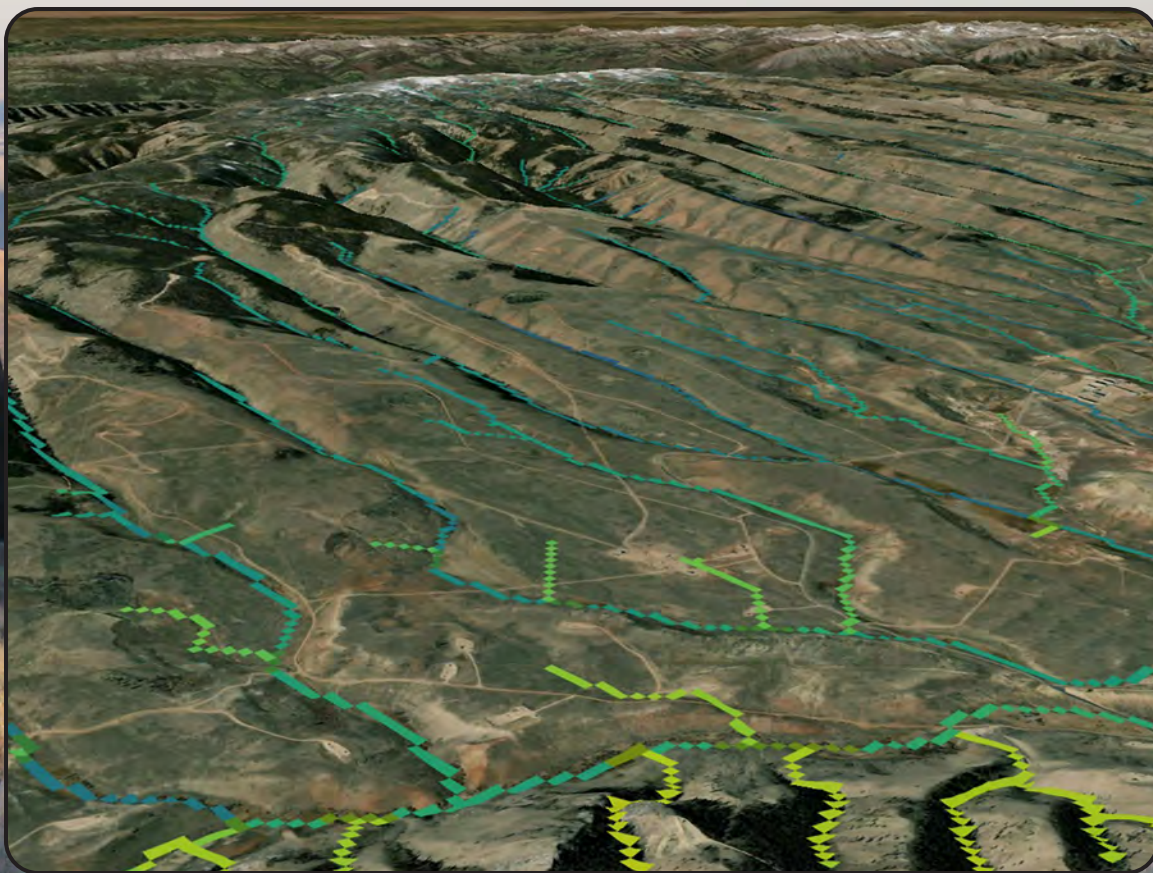


A Machine Learning Approach to Modeling Streamflow with Sparse Data in Ungaged Watersheds on the Wyoming Range, Wyoming, 2012–17



Scientific Investigations Report 2021–5093

Covers:

Front: Image showing a watershed on the Wyoming Range with gridded monthly streamflow predictions in May 2017 using a machine learning streamflow model. Base-map imagery is the intellectual property of Esri and is used herein under license. Copyright © 2021 Esri and its licensors. All rights reserved.

Back: Photograph showing a streamflow sampling site on a small stream on the Wyoming Range, taken near Big Piney, Wyoming, on July 6, 2017, by Cheryl Eddy-Miller, U.S. Geological Survey.

Background: Photograph showing the watershed of a small stream on the Wyoming Range, taken near Big Piney, Wyoming, on November 8, 2015, by Cheryl Eddy-Miller, U.S. Geological Survey.

A Machine Learning Approach to Modeling Streamflow with Sparse Data in Ungaged Watersheds on the Wyoming Range, Wyoming, 2012–17

By Ryan R. McShane and Cheryl A. Eddy-Miller

Scientific Investigations Report 2021–5093

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

U.S. Geological Survey, Reston, Virginia: 2021

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/>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

McShane, R.R., and Eddy-Miller, C.A., 2021, A machine learning approach to modeling streamflow with sparse data in ungaged watersheds on the Wyoming Range, Wyoming, 2012–17: U.S. Geological Survey Scientific Investigations Report 2021–5093, 29 p., <https://doi.org/10.3133/sir20215093>.

Associated data for this publication:

McShane, R.R., and Eddy-Miller, C.A., 2021, Input data, model output, and R scripts for a machine learning streamflow model on the Wyoming Range, Wyoming, 2012–17: U.S. Geological Survey data release, <https://doi.org/10.5066/P9XCP1AE>.

U.S. Geological Survey, 2021, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, <https://doi.org/10.5066/F7P55KJN>.

ISSN 2328-0328 (online)

Acknowledgments

Max Kuhn is thanked for creating the “caret” package in R, which greatly simplified calibration and validation of the machine learning streamflow model. Many universities and State and Federal agencies are greatly appreciated for producing time series and geospatial data for use in modeling streamflow at fine resolution and broad extent.

Annika Walters, with the Wyoming Cooperative Fish and Wildlife Research Unit (University of Wyoming and U.S. Geological Survey), and her graduate students, Richard Walker and Carlin Gerard, were instrumental in providing streamflow observations used to fit the machine learning streamflow model. Several U.S. Geological Survey colleagues provided reviews that improved the report and accompanying data release.

Contents

Acknowledgments	iii
Abstract	1
Introduction	1
Description of Study Area	2
Purpose and Scope	4
Methods for Machine Learning Approach to Modeling Streamflow	4
Streamflow Observations	4
Predictor Variables	7
Processes of Temporal and Spatial Conditioning	7
Evaluation of Predictor Variable Conditioning	10
Development of Machine Learning Streamflow Model	11
Evaluation of Predictive Performance and Variable Importance	12
Explanation and Prediction of Streamflow	12
Assessment of Model Sensitivity	12
Results of Machine Learning Approach to Modeling Streamflow	13
Predictive Performance	13
Variable Importance	15
Model Sensitivity	18
Streamflow Predictions	21
Utility and Limitations of Modeling Approach	23
Summary	25
References Cited	26

Figures

1. Map showing the study area on the Wyoming Range, including the watersheds, channel networks, and sampling sites used in modeling monthly streamflow and the U.S. Geological Survey streamgage closest to the study area3
2. Graphs showing observed streamflow against drainage area of sampling sites by watershed and year for each month sampled during 2012–176
3. Diagrams showing an example of a dynamic climatic variable, snow water equivalent, shown before and after temporal conditioning with a moving average of cell values10
4. Diagrams showing an example of a static physiographic variable, forest cover, shown before and after spatial conditioning with an area-averaged accumulation of cell values11
5. Diagrams showing an example of a static physiographic variable, number of surficial geology contacts, shown before and after spatial conditioning with a distance-decayed accumulation of cell values.....11
6. Graph showing relation between observed and predicted streamflow13
7. Boxplots showing distribution of observed and predicted streamflow and residual values for each month sampled during 2012–1714
8. Graphs showing observed and predicted streamflow at the three sites with the most streamflow observations15

9.	Graph showing relative importance of the top 20 variables used in the model fitted to all data	15
10.	Graphs showing partial dependence of streamflow on the top 20 most important variables used in the model fitted to all data	16
11.	Graphs showing relative importance of the top five variables used in models fitted to data only from each watershed.....	17
12.	Graphs showing relative importance of the top five variables used in models fitted to data only from each year	17
13.	Diagrams showing intra-annual variation in streamflow predictions on part of a channel network for each month of 2017	22
14.	Diagrams showing interannual variation in streamflow predictions on part of a channel network for each year in August.....	23

