics (Hastie and others, 2009).

Output Data Visualization

The output from the indirect water-use analysis is postprocessed like the direct water-use analysis. The output is a tabular and graphical depiction of estimated coefficients for water use in units of million gallons per well. The R script can be modified to provide the outputs in units of per barrel of oil or water use per cubic foot of natural gas, if desired. The output table format is similar to that of the direct water-use table, with the exception that there is only one row of output that represents the total of indirect water uses (table 4). If the data were available, the graphical output could be similar to the direct water-use analysis if data were reported on an individual indirect purpose rather than summarized into a total indirect purpose. An example hypothetical graphical output showing the total indirect water use is shown in figure 5.

Ancillary Water Use

Ancillary water use is defined as the water used to support any COG development that is not categorized for direct or indirect purposes (Valder and others, 2018). Examples of ancillary water use include utility power generation and domestic uses (Valder and others, 2018). The ancillary water use is the most difficult category to summarize because this ancillary category would be anything remaining that would increase water use in an area, regardless of whether it was directly or indirectly related to COG. This section provides an overview of the sources of data that could be used for estimating ancillary water use and a detailed description of the analytical approach to estimate ancillary water use using

Table 3. Hypothetical example of the output table generated for the direct water-use category showing the coefficients of the mean and 10th, 50th, and 90th percentiles in volumes of water use.

[Hypothetical values in table are in million gallons per well]

Direct water use	Mean			10th percentile			50th percentile			90th percentile		
	Simulated	Lower confidence limit	Upper confidence limit	Simulated	Lower confidence limit	Upper confidence limit	Simulated	Lower confidence limit	Upper confidence limit	Simulated	Lower confidence limit	Upper confidence limit
Drilling	0.013	0.013	0.014	0.012	0.011	0.013	0.013	0.013	0.014	0.015	0.014	0.016
Cementing	0.017	0.017	0.017	0.015	0.014	0.016	0.017	0.016	0.017	0.019	0.018	0.021
Hydraulic fracturing	5.540	4.942	6.138	2.383	1.372	3.393	4.651	3.560	5.743	9.956	6.726	13.187
Total direct water use	**5.569**	**4.971**	**6.168**	**2.409**	**1.399**	**3.418**	**4.681**	**3.587**	**5.774**	**9.990**	**6.758**	**13.223**

Table 4. Hypothetical example of the output table generated for the indirect water-use category showing the 10th, 50th, and 90th percentile coefficient values in volumes of water use.

[Hypothetical values in table are in million gallons per well]

Indirect water use	Mean			10th percentile			50th percentile			90th percentile		
	Simulated	Lower confidence limit	Upper confidence limit	Simulated	Lower confidence limit	Upper confidence limit	Simulated	Lower confidence limit	Upper confidence limit	Simulated	Lower confidence limit	Upper confidence limit
Total indirect water use	0.901	0.840	0.962	0.642	0.413	0.872	0.938	0.753	1.123	1.114	0.989	1.238