Tables

1.	Number of sites and months sampled by watershed and year and spatial and temporal distribution among watersheds and years of number of streamflow observations by month sampled during 2012–17.....	5
2.	Summary statistics of streamflow observations by watershed and year.....	6
3.	Static physiographic and anthropogenic variables and dynamic climatic variables used in the machine learning streamflow model.....	8
4.	Distribution of goodness-of-fit metrics from repeated cross-validation of the machine learning streamflow model, fitted to all streamflow observations, computed for each site independently	14
5.	Relation of streamflow to some dynamic and static variables before and after temporal or spatial conditioning	19
6.	Goodness-of-fit metrics for models fitted to different combinations of static and dynamic variables before or after spatial or temporal conditioning, indicating qualitative model sensitivity to selection and conditioning of the predictor variables.....	19
7.	Goodness-of-fit metrics for models trained with observations from all watersheds and years and tested on observations from each watershed or year separately or for models trained with observations from all except one watershed or 1 year left out sequentially and tested on observations from the left-out watershed or year, indicating qualitative model sensitivity to data used from a different area or period	20
8.	Goodness-of-fit metrics for the model fitted to all observations or for models fitted to observations progressively reduced by percentage of sites or months available, indicating qualitative model sensitivity to quantity of the data used.....	21

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
Length		
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
square mile (mi ²)	2.590	square kilometer (km ²)
Volume		
cubic foot (ft ³)	0.02832	cubic meter (m ³)
Flow rate		
cubic foot per second (ft ³ /s)	0.02832	cubic meter per second (m ³ /s)

International System of Units to U.S. customary units

Multiply	By	To obtain
Length		
centimeter (cm)	0.3937	inch (in.)
millimeter (mm)	0.03937	inch (in.)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
Area		
square kilometer (km ²)	0.3861	square mile (mi ²)
Volume		
cubic meter (m ³)	35.31	cubic foot (ft ³)
Flow rate		
cubic meter per second (m ³ /s)	35.31	cubic foot per second (ft ³ /s)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$^{\circ}\text{F} = (1.8 \times ^{\circ}\text{C}) + 32.$$

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

$$^{\circ}\text{C} = (^{\circ}\text{F} - 32) / 1.8.$$

Datum

Vertical coordinate information is referenced to the North American Vertical Datum of 1988, unless noted otherwise.

Horizontal coordinate information is referenced to the North American Datum of 1983.

Elevation, as used in this report, refers to distance above the vertical datum.

Abbreviations

DEM	digital elevation model
HUC12	12-digit hydrologic unit code
logNSE	Nash-Sutcliffe efficiency with log-transformed data
MAD	median absolute deviation
MLFLOW	machine learning streamflow
NSE	Nash-Sutcliffe efficiency
PBIAS	percent bias
r	Pearson correlation coefficient
R^2	coefficient of determination
USGS	U.S. Geological Survey
WLCI	Wyoming Landscape Conservation Initiative

A Machine Learning Approach to Modeling Streamflow with Sparse Data in Ungaged Watersheds on the Wyoming Range, Wyoming, 2012–17

By Ryan R. McShane and Cheryl A. Eddy-Miller

Abstract

Scant availability of streamflow data can impede the utility of streamflow as a variable in ecological models of aquatic and terrestrial species, especially when studying small streams in watersheds that lack streamgages. Streamflow data at fine resolution and broad extent were needed by collaborators for ecological research on small streams in several ungaged watersheds of southwestern Wyoming, where streamflow data are sparse.

To improve the utility of sparse streamflow data to ecological research in ungaged watersheds, we developed a machine learning approach in R for modeling spatially and temporally continuous monthly streamflow from 2012 through 2017 in three semiarid montane-steppe watersheds (with drainage areas of 26–55 square miles and mean elevations of 8,031–8,455 feet) on the Wyoming Range in the upper Green River Basin. A machine learning streamflow (MLFLOW) model was calibrated and validated with 971 discrete streamflow observations and 24 static and dynamic predictor variables derived from geospatial and time series data on climatic, physiographic, and anthropogenic characteristics affecting streamflow. The predictor variables were temporally and spatially conditioned to amplify the relation of predictor variables to monthly streamflow.

The MLFLOW model had satisfactory agreement between observed and predicted streamflow (coefficient of determination [R^2]=0.80, Nash-Sutcliffe efficiency [NSE]=0.79, NSE with log-transformed data [logNSE]=0.82, and percent bias [PBIAS]=0.7 percent). NSE and logNSE indicated the MLFLOW model performed equally well for high and low flows, and PBIAS indicated the MLFLOW model did not overpredict or underpredict monthly streamflow. Streamflow predictions seemed to well represent the annual hydrograph within the study area during the study period.

The most important variables (statistically important in the MLFLOW model) for explaining monthly streamflow were temporally and spatially conditioned dynamic climatic variables, mostly precipitation and snow water equivalent. Importance of the static and dynamic variables did not differ substantially among the three watersheds but

differed considerably among the 6 years. Monthly streamflow increased with increasing precipitation, snow water equivalent, and drainage area but decreased with increasing forest cover, elevation, evapotranspiration, and temperature.

The MLFLOW model was most sensitive to selection of dynamic climatic variables. Unconditioned dynamic climatic variables alone explained 54 percent of the variance (R^2 =0.54) in monthly streamflow, whereas adding static physiographic and anthropogenic variables only explained 12 percent more of the variance (R^2 =0.66). Also, spatial conditioning of all variables together with temporal conditioning of dynamic variables increased the variance explained in the MLFLOW model by another 14 percent (R^2 =0.80). The MLFLOW model also had greater sensitivity to temporal than to spatial differences in the data. For the MLFLOW model trained with observations from all watersheds and years or for models trained with observations from all except one watershed or 1 year left out sequentially, performance was better in testing on observations from each watershed than from each year separately. Also, performance was better for models fitted to fewer sites than to fewer months of observations.

The greatest utility of the modeling approach is the ease of use and the speed of processing input data, running the model, and interpreting the model output, whereas the greatest limitation is the need for spatially and temporally representative streamflow observations to drive the model. Although familiarity with R is necessary, only a working knowledge of hydrology (for selecting appropriate predictor variables and evaluating the quality of streamflow observations) and a rudimentary understanding of machine learning models are needed. Therefore, this modeling approach is practicable for other scientists who work with water but who are not hydrologists.

Introduction

Streamflow is a necessary ecological resource for many animal and plant species. However, scant availability of streamflow data can impede the utility of streamflow as a variable in ecological models of aquatic and terrestrial species,

especially when studying small streams (low stream order or low flow, or both) in watersheds that lack streamgages. Much ecological research on small streams is concentrated on species-habitat relations throughout the channel network of a watershed, but streamflow data, needed at fine resolution and broad extent for this research, are typically sparse. For instance, collaborators with the Wyoming Cooperative Fish and Wildlife Research Unit (University of Wyoming and U.S. Geological Survey [USGS]) needed more detailed streamflow information for researching the effects of multiple stressors (including low flows) on fish and invertebrates in several ungaged watersheds in the upper Green River Basin (not shown) in southwestern Wyoming (fig. 1), where streamflow data are sparse (Girard and Walters, 2018; Walters and others, 2019; Walker and others, 2020).

Several approaches to modeling streamflow at various scales have been developed that may improve the utility of sparse streamflow data to ecological research in small streams. One approach is a water-balance model that estimates various elements of the hydrologic cycle (including runoff generation), such as the USGS Thornthwaite water-balance model (McCabe and Markstrom, 2007). Another approach is a physically based hydrologic (rainfall-runoff) model that simulates water and energy fluxes between the atmosphere and the land surface, such as the USGS Precipitation-Runoff Modeling System (Leavesley and others, 1983; Markstrom and others, 2015). However, these two approaches require streamflow data of adequate quality and in sufficient quantity, typically from USGS streamgages, that may not be available for the area or period of interest, and the approaches require expertise in hydrologic modeling that may be impractical for a scientist without much knowledge of hydrology or experience using complicated applications for modeling streamflow.

A third streamflow modeling approach is a machine learning model that can fit potentially complex relations between streamflow observations and environmental predictor variables. Machine learning models are increasingly used in the water sciences in general (Shen and others, 2018) and specifically for modeling streamflow (Bellos and Carbajal, 2020). However, many applications of machine learning to streamflow modeling are reliant on streamflow data from streamgages for use in model calibration and validation. These data are typically more available for larger, higher-order streams, and calibration and validation of models on low-order streams has been uncommon. For example, Miller and others (2018) used a random forest model fitted to USGS streamgage data (with few data from low-order streams) to predict monthly streamflow for the conterminous United States. However, in contrast to this reliance on streamgage data, Jaeger and others (2019) used a random forest model fitted to thousands of observations of wet or dry stream conditions to predict the annual probability of a channel maintaining streamflow throughout the year in the Columbia River Basin (not shown).

Of the three approaches to modeling streamflow, machine learning models may be most adaptable to the needs of ecological researchers working in ungaged watersheds and may have fewer initial learning impediments to developing a working streamflow model. Moreover, using a more complex water-balance model or physically based hydrologic model, even if the requisite data are available, does not necessarily result in better streamflow predictions than using a simpler machine learning model. Therefore, we determined that a machine learning model was the most appropriate streamflow modeling approach for this study.