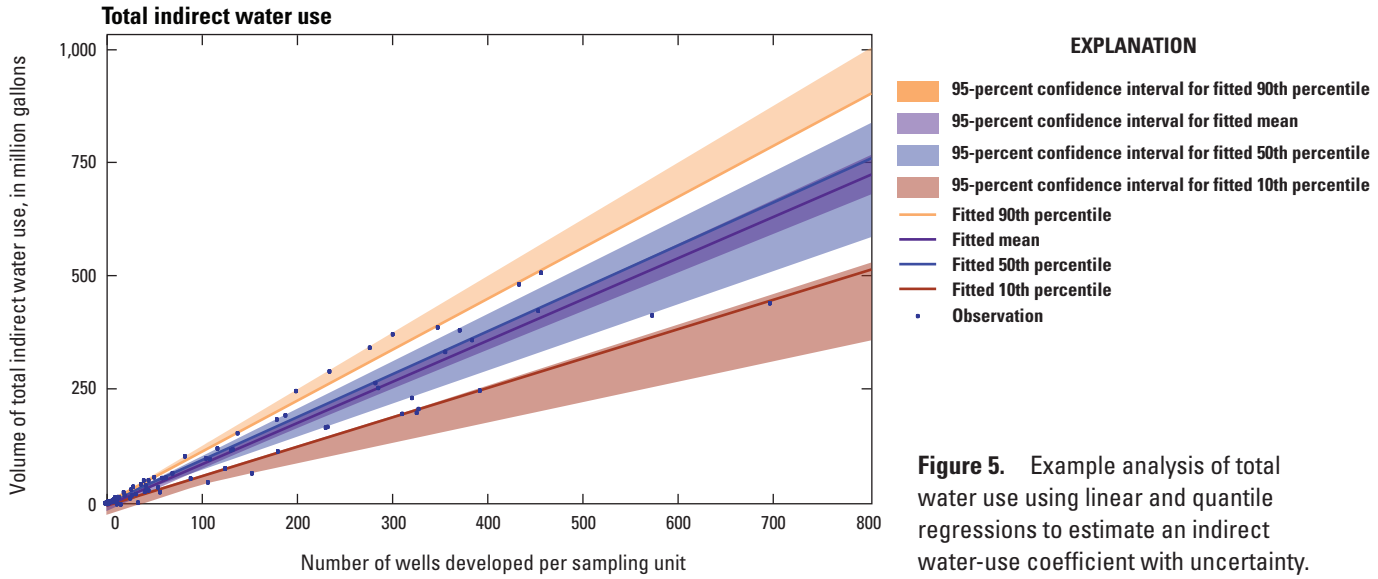


Figure 5. Example analysis of total water use using linear and quantile regressions to estimate an indirect water-use coefficient with uncertainty.

the R script (appendix). A synopsis of the ancillary water-use analysis is shown in figure 2.

Data Sources

The data used for estimating ancillary water use for COG development come from a variety of sources. The type of data collected and used depends on the analytical approach used to estimate ancillary water use for a particular reservoir, basin, or region of interest. Water withdrawal data from permit applications and licenses may be available for municipal, irrigation, industrial, rural water, and multiple use types that could be used to determine ancillary water use. Climate data for the region and the time of interest also could be obtained to remove any effect of climate variation in the ancillary water-use analyses. A dataset that includes the pre-COG development could be beneficial in analyzing, for example, a time series of wells drilled for oil and gas production. Local and State agencies would be likely sources of data that would include the water withdrawals for domestic, industrial, irrigation, multiple use, municipal, power generation, and rural water. In addition, Federal agencies may provide additional data, such as gridded air temperature and precipitation products, which can be used to correct ancillary water-use analyses for climate variability (PRISM Climate Group, 2019). A time series of wells completed per year could be used to identify when COG development started in a particular area and reservoir. Example data for estimating ancillary water use are available as a USGS data release (Dutton and others, 2019) for the Williston Basin. The methodology of how the various datasets or similar datasets can be used to estimate ancillary water use is described in the following sections.

Analytical Approach

The analytical approach for estimating the ancillary water use does not depend on the indirect water-use analysis; however, it is dependent upon the direct water-use analysis. Ancillary water use can be estimated for the entire COG reservoir but analyzing the uncertainty in the parameter estimates will be less robust if an analysis of direct water use is not completed. A synopsis of the ancillary water-use analysis is shown in figure 2.

Input Data Standardization

The inputs for ancillary water use are preprocessed and formatted with the same data structure as that used for the direct and indirect water-use analysis. The input data are formatted as a single object for all ancillary uses. Similar to the data for indirect purposes, the data for ancillary uses may not be for a specific well. Instead, the data reported as ancillary water use may be obtained from State water permits that were approved as a water appropriation by the applicable State organization. Examples of ancillary water use that are not directly or indirectly related to COG development can be grouped into the same water-use categories as the USGS 5-year national water-use compilations. These categories include, in part, mining, industrial, public supply, and domestic self-supplied categories. If datasets are not available for ancillary uses, then the R script will still calculate estimates for direct and indirect water uses.

Data Processing, Interpretation, and Uncertainty Analysis

Data processing of the ancillary water-use analysis is similar to that of the indirect water-use analysis. Data for ancillary

water use are summarized from the water-use permits that have been approved for uses other than direct COG water-use activities, such as permits designated as irrigation of domestic uses. Water use reported by the permits is analyzed by fitting a multiple linear regression with the volume of water use as the response and a combination of the year of observation, total annual precipitation, and mean annual temperature as the predictors for years preceding the breakpoint analysis for the reservoir. Air temperature and precipitation variables are used in the analysis to remove the effect of climate variability on ancillary water use. Data summarized as single values in the volume of water use per year are input for each ancillary use. The fitted regression analysis is applied to predict the ancillary water use for years that are post-COG development. The residual water use is in volume per year, which if these data are reported as the ancillary water use in a volume per year per county, the residual water use per year can be distributed per county based on the ratio of the number of wells developed and the total number of wells developed for all counties. The fitted regression predicting the ancillary water use is based on a simulated mean of the distribution of the data. Uncertainty in the ancillary water-use data can be quantified by the upper and lower limits of a 95-percent confidence interval around the parameter estimate for each ancillary use.

Estimating the water use for ancillary uses is similar to estimating direct and indirect uses. Models of the mean and the 10th, 50th, and 90th percentiles of the sampling distribution (Koenker, 2018) are fit using simple linear regression and quantile regression, and 95-percent confidence intervals applied to the parameter estimates are used to evaluate uncertainty. Leave-one-out cross validation (Hastie and others, 2009) is used to validate the fitted linear and quantile regressions. The performance of the regressions in estimating the volume of ancillary water use against the number of wells developed is assessed using goodness-of-fit metrics (Hastie and others, 2009).

Output Data Visualization

The resulting output from the ancillary water-use analysis is summarized in tabular and graphical forms as estimated coefficients of water use in million gallons per well. The output results are postprocessed similarly to the direct and indirect water-use analyses. The R script can be modified to estimate water use per barrel of oil or water use per cubic foot of natural gas. The output table has a similar format compared to the direct or indirect water-use analyses, summarizing each of the individual ancillary uses (for example, public supply or domestic; table 5). The figures generated are similar to the direct water-use analysis; there is one figure generated for each parameter of the sampling distributions for each ancillary use. The generated figures plot the volume of water use compared to the number of wells. A hypothetical example of the resulting output plot is shown in figure 6.

Unlike the direct and indirect water-use estimates, which can only have positive water-use coefficients, water-use coefficients for ancillary uses may be positive or negative. Ancillary water use represents additional water uses that might not have been observed without COG development; for example, if the industrial water-use category has a positive estimated coefficient, this would mean that water use in that category increased in relation to COG development. This increase above expected industrial water use would be attributed to COG development beyond water use for direct and indirect purposes. Alternatively, ancillary water use can represent a reduction in water use with COG development present. If the estimated coefficient is negative, then less water was used during periods of COG development than would be expected. The interpretation of a negative coefficient for ancillary water use means that COG development decreased the water consumed in a specific water-use category.

Water-Use Coefficients and Uncertainty

Estimates of direct, indirect, and ancillary water use can be correlated to the number of COG wells developed, the barrels of oil or cubic feet of gas produced, or the number of persons in the specific COG reservoir. Simple linear regressions relating the direct, indirect, or ancillary water use to the well count, oil and gas production, or population are fit as follows:

$$y=\beta_0+\beta_1 x+\varepsilon \quad (1)$$

where

- y is the predicted direct, indirect, or ancillary water use;
- β_0 is the y-intercept;
- β_1 is the sensitivity of y to a change in x;
- x is the observed number of wells developed, volume of oil or gas produced, or number of persons in a COG reservoir; and
- ε is the random error.

Confidence intervals bracketing the parameter estimates are used to assess uncertainty in the direct, indirect, and ancillary water use as it relates to the number of wells developed, the volume of oil or gas produced, or the number of persons in a COG reservoir area. Linear regression analysis simulates the mean of the distribution of the data. Additional information pertaining to water-use data can be simulated using quantile regression, which may be important for understanding direct, indirect, or ancillary water use associated with other parameters of the sampling distribution. Examples include simulating the 90th percentile of the distribution for predicting water use on a per-well basis or simulating the 50th percentile (that is, the median) for information at the per-well scale.

The following example demonstrates a hypothetical analysis, used for illustration purposes only, for direct water use. This hypothetical example includes simulating water use per well as a mean value using linear regression and as 10th-,

Table 5. Hypothetical example of the output table generated for the ancillary water-use category showing the 10th, 50th, and 90th percentile coefficient values in volumes of water use.