The primary objective of this study was to develop an approach to modeling streamflow with sparse data in ungaged watersheds in southwestern Wyoming, predicting spatially and temporally continuous monthly streamflow for all the study area and study period (fig. 1). A secondary objective was to explain the environmental drivers of monthly streamflow within the study area during the study period. We accomplished these two objectives by developing a machine learning streamflow (MLFLOW) model calibrated and validated with discrete streamflow observations and static and dynamic predictor variables derived from geospatial and time series data on climatic, physiographic, and anthropogenic characteristics affecting streamflow. The resulting model predictions and explanation of streamflow are applied at the reach scale (tens of meters), of use particularly on small streams.

This study also is a contribution to the Wyoming Landscape Conservation Initiative (WLCI), a program with the mission “to implement a long-term, science-based program of assessing, conserving, and enhancing fish and wildlife habitats while facilitating responsible energy and other development through local collaboration and partnerships” (Bowen and others, 2014, p. 2). Much of the WLCI area in southwestern Wyoming consists of small streams in watersheds without streamgages, which complicates management of species dependent on these streams. Therefore, the modeling approach developed in this study can support species management decisions throughout the WLCI area.

Description of Study Area

The channel networks of 3 watersheds and 125 streamflow sampling sites on the Wyoming Range in southwestern Wyoming (fig. 1) were used for modeling monthly streamflow. The three watersheds are identified by the following USGS 12-digit hydrologic unit codes (HUC12; U.S. Geological Survey, 2016): 140401010907 (Lower South Piney Creek), 140401011102 (North Fork Dry Piney Creek), and 140401011103 (Dry Piney Creek). In this report, these three HUC12s are referred to as watersheds 1–3, respectively. The three watersheds are in the western WLCI area, which covers much of southwestern Wyoming (fig. 1).

The three watersheds are in a sparsely populated region and have many similar physical properties. Watersheds 1 and 2 are similar in drainage area (34 and 26 square miles

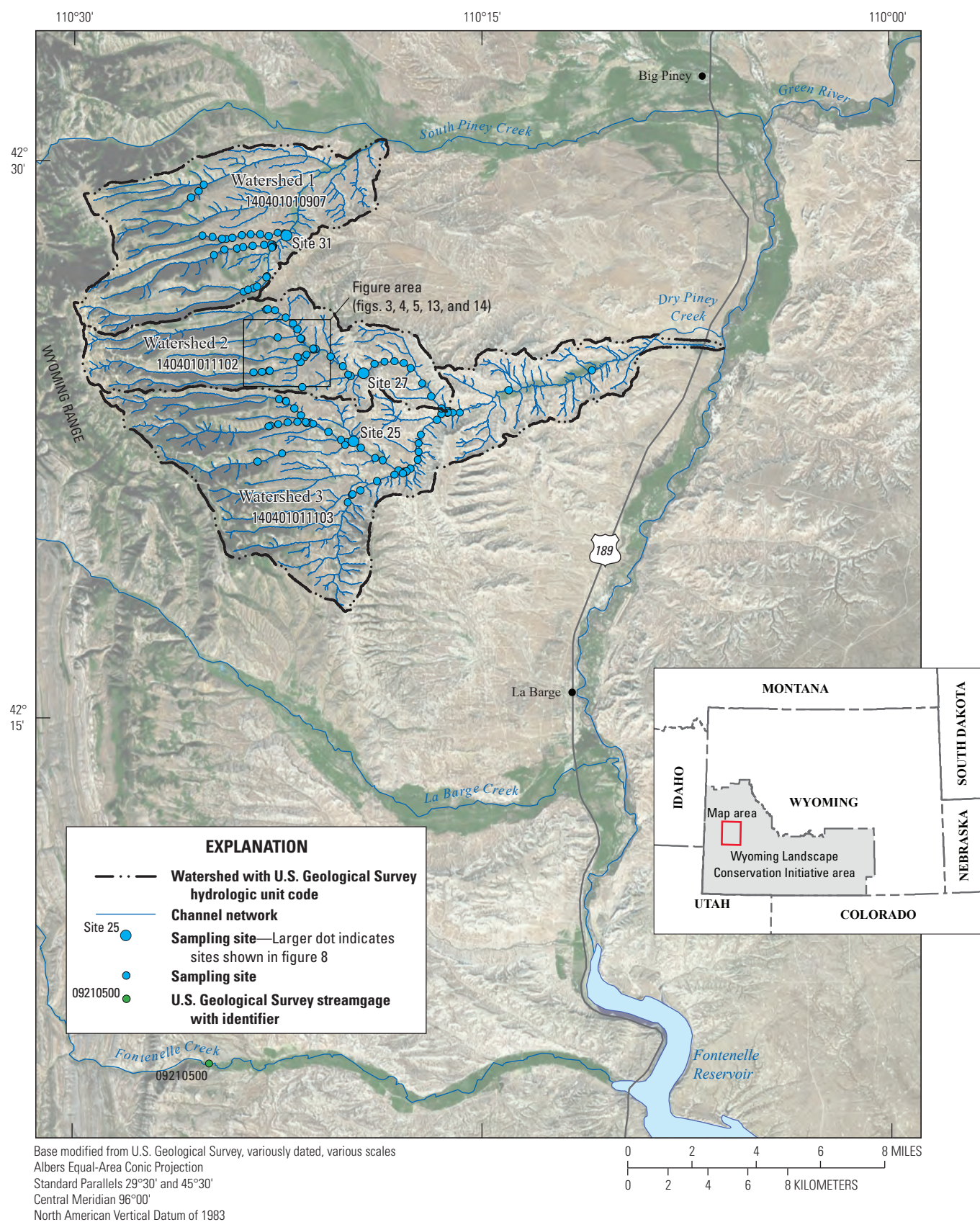


Figure 1. The study area on the Wyoming Range, including the watersheds, channel networks, and sampling sites used in modeling monthly streamflow and the U.S. Geological Survey streamgage closest to the study area.

[mi²], respectively) and mean elevation (8,428 and 8,455 feet, respectively), whereas watershed 3 is larger (55 mi²) and lower (8,031 feet) (U.S. Geological Survey, 2019a). The watersheds have a similar semiarid climate characterized by warm summers with precipitation from occasional thunderstorms and by cold winters with snow typically covering the montane-steppe landscape from November through March. Watersheds 1 and 2 have a mean annual temperature of 38.1 degrees Fahrenheit and a total annual precipitation of 10.6 inches, whereas watershed 3 is cooler (37.2 degrees Fahrenheit) and wetter (13.7 inches) (PRISM Climate Group, 2018). Geology of the watersheds is mostly sedimentary rocks of Quaternary and Tertiary age (Oriel and Platt, 1980). The watersheds are on the eastern Overthrust Belt (not shown), and the surficial units in the upper elevations and Cretaceous Mountain (not shown; to the east of watersheds 1 and 2) are much older, dating to the Cambrian. Additionally, the Overthrust Belt has many surficial and buried faults that transmit groundwater into the watersheds through springs and seeps (Zielinski and others, 1985). Land cover in the watersheds is mostly shrub/scrub (66–72 percent) and forest (20–25 percent), and irrigated land use ranges from about 0.5 to 2.5 percent (Multi-Resolution Land Characteristics Consortium, 2017). Oil and gas development began in the region in the early 1900s and is ongoing. The energy development footprint increases from watershed 1 to 3 (Walters and others, 2019). Total length of streams in the three watersheds is 372 miles, and about 75 percent (277 miles) consisted of first- and second-order streams (U.S. Geological Survey, 2019a). Annual hydrographs are characterized by snowmelt high flows in late spring and early summer (May and June) followed by low flows during summer through the following spring; the lowest flows are in fall and winter.

The only USGS streamgage close to the study area and operating during the study period (2012–17) was on Fontenelle Creek near Herschler Ranch, near Fontenelle, Wyoming (USGS 09210500; [fig. 1](#)), which has a drainage area of 152 mi² and an elevation of 6,950 feet (above the National Geodetic Vertical Datum of 1929; U.S. Geological Survey, 2021). The watershed monitored by this streamgage has a larger drainage area and lower mean elevation than all three watersheds of the study area. This difference in drainage area and elevation and the distance of this streamgage from the study area made the streamgage unfit for calibrating and validating a streamflow model.