[Hypothetical values in table are in million gallons per well]

Ancillary water use	Mean			10th percentile			50th percentile			90th percentile		
	Simulated	Lower confidence limit	Upper confidence limit	Simulated	Lower confidence limit	Upper confidence limit	Simulated	Lower confidence limit	Upper confidence limit	Simulated	Lower confidence limit	Upper confidence limit
Domestic	0.138	0.103	0.173	−0.153	−0.293	−0.013	0.189	0.047	0.332	0.228	0.192	0.263
Industrial	0.425	0.273	0.577	0.118	0.096	0.139	0.338	0.067	0.608	2.279	1.470	3.088
Irrigation	−8.905	−9.857	−7.953	−16.920	−21.252	−12.589	−8.465	−9.485	−7.444	−4.419	−7.324	−1.515
Mining	0.249	0.153	0.345	−0.016	−0.101	0.069	0.317	0.080	0.555	0.988	0.286	1.689
Public supply	1.435	1.152	1.719	0.133	−0.408	0.675	1.092	0.843	1.341	3.944	1.675	6.214
Thermoelectric power	−2.016	−2.451	−1.580	−5.733	−9.201	−2.265	−1.882	−2.415	−1.349	−0.837	−1.447	−0.227

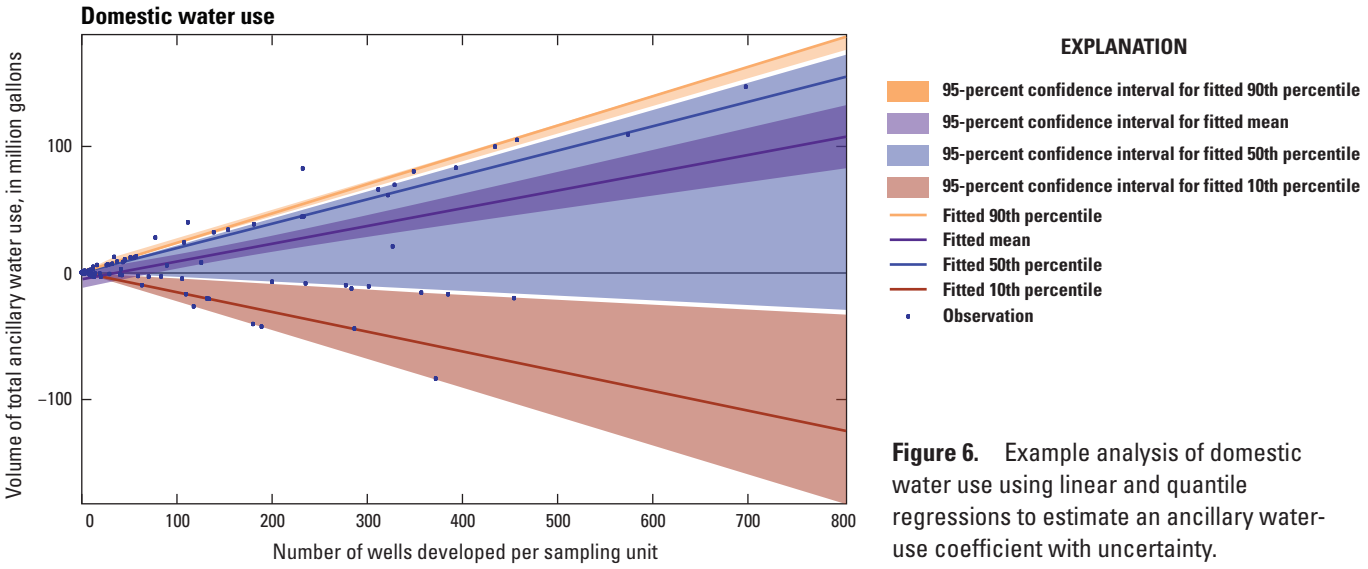


Figure 6. Example analysis of domestic water use using linear and quantile regressions to estimate an ancillary water-use coefficient with uncertainty.

50th-, and 90th-percentile values using quantile regression (fig. 4). The analysis yields a β_1 coefficient with a value of 3.72 million gallons (Mgal) per well for the mean estimate, and 1.65, 3.10, and 6.69 Mgal per well for the 10th-, 50th-, and 90th-percentile estimates, respectively. However, the β_1 coefficients are estimated from water-use data that span several years; a 95-percent confidence interval around the β_1 coefficient for the mean estimate is 3.39–4.05 Mgal per well, and a 95-percent confidence interval around the β_1 coefficient for the 10th-, 50th-, and 90th-percentile estimates are 1.05–2.25, 2.50–3.70, and 4.86–8.53 Mgal per well, respectively. Linear and quantile regressions, with confidence intervals around the coefficients estimates, can provide a reasonably robust analysis of direct, indirect, and ancillary water use.

Sensitivity of the analytical approach described herein to the availability of specific data sources is identified by Valder and others (2018), where an analysis may be most sensitive to direct water-use data for hydraulic fracturing purposes. Even with sparse data for a COG reservoir, a partial assessment may be achievable; however, the more comprehensive the data available for a COG reservoir are, the more robust an assessment of water use associated with COG development can be. The partial assessment will still provide important information on water use within the COG reservoir; however, the results may have larger confidence intervals associated with the resulting interpretations. With limited data, outcome 6 in figure 5 of Valder and others (2018) would be the most complete assessment possible, but as particular sources of water-use data become more limited, especially those for permits of appropriated water, the more incomplete an assessment becomes. Although a total accounting of all water use may be most sensitive to direct water-use data for hydraulic fracturing purposes, the general utility of an assessment is more sensitive to being able to estimate the indirect and ancillary water use, which can be used to characterize water use associated with COG development.

Data and Analytical Framework Limitations

The analytical framework has several limitations because of the assumptions and simplifications necessary to calculate water-use changes caused by COG development. Numerical simulations representing a complex system are, by design, created to simplify natural conditions. These simplifications rely on assumptions and estimates, which include differing degrees of uncertainty to be introduced in the numeric output. Several limitations and assumptions may need to be made to fully characterize the water use associated with COG production as it relates to direct, indirect, and ancillary water uses. Much like other analytical models, the approach described herein, and the resulting output are dependent on the availability of data. Commonly, when the available data overlap the water-use categories, overestimations and underestimations could occur. Limitations on interpreting the output may include, in part, the possibility that water-use data for direct purposes is underreported, which in turn, means that indirect water use is overestimated in the analysis; water-use data may be categorized as both direct and indirect purposes, which would mean that indirect water use is overestimated, and difficulties in determining permitted water appropriations related to COG development could result in the overestimation of the indirect water, whereas ancillary water use is underestimated.

The indirect and ancillary water-use analyses depend on estimating a temporal breakpoint in water use for direct purposes, which is used for separating data into water pre- and post-COG development (fig. 3). The year indicated by the breakpoint analysis can be modified if additional information, such as permits, supports pre-COG development, which may be relevant in the water-use estimations. The year selected with the breakpoint analysis can affect the parameter estimates from the simple linear regression fit to the years preceding the

breakpoint, which in turn, will affect the residual water use for the years after the breakpoint. Water-use trends before the breakpoint are assumed to be extrapolated forward as the post-COG development occurs.

The analytical approach is limited by the statistical analysis applied to the data. Simple linear or quantile regression may not be the best statistical approach as compared to a more complex nonlinear or multivariate approach. A more complex analysis could have less bias but a greater variance in predictions, resulting in overfitting of the data. Alternatively, simple linear and quantile regressions may underfit the data because they will result in predictions with decreased variance but increased bias. The simple linear statistical models used in this analytical approach more closely align the resulting outputs with the water-use coefficients associated with COG development. Additionally, this analysis may be applied to estimate the water use for categories used in the USGS 5-year national water-use compilations.

The linear and quantile regressions in the water-use analysis are validated using leave-one-out cross validation. Other validation procedures also may be appropriate, including bootstrap, jackknife, or k-fold cross validation (Hastie and others, 2001, 2009). The performance of the linear and quantile regressions of water use is assessed with the goodness-of-fit metrics root mean square error, mean absolute error, and coefficient of determination. Although other metrics also may be applied, these metrics are commonly used to evaluate hydrologic models and are recognized to capture important information about the validity of a model's predictions from the observed data (Moriasi and others, 2007).

Uncertainties in the estimated water-use coefficients from the linear and quantile regressions are analyzed using confidence intervals, which represent a range that may include the mean (or a percentile) value of multiple observations. Access to datasets, spatial or temporal, is a limiting factor for most numerical assessments. Limited availability and the inconsistency in data collection among agencies can create more uncertainty within output results than robust data sources. Although a partial assessment may be completed with limited or sparse data, the results could have a greater magnitude of confidence intervals associated with the results. Based on the results of potential data limitations or uncertainties, a modification of the script (appendix) may be necessary to apply the analytical approach to another COG reservoir, depending on the type of data available.