Purpose and Scope

A machine learning approach was developed in R (R Core Team, 2021) for modeling monthly streamflow from 2012 through 2017 in three watersheds on the Wyoming Range. This study was intended to investigate the utility and limitations of applying a machine learning approach to modeling streamflow with sparse data (for example, streamflow measurements made at different times and places instead of a

time series of streamflow observations made at a streamgage). This study was not intended to investigate performance differences among various machine learning models used for modeling streamflow. Methodology of the modeling approach is described, including temporal and spatial conditioning processes applied to the input data and calibration and validation of the MLFLOW model. Results of the MLFLOW model are presented with discussion of the model output, including predictive performance at the sampling sites (with comparisons to other streamflow modeling studies), variable importance for explaining monthly streamflow, sensitivity to the predictor variables and streamflow observations used in fitting the model, and monthly streamflow predictions on the channel networks. The input data, model output, and R scripts for the MLFLOW model are available in an accompanying data release (McShane and Eddy-Miller, 2021).

Methods for Machine Learning Approach to Modeling Streamflow

The MLFLOW model was developed to predict monthly streamflow using static and dynamic variables derived from geospatial and time series data and to explain the multidimensional relation of monthly streamflow to the predictor variables. The following sections describe characteristics of the input data (streamflow observations and predictor variables) and procedures used to process data temporally and spatially for use in the MLFLOW model, development of the MLFLOW model, evaluation of the model output for predictive performance and variable importance, and assessment of model sensitivity to how the data were used in fitting the MLFLOW model. All data preparation, streamflow modeling, and model analysis were done by using geoprocessing tools in ArcGIS (Esri, 2019) and TauDEM (Tarboton, 2016) and by using several packages in R (R Core Team, 2021).

Streamflow Observations

No USGS streamgages, with time series of streamflow, were available in the study area for use in developing the MLFLOW model. One nearby streamgage (less than 20 miles from the study area) was operational during the study period (2012–17), but the streamgage was on a river with a larger drainage area and lower mean elevation than of the three watersheds of the study area. Therefore, this study relied on available discrete streamflow observations to calibrate and validate the MLFLOW model.

Streamflow measurements were made at 125 sites ([fig. 1](#)) in 35 months during 2012–17, totaling 971 discrete observations ([table 1](#)); only a single measurement was made at any site in any month. These observations were not made on a consistent spatiotemporal basis but together are representative of the annual hydrograph on a monthly timestep, including

high flows in late spring and early summer (fig. 2B–E) and low flows from late summer through early spring (fig. 2F–H, A). Therefore, the observations were assumed to qualitatively represent monthly streamflow. Ecologists with the Wyoming Cooperative Fish and Wildlife Research Unit (University of Wyoming and USGS) made 763 observations at 101 sites in 30 months during 2012–17 for various research on aquatic species distribution and abundance. Most (736) of the observations were made during May through August of 2012–17, and the remainder of the observations were made during April of 2012 and 2017 and September of 2013–16. USGS hydrologists made 208 observations at 24 sites in 10 months during 2015–17 for research on groundwater/surface-water relations. Most (184) of the observations were made during June, August, and November of 2015 and 2016 and July and September of 2017; the remainder of the observations were made during April 2016 and February 2017. Streamflow measurements were made following standard USGS techniques and methods for streamflow measurement and computation using current-velocity meters, portable flumes, or pressure transducers (Rantz and others, 1982; Sauer and Turnipseed, 2010). The monthly streamflow observation data are available in the accompanying data release (McShane and Eddy-Miller, 2021).

The number of sites where streamflow measurements were made was different among the three watersheds and among the 6 years of sampling, whereas the number of months when streamflow measurements were made was similar among watersheds and among years. On average, sampling involved 42 sites and 33 months per watershed and 88 sites and 6 months per year (table 1). No sites were sampled in every month, but every site was sampled in at least 2 months, and of the 35 months sampled during 2012–17, sites were sampled in 8 months on average (McShane and Eddy-Miller, 2021). The sites sampled in the greatest number of months were sites 25, 27, and 31 (fig. 1), which were sampled in 28, 25,

and 28 months, respectively. Additionally, the watershed (3) or year (2013) with the most streamflow observations had about 100 more observations than the watershed (1) or year (2014) with the fewest streamflow observations (table 2).

Observed flows at the sampling sites ranged from 0 to 19.67 cubic feet per second (ft³/s) and averaged 2.39 ft³/s (table 2), and drainage area of the sampling sites ranged from 0.64 to 78.1 mi² and averaged 14.3 mi² (fig. 2). A robust streamflow/drainage-area relation among the sampling sites was not readily apparent in every month sampled (fig. 2A–H), most likely because sites with different drainage areas were sampled in different months, confounding the relation between streamflow and drainage area with the relation between streamflow and month. Sites with the largest drainage areas (greater than 67 mi²) were sampled only in watershed 3 and only during summer (June through August, fig. 2D–F). Furthermore, sites with smaller drainage areas (less than 41 mi²) were sampled in every month (fig. 2A–H) but only in 1 or 2 years in February, April, and November (table 1; fig. 2A, B, H). Additionally, in some months (for example, May, fig. 2C), the streamflow-drainage area relation seemed to be more robust by watershed, where streamflow increased with increasing drainage area, than by year, when streamflow decreased with drainage area in some years but increased in others.

The distribution in values of streamflow observations varied more among the 6 years of sampling than among the three watersheds. For example, mean streamflow differed by only 31 percent among watersheds but by 740 percent among years (table 2). In general, streamflow values were low—median streamflow ranged from 0.4 to 4.01 ft³/s among watersheds and years; furthermore, two watersheds and 3 years had zero-flow observations (table 2). In addition, climatically, 2012 and 2013 were dry years, whereas 2017 was a wet year. These 2 dry years and 1 wet year had very low and very high mean streamflow, respectively, compared with the other 3 years (2014–16, table 2).

Table 1. Number of sites and months sampled by watershed and year and spatial and temporal distribution among watersheds and years of number of streamflow observations by month sampled during 2012–17.

[--, not sampled]

Watershed or year	Number of sites sampled	Number of months sampled	Number of streamflow observations							
			February	April	May	June	July	August	September	November
Watershed 1	37	32	3	1	16	87	96	69	20	22
Watershed 2	37	33	3	5	9	80	91	73	11	9
Watershed 3	51	34	7	8	11	131	119	69	16	15
2012	80	5	--	1	5	45	40	79	--	--
2013	95	5	--	--	5	74	92	28	3	--
2014	71	5	--	--	7	19	55	11	6	--
2015	95	6	--	--	8	53	45	35	8	22
2016	95	7	--	11	7	83	19	31	7	24
2017	94	7	13	2	4	24	55	27	23	--

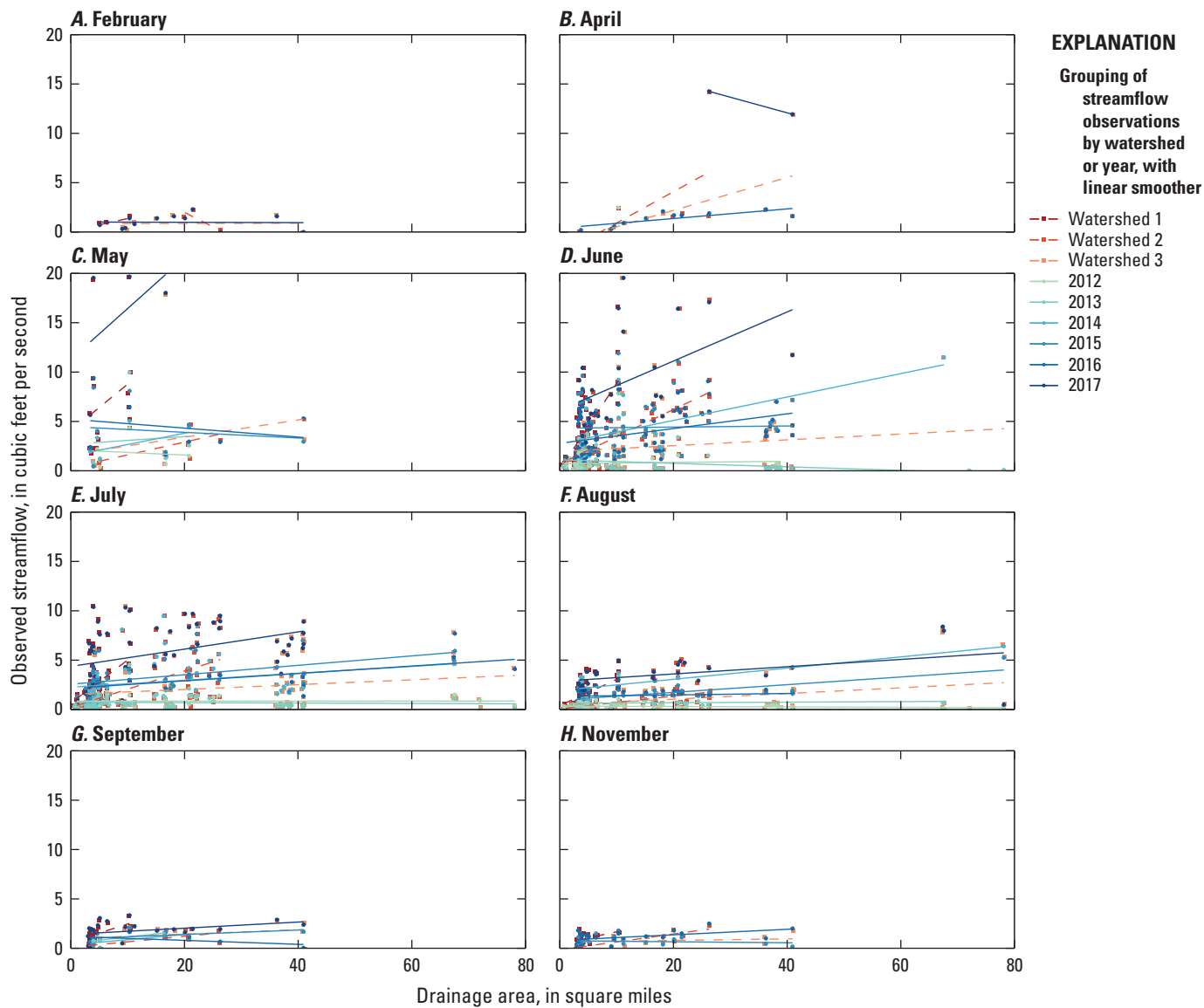


Figure 2. Observed streamflow against drainage area of sampling sites by watershed and year for each month sampled during 2012–17. A, February; B, April; C, May; D, June; E, July; F, August; G, September; and H, November.

Table 2. Summary statistics of streamflow observations by watershed and year.

Watershed or year	Number of streamflow observations	Summary statistics of streamflow observations (cubic foot per second)			
		Minimum	Median	Mean	Maximum
Watershed 1	314	0.05	1.81	2.68	19.67
Watershed 2	281	0	1.40	2.53	17.08
Watershed 3	376	0	1.13	2.05	19.54
2012	170	0	0.40	0.62	4.34
2013	202	0	0.46	0.84	9.97
2014	98	0.05	2.32	2.97	11.49
2015	171	0.13	2.01	2.89	11.90
2016	182	0.04	1.98	2.72	9.39
2017	148	0	4.01	5.21	19.67

Predictor Variables

Based on the potential to affect streamflow, 24 variables describing physiographic, anthropogenic, and climatic characteristics were chosen as predictors in the MLFLOW model—20 variables described static physiographic and anthropogenic conditions, and 4 variables described dynamic climatic conditions (table 3). Vector data (surficial geology permeability index and number of surficial geology contacts, bedrock geology faults, springs, water diversions, roads, and water wells, table 3) were converted to raster data. Raster data describing topography (drainage area, elevation, slope-area ratio, and slope, table 3) were derived from the 30-meter (m) resolution National Elevation Dataset digital elevation model (DEM) in the USGS National Hydrography Dataset (U.S. Geological Survey, 2019a). All variables were snapped to the 30-m DEM extent. Any variable with lower spatial resolution than the 30-m DEM resolution, including base-flow index, depth to soil restrictive layer, depth to water table, and every climate variable (table 3), was resampled, which was necessary for processing the data for use in the MLFLOW model. This resampling produced data with higher spatial resolution than the original data resolution but with minimal change in the distribution of values of the original data. Each climate variable was provided as a monthly value—total for evapotranspiration and precipitation, mean for temperature, and first day of the month for snow water equivalent. Values of each predictor variable were produced throughout the channel networks of the three watersheds, which consisted of every cell of the flow accumulation grid in the medium resolution (30 m) National Hydrography Dataset with a value of 100 or more, equivalent to a drainage area of 0.09 square kilometer or more. Data on the predictor variables are available in the accompanying data release (McShane and Eddy-Miller, 2021), including raw values of the variables and values of the variables with temporal and spatial conditioning—the conditioning processes are described in the following section.

Processes of Temporal and Spatial Conditioning

Streamflow is affected by antecedent climatic conditions. For example, precipitation during an earlier period can have an enduring effect on streamflow. Therefore, to generate a diverse lagged effect of climatic conditions on streamflow, a temporal conditioning process was applied to the climate variables. The process consisted of moving averages of the time series data that ranged from the prior month to the prior year. Data on evapotranspiration, precipitation, snow water equivalent, and temperature were temporally conditioned into five variants for each climate variable, and every variant was used in fitting the MLFLOW model—moving-average (mean) values of the current month together with the previous 1, 3, 6, 9, or 12 months (codes with suffix “01” through “12,” table 3). Data on the four climate variables without conditioning (raw values of the data; codes with suffix “00,” table 3) also were used in model

fitting. Using May 2017 as an example, variants for snow water equivalent were generated to describe, for any cell, unconditioned (raw) snow water equivalent for the current month (fig. 3A); mean snow water equivalent for the current and prior month (fig. 3B); and mean snow water equivalent for the current and prior 3, 6, 9, and 12 months (fig. 3C–F, respectively). For any cell in the resulting grids, the different conditioned variants (fig. 3B–F) could have a value greater or less than the raw value of the current month (fig. 3A), which might be more explanatory, in the MLFLOW model, of streamflow in the current month.

The relation of a predictor variable to streamflow depends on the area upslope of a channel and the distance to a channel downslope. Therefore, two geospatial processes were used to account for area and distance effects on streamflow. An area-averaged accumulation of the spatial data generated upstream effects of a variable. A distance-decayed accumulation of the spatial data generated more localized effects of a variable. Practical software tools for implementing these geospatial (flow-conditioning) processes are available in Barnhart and others (2020).

The process for area-averaged accumulation involved weighting a flow accumulation grid with values of a variable and dividing by the total area of all upslope grid cells. The resulting grid described the average value of the variable for the area upslope of each grid cell. Area-averaged conditioning was most applicable to more continuous variables, such as vegetation and climate, whose effect on streamflow was anticipated mostly from areas upstream. Spatial conditioning with an area-averaged accumulation was applied to 17 variables (codes with suffix “a,” table 3) as follows: 6 geology (soil bulk density, soil saturated hydraulic conductivity, depth to soil restrictive layer, surficial geology permeability index, soil saturated water content, and depth to water table); 1 hydrology (base-flow index); 3 topography (elevation, slope, and slope-area ratio); 3 vegetation (cropland cover, forest cover, and wetland cover); and 4 climate (evapotranspiration, precipitation, snow water equivalent, and temperature). Using forest cover as an example, the unconditioned value of forest cover for any cell was 0 or 100 percent (fig. 4A), but area-averaged conditioning produced a continuous value for every cell based on the averaged downslope accumulation of forest cover (fig. 4B). Streamflow would be less affected by the forest cover at a specific grid cell on the channel network than by the general forest cover upstream from the cell.

The process for distance-decayed accumulation involved weighting a flow accumulation grid with values of a variable and multiplying by a decay grid, which was implemented as a negative exponential function of the distance of any grid cell to the nearest channel downslope (traveling along the flow direction grid). The resulting grid described the total value of the variable for all cells upslope of each grid cell attenuated by distance. Distance-decayed conditioning was most applicable to discrete variables, such as human infrastructure or geologic structures, which were expected to affect streamflow more

Table 3. Static physiographic and anthropogenic variables and dynamic climatic variables used in the machine learning streamflow model.

[Suffixes of variable codes (following an underscore) with the following definitions: a, area-averaged spatial conditioning; d, distance-decayed spatial conditioning; 00, 01, 03, 06, 09, 12, temporal conditioning of dynamic climatic variables with a moving average of the current month (00) and prior 1, 3, 6, 9, or 12 months (01–12), respectively. DEM, digital elevation model]

Variable		Code	Source
Static physiographic and anthropogenic variables used to fit models			
Geology	Soil bulk density	bulkDens_a	Chaney and others (2019); POLARIS (2019)
	Number of surficial geology contacts	contacts_d	Wyoming State Geological Survey (2015)
	Number of bedrock geology faults	faults_d	Wyoming State Geological Survey (2014)
	Soil saturated hydraulic conductivity	hydCond_a	Chaney and others (2019); POLARIS (2019)
	Depth to soil restrictive layer	restrLayer_a	Soil Survey Staff (2016, 2017)
	Surficial geology permeability index	surfPerm_a	Stoeser and others (2005); Bartolino and others (2019)
	Soil saturated water content	waterCont_a	Chaney and others (2019); POLARIS (2019)
	Depth to water table	waterTable_a	Soil Survey Staff (2016, 2017)
Hydrology	Base-flow index	baseFlow_a	Wolock (2003)
	Number of springs	springs_d	U.S. Geological Survey (2019b)
Topography	Drainage area	area	Computed with 30-meter DEM (U.S. Geological Survey, 2019a) using TauDEM (Tarboton, 2016)
	Elevation	elevation_a	30-meter DEM (U.S. Geological Survey, 2019a)
	Slope	slope_a	Computed with 30-meter DEM (U.S. Geological Survey, 2019a) using TauDEM (Tarboton, 2016)
	Slope-area ratio	slopeArea_a	Computed with 30-meter DEM (U.S. Geological Survey, 2019a) using TauDEM (Tarboton, 2016)
Vegetation	Cropland cover	cropland_a	Massey and others (2017)
	Forest cover	forest_a	Homer and others (2015); Multi-Resolution Land Characteristics Consortium (2017)
	Wetland cover	wetland_a	Homer and others (2015); Multi-Resolution Land Characteristics Consortium (2017)
Human	Number of water diversions	diversions_d	Wyoming Water Development Office (2007)
	Number of roads	roads_d	O'Donnell and others (2014)
	Number of water wells	wells_d	Wyoming State Engineer's Office (2016)