Summary

An analytical framework to estimate water use associated with continuous oil and gas (COG) development was developed for the U.S. Geological Survey Water Availability and Use Science Program. This framework was developed to better understand the relation between the production of COG resources for energy and the amount of water needed to sustain this type of energy development in the United States. The total mean undiscovered, technically recoverable volume of COG has increased, highlighting the continued need to develop approaches to better characterize water use associated with COG development.

The analytical framework can be used to estimate water use associated with COG development for three water-use components—direct, indirect, and ancillary water use—that are related to the life cycle of COG development. Direct water use is defined as water used in a wellbore to complete a well, including the water used for drilling, cementing, stimulating, and maintaining the well during production. Indirect water use is the water used at or near the well site, including water used for dust abatement, for cleaning equipment, and for crew and staff use. Ancillary water use is 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 analytical framework includes the data inputs, the processes involved in estimating the water-use coefficients and analyzing their uncertainties, and the outputs. The analytical framework was developed as an R script, which contains the statistical models used to estimate water-use components.

The availability of data across COG reservoirs in the United States is variable and presents challenges associated with estimating water use associated with COG reservoirs; thus, the R script can be modified according to the types of data available within a COG reservoir, the extent and resolution of data available for each component, and the desired output of the water-use assessment. The script was written so that the units of the data in the script were standardized. Water-use estimates are simulated for the mean and 10th, 50th, and 90th percentiles of the distributions of the data. Uncertainties are quantified with confidence intervals around the estimated coefficients. Uncertainty for estimated or simulated data can be calculated with the R script by providing a range of representative values that are within the appropriate confidence intervals of the mean of the data.

Examples of sources of input data that may be available are provided for direct, indirect, and ancillary water uses. The preprocessing structure of inputs for use in the R script is described. The processing of data includes a breakpoint assessment for pre- and post-COG development. Linear and quantile regressions are applied to fit models of the mean and selected percentiles of the sampling distribution. The resulting output for the direct, indirect, and ancillary water uses from the R script includes tabular and graphical output.

Water-use coefficients can be developed using simple linear regressions relating the direct, indirect, or ancillary water-use estimates to various parameters such as well counts, barrels of oil and gas produced, or population within a COG reservoir. Uncertainties in the estimated water-use coefficients can be analyzed using confidence intervals. The availability and quality of data for a particular reservoir will affect model limitations. Limited availability and the inconsistency in data collection among agencies can create more uncertainty within output results than robust data sources.

References Cited

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.]

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., 2019, Data to estimate water use associated with continuous oil and gas development, Williston Basin, United States, 1980–2017: U.S. Geological Survey data release, https://doi.org/10.5066/P9CPKRLW.

FracFocus, 2018, FracFocus data download: FracFocus Chemical Disclosure Registry, accessed April 13, 2018, at https://fracfocus.org/data-download.

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 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.]

Hastie, T., Tibshirani, R., and Friedman, J., 2001, The elements of statistical learning—Data mining, inference, and prediction: New York, Springer, 745 p.

Hastie, T., Tibshirani, R., and Friedman, J., 2009, The elements of statistical learning—Data mining, inference, and prediction 2d ed.: New York, Springer, 745 p. [Also available at https://web.stanford.edu/~hastie/ElemStatLearn/.]

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, US well data: IHS Markit™ Well Database, accessed April 16, 2018, at https://ihsmarkit.com/products/us-well-data.html. [Available from IHS Markit™, 15 Inverness Way East, Englewood, Colo., 80112.]

Jiang, M., Hendrickson, C.T., and VanBriesen, J.M., 2014, Life cycle water consumption and wastewater generation impacts of a Marcellus shale gas well: Environmental Science & Technology, v. 48, no. 3, p. 1911–1920. [Also available at https://doi.org/10.1021/es4047654.]

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. [Also available at https://cran.r-project.org/web/packages/quantreg/.]

Lutey, T., 2017, As the Bakken heats up again, Williston prepares for more growth: Billings Gazette, March 5, 2017, accessed March 4, 2019, at https://billingsgazette.com/news/as-the-bakken-heats-up-again-williston-prepares-for-more/article_a1f87308-00c6-5e91-b560-31459c6b822b.html.

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. [Also available at https://cran.r-project.org/web/packages/segmented/.]

North Dakota Department of Mineral Resources, 2019, North Dakota drilling and production statistics: North Dakota Department of Mineral Resources web page, accessed February 15, 2019, at https://www.dmr.nd.gov/oilgas/stats/statisticsvw.asp.