Table 3. Static physiographic and anthropogenic variables and dynamic climatic variables used in the machine learning streamflow model.—Continued

[Suffixes of variable codes (following an underscore) with the following definitions: a, area-averaged spatial conditioning; d, distance-decayed spatial conditioning; 00, 01, 03, 06, 09, 12, temporal conditioning of dynamic climatic variables with a moving average of the current month (00) and prior 1, 3, 6, 9, or 12 months (01–12), respectively. DEM, digital elevation model]

Variable		Code	Source
Dynamic climatic variables used to fit models			
Climate	Evapotranspiration	eta_00_a, eta_01_a, eta_03_a, eta_06_a, eta_09_a, eta_12_a	Senay and others (2013); U.S. Geological Survey (2018)
	Precipitation	ppt_00_a, ppt_01_a, ppt_03_a, ppt_06_a, ppt_09_a, ppt_12_a	Daly and others (2008); PRISM Climate Group (2018)
	Snow water equivalent	swe_00_a, swe_01_a, swe_03_a, swe_06_a, swe_09_a, swe_12_a	Barrett (2003); National Operational Hydrologic Remote Sensing Center (2004)
	Temperature	tmp_00_a, tmp_01_a, tmp_03_a, tmp_06_a, tmp_09_a, tmp_12_a	Daly and others (2008); PRISM Climate Group (2018)

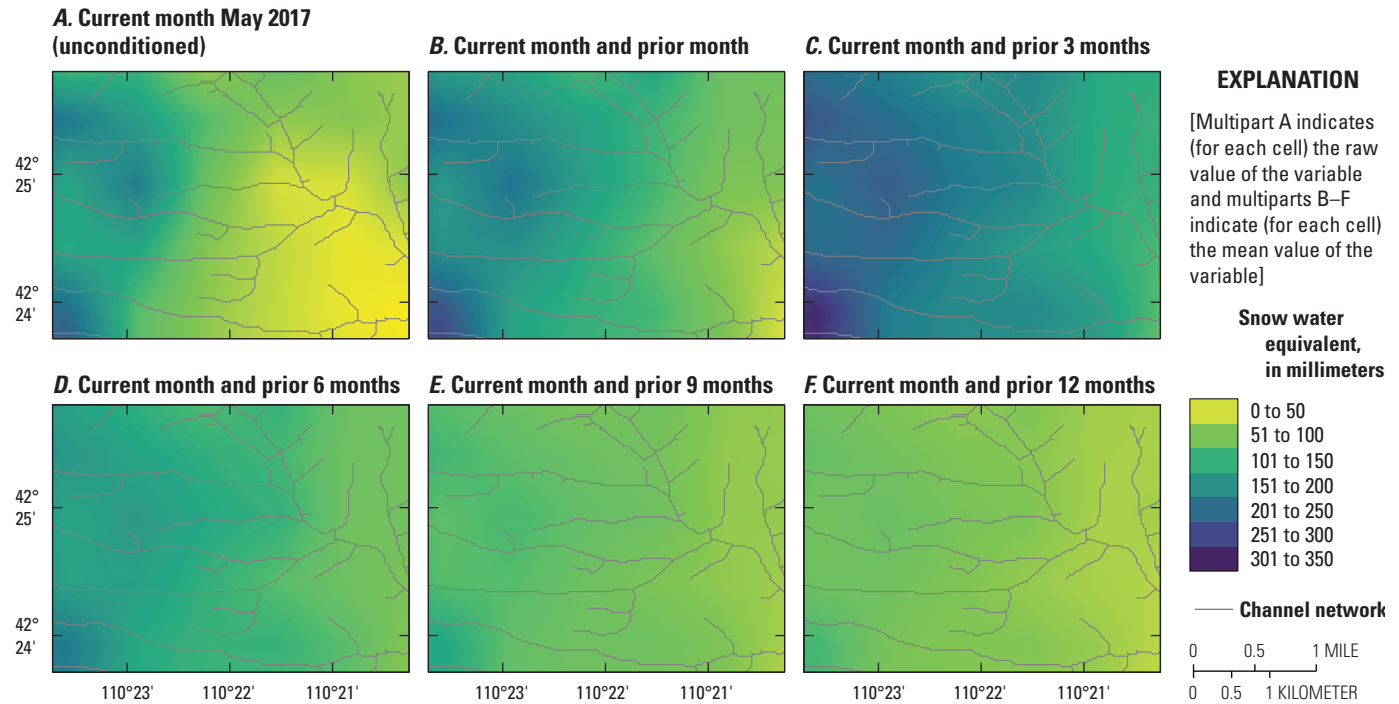


Figure 3. An example of a dynamic climatic variable, snow water equivalent, shown before and after temporal conditioning with a moving average of cell values (figure area, [fig. 1](#)). *A*, current month May 2017; *B*, current and prior month; *C*, current and prior 3 months; *D*, current and prior 6 months; *E*, current and prior 9 months; and *F*, current and prior 12 months.

in proximate channels. Spatial conditioning with a distance-decayed accumulation was applied to six variables (codes with suffix “d,” [table 3](#)) as follows: two geology (number of surficial geology contacts and number of bedrock geology faults); one hydrology (number of springs); and three human (number of water diversions, number of roads, and number of water wells). Using surficial geology contacts as an example, the unconditioned value of a geologic contact for any cell was 0 or 1 ([fig. 5A](#)), but distance-decayed conditioning produced a continuous value for every cell based on the decayed downslope accumulation of geologic contacts ([fig. 5B](#)). Streamflow would be most affected in a channel closest to the geologic contact, decreasing in effect farther downstream; moreover, a geologic contact would have less effect on streamflow as upslope distance from a channel increased.

Evaluation of Predictor Variable Conditioning

The dynamic and static variables were temporally and spatially conditioned to amplify the relation of the predictor variables to the response variable (monthly streamflow). Pearson correlation coefficient (r) was used to evaluate the effect of temporal and spatial conditioning on the predictor-response relation. Values of r were compared between a variable without conditioning and the equivalent variable with temporal or spatial conditioning (for example, unconditioned forest cover and forest cover with area-averaged spatial conditioning).

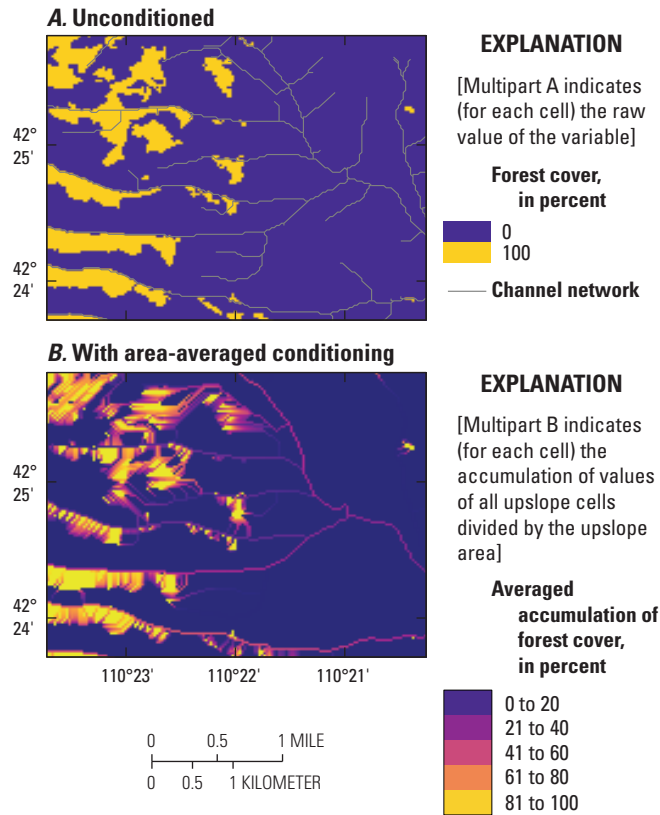


Figure 4. An example of a static physiographic variable, forest cover, shown before and after spatial conditioning with an area-averaged accumulation of cell values (figure area, [fig. 1](#)). A, unconditioned and B, with area-averaged conditioning.

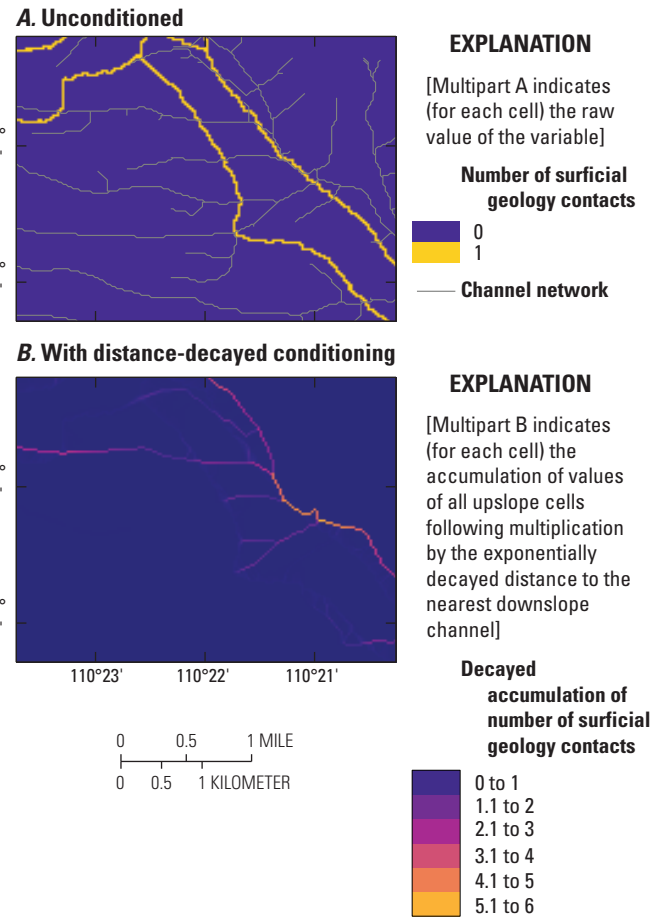


Figure 5. An example of a static physiographic variable, number of surficial geology contacts, shown before and after spatial conditioning with a distance-decayed accumulation of cell values (figure area, [fig. 1](#)). A, unconditioned and B, with distance-decayed conditioning.

Development of Machine Learning Streamflow Model

The MLFLOW model used a gradient boosting machine (Friedman, 2001, 2002; also known as boosted trees or generalized boosted models) because of the potential for the machine learning model to fit complex relations between the predictor variables (24 geology, hydrology, topography, vegetation, human, and climate variables, [table 3](#)) and the response variable (monthly streamflow; streamflow observations, [table 1](#)). Gradient boosting machines are one of many machine learning models, including random forests, support vector machines, and neural networks, that are used in various hydrologic modeling (Shen and others, 2018), including modeling of streamflow (Bellos and Carbajal, 2020). Gradient boosting machines were applied in this study using the “caret” (Kuhn 2008, 2020) and “gbm” (Greenwell and others, 2020; Ridgeway, 2020) packages in R (R Core Team, 2021).

A gradient boosting machine is a machine learning model used to solve regression or classification problems. The gradient boosting machine produces a predictive model that is an ensemble of many weaker models, which are typically implemented as decision trees. Gradient boosting is treated statistically as a numerical optimization problem with an objective of minimizing the loss of the model by adding weak learners using a functional gradient descent (Friedman, 2002). Gradient boosting machines are applied with the following features in R (Ridgeway, 2020): a loss function (squared error) to optimize; a weak learner (decision trees) for predicting; and an additive model (stochastic gradient descent) for adding weak learners to minimize the loss function. However, gradient boosting is a greedy model, meaning gradient boosting can readily overfit the data (Friedman, 2001). To limit the model from overfitting the data, the following four hyperparameters of gradient boosting machines can be tuned using packages in R (Kuhn, 2020, Greenwell and others, 2020): (1) the learning rate (shrinkage) of the decision trees; (2) the maximum interaction

depth (number of branches, or intermediate nodes) of a tree; (3) the minimum number of observations per leaf, or terminal node, of a tree; and (4) the number of iterations (trees) fit. The hyperparameters of the models in this study were tuned as follows: 0.005, 0.01, or 0.05 for shrinkage; 1 or 5 for maximum interaction depth; 5 or 10 for minimum number of observations; and 500 for number of trees. More detail on modeling with gradient boosting machines in general and specifically in R is available in Kuhn and Johnson (2013).

Development of the MLFLOW model proceeded in an iterative process of model calibration and validation. Streamflow observations (monthly streamflow) used to fit the model were split into training and testing samples using repeated k -fold cross-validation (Hastie and others, 2009). The data were randomly split into 10 folds, with each fold containing 10 percent of the data. The model was iteratively trained with 9 folds and tested against 1 fold left out. The splitting of data into 10 folds was repeated 5 times, totaling 50 iterations (resamples) of training and testing. For each resample, the model fitted 500 trees, with each tree iteratively learning from the previously fitted tree.

Evaluation of Predictive Performance and Variable Importance

Predictive performance of the MLFLOW model was evaluated with the following four goodness-of-fit metrics: coefficient of determination (R^2), Nash-Sutcliffe efficiency (NSE), NSE with log-transformed data (logNSE), and percent bias (PBIAS). R^2 measures the proportion of the data variance that can be explained by the model; values greater than 0.5 may be considered acceptable (Legates and McCabe, 1999). NSE measures the relation of the residual variance to the data variance (Nash and Sutcliffe, 1970). Because NSE is computed with squared values of predictions and observations, the metric is biased toward predictions agreeing with observations at larger values. Log transformation of data reduces the difference between large and small values, so logNSE is more responsive to agreement between predictions and observations at smaller values. PBIAS measures the average tendency that predicted values are greater or less than observed values (Moriasi and others, 2007). For the purposes of hydrologic modeling of monthly streamflow, values of NSE and logNSE greater than 0.5 and PBIAS less than 25 percent are considered satisfactory (Moriasi and others, 2007).

The MLFLOW model was developed not only as a predictive model but also as an explanatory model. The explanatory power of predictor variables in the model was evaluated with variable importance, which was computed as the reduction in squared error in any decision tree fitted using a predictor variable, averaged for every tree that included the variable (Friedman, 2001). Variable importance was evaluated for the MLFLOW model fitted to all data and for models fitted to data from each watershed or year separately. Moreover, the relation between predictor variables and monthly streamflow was

interpreted with partial dependence plots, which emphasize the effect of each predictor variable by marginalizing the effect of all other predictor variables in the model (Friedman, 2001).

Explanation and Prediction of Streamflow

The MLFLOW model was initially fitted to all data. This initial model was used to explain monthly streamflow in relation to the predictor variables for all the study area and study period (2012–17), and predictive performance and variable importance (including partial dependence plots) were evaluated. Furthermore, predictive performance was evaluated independently for every site with at least 8 streamflow observations, which was the median number of observations per site of all 125 sites sampled. Nine additional models were fitted, one model for each watershed (using data from all years in each model) and one model for each year (using data from all watersheds in each model). These additional models were used to explain spatial or temporal variation in the relation of monthly streamflow to the predictor variables, and variable importance of each model was evaluated. Lastly, the initial model, fitted to all data, was used to predict monthly streamflow at every cell on the channel networks in every month of 2012–17. The predictions were assumed to qualitatively represent natural streamflow because the MLFLOW model had no predictor variables or mechanism for quantifying the effect of human infrastructure on streamflow. The monthly streamflow prediction data for each cell (17,518 cells) and month (72 months) are available in the accompanying data release (McShane and Eddy-Miller, 2021). Further models were fitted for qualitatively assessing model sensitivity, which is explained in the following section.

Assessment of Model Sensitivity

Multiple models were fitted to qualitatively assess sensitivity of the MLFLOW model to variations of the predictor variables and streamflow observations. The MLFLOW model was applied using the following three variations of the data: (1) dynamic or static predictor variables with or without temporal or spatial conditioning, (2) streamflow observations differentially grouped by watershed or year, and (3) streamflow observations progressively reduced by percentage of sites or months available. Models developed using these data variations were qualitatively assessed using goodness-of-fit metrics (R^2 , NSE, logNSE, and PBIAS).

Models fitted to different combinations of predictor variables were used to assess sensitivity to selection and conditioning of the predictor variables—static versus dynamic; conditioned versus unconditioned. The following five combinations of the variables were used to fit models: (1) unconditioned climatic (dynamic) variables (4 climate variables); (2) dynamic variables with temporal conditioning (20 moving-average variants of the climate variables); (3) unconditioned dynamic and physiographic and anthropogenic (static)

variables (24 variables); (4) dynamic and static variables with spatial conditioning (24 variables); and (5) static variables with spatial conditioning and dynamic variables with spatial and temporal conditioning (44 variables, including the 20 moving-average variants of the climate variables) (table 3).

Model sensitivity to fitting data from a different area (watershed) or period (year) was assessed by differentially grouping streamflow observations by watershed or year for use in model training and testing. The following two approaches were used in grouping the observations: (1) training with observations from all groups (all watersheds and years) and testing on observations from each watershed (1, 2, or 3) or year (2012, 2013, 2014, 2015, 2016, or 2017) separately and (2) training with observations from all groups except one group left out sequentially and testing on observations from the left-out group; for example, training with data from watersheds 1 and 2 and testing on data from watershed 3, or training with data from 2012 through 2016 and testing on data from 2017.

Model sensitivity to the quantity of data fitted was assessed by progressively reducing streamflow observations by percentage of sites or months available for use in model fitting. Models were fitted using observations randomly sampled, with 50 iterations, in two percentages as follows: (1) observations from only 50 percent of the sites or months (63 sites or 18 months on average per random sample) and (2) observations from only 15 percent of the sites or months (about 19 sites or 6 months on average per random sample).

Results of Machine Learning Approach to Modeling Streamflow

The MLFLOW model performed satisfactorily—streamflow predictions generally agreed with observations at low and high flows. Most of the important predictors of monthly streamflow were temporally conditioned dynamic climatic variables. Temporal and spatial conditioning of the predictor variables improved the explanatory power of the variables in the MLFLOW model. The next three sections discuss predictive performance, variable importance, and model sensitivity to variations of the predictor variables and streamflow observations (including comparison of the MLFLOW model with other streamflow modeling approaches). Spatial and temporal variability of the streamflow predictions and utility and limitations of the MLFLOW model are discussed in the last two sections.

Predictive Performance

The MLFLOW model fitted to all data had satisfactory agreement between observed and predicted streamflow ($R^2=0.80$, $NSE=0.79$, $\log NSE=0.82$, and $PBIAS=0.7$ percent, fig. 6). The equivalence between NSE (0.79) and $\log NSE$

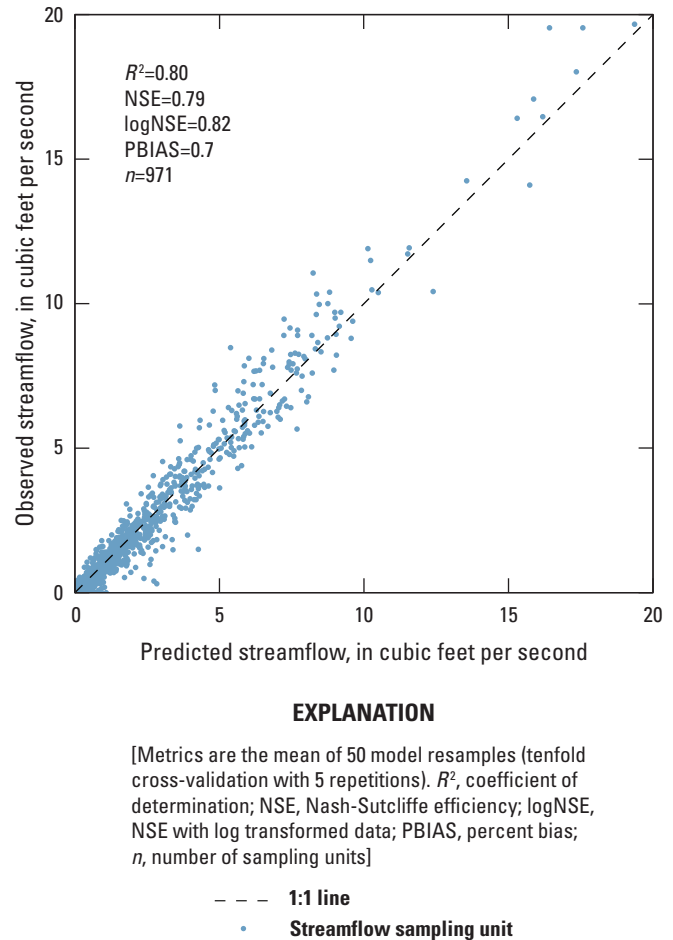


Figure 6. Relation between observed and predicted streamflow.

(0.82) indicated the MLFLOW model performed equally well for high and low flows. $PBIAS$ (0.7 percent) indicated the MLFLOW model did not overpredict or underpredict monthly streamflow in general. For streamflow observations less than 10 ft³/s, the model on average overpredicted by less than 1 percent. For streamflow observations toward 20 ft³/s, the model on average underpredicted by about 5 percent; however, the model had fewer streamflow observations to fit toward the upper range of values.

The MLFLOW model performed equally well for all months with streamflow observations. For each month, the mean of the distribution of residuals (observed minus predicted streamflow) of the model was about zero (fig. 7), indicating model predictions for months with fewer observations (such as February and April) were not biased by the model potentially overfitting data for months with more observations (such as June and July). For most months, the median of model predictions was higher than the median of streamflow observations, and the distribution of model predictions for all months was more limited than the distribution of streamflow observations (fig. 7), suggesting the MLFLOW model was fitting toward the mean of the data.

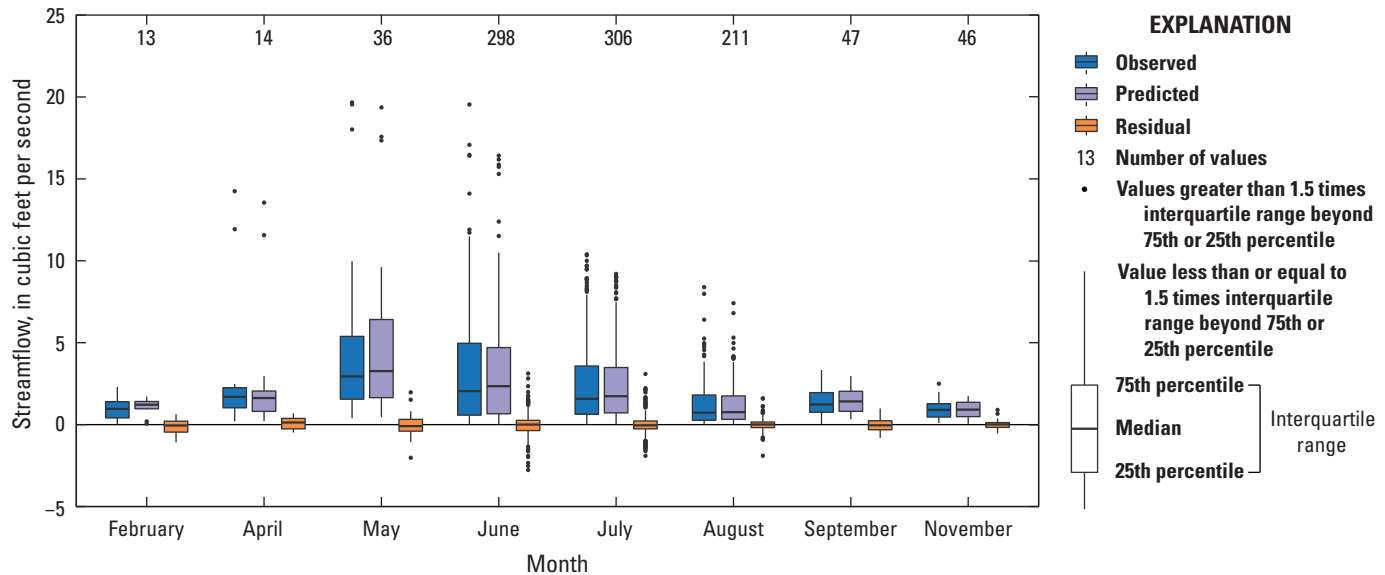


Figure 7. Distribution of observed and predicted streamflow and residual values (observed minus predicted) for each month sampled during 2012–17.

Goodness-of-fit metrics computed independently for every site with at least eight streamflow observations also indicated acceptable performance of the MLFLOW model (table 4). Medians of the goodness-of-fit metrics for the 85 sites analyzed were 0.81 for R^2 , 0.74 for NSE, 0.77 for logNSE, and 2.3 percent for PBIAS (table 4), not much different from the metrics for the MLFLOW model fitted to all data ($R^2=0.80$, $NSE=0.79$, $\logNSE=0.82$, and $PBIAS=0.7$ percent, fig. 6). Median absolute deviation (MAD) of the metrics indicated some variation