North Dakota Industrial Commission, 2019, North Dakota Oil and Gas Division: North Dakota Department of Mineral Resources web page, accessed April 15, 2019, at https://www.dmr.nd.gov/oilgas/.

North Dakota State Water Commission, 2015, North Dakota State Water Commission map services: North Dakota State Water Commission web page, accessed April 15, 2019, at http://www.swc.nd.gov/info_edu/map_data_resources/mapservices.html.

PRISM Climate Group, 2019, PRISM climate data: Oregon State University web page, accessed July 9, 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. [Also available at https://www.R-project.org/.]

Schmoker, J.W., 2005, U.S. Geological Survey assessment concepts for continuous petroleum accumulations, chap. 13 *of* USGS Southwestern Wyoming Province Assessment Team, comps., 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 on CD-ROM and at https://pubs.usgs.gov/dds/dds-069/dds-069-d/REPORTS/69_D_CH_13.pdf.]

U.S. Census Bureau, 2017, QuickFacts—United States: U.S. Census Bureau web page, accessed February 15, 2019, at https://www.census.gov/quickfacts/fact/.

U.S. Department of Energy, 2019, Shale research and development: U.S. Department of Energy web page, accessed April 1, 2019, at https://www.energy.gov/fe/science-innovation/oil-gas-research/shale-gas-rd.

U.S. Energy Information Administration, 2016, Summary maps—Shale gas and oil plays, lower 48 states: U.S. Energy Information Administration web page, accessed February 15, 2019, at https://www.eia.gov/maps/maps.htm.

U.S. Energy Information Administration, 2018, August 2018 monthly energy review: Washington, D.C., Office of Energy Statistics, U.S. Energy Information Administration, 244 p., [Also available at https://www.eia.gov/TOTALENERGY/DATA/MONTHLY/archive/00351808.pdf.]

U.S. Environmental Protection Agency, 2017, Drinking water requirements for states and public water systems—Information about public water systems: U.S. Environmental Protection Agency web page, accessed March 29, 2019, at https://www.epa.gov/dwreginfo/information-about-public-water-systems.

U.S. Geological Survey Energy Resources Program, 2018, Assessment summary maps/tables: national oil and gas assessment: U.S. Geological Survey web page, accessed March 11, 2019, at https://www.usgs.gov/centers/cersc/science/united-states-assessments-undiscovered-oil-and-gas-resources?qt-science_center_objects=9#qt-science_center_objects.

U.S. Geological Survey National Assessment of Oil and Gas Resources Team and Biewick, L.R.H., comp., 2014, Map of assessed tight-gas resources in the United States, 2014: U.S. Geological Survey Digital Data Series 69–HH, 6 p., 1 pl., GIS data package, accessed March 15, 2019, at https://dx.doi.org/10.3133/ds69HH.

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., accessed February 27, 2019, at https://doi.org/10.3133/sir20185027.

Appendix. R Script

A zipped archive, COGWaterUseTool.zip, contains the following:

- Files—README.txt; run_analysis.R (the main script, which is used for running the analysis); and munge_data_release.R (a script for preparing data from the accompanying data release).
- Folders—Analysis, Data, Functions, Plots, and Raw. The Functions folder contains scripts wrangle.R, model.R, and visualize.R (additional scripts containing functions that are called in the main script to run the analysis).The Raw folder will need to be populated by the user with applicable datasets. The Analysis, Data, and Plots folders will be populated when the scripts are run.

The zipped archive can be downloaded at https://doi.org/10.3133/sir20195100. Although these data have been processed successfully on a computer system at the U.S. Geological Survey, no warranty expressed or implied is made regarding the display or utility of the data for other purposes, nor on all computer systems, nor shall the act of distribution constitute any such warranty. The U.S. Geological Survey or the U.S. Government shall not be held liable for improper or incorrect use of the data described and/or contained herein. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Aerial photograph of Bakken Formation well in North Dakota. Photograph by Vern Whitten Photography, used with permission.

For more information about this publication, contact:
Director, USGS Dakota Water Science Center
821 East Interstate Avenue, Bismarck, ND 58503
1608 Mountain View Road, Rapid City, SD 57702
605–394–3200

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

Publishing support provided by the
Rolla Publishing Service Center

USGS

Valder and others—**Analytical Framework to Estimate Water Use Associated with Continuous Oil and Gas Development**—SIR 2019–5100

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20195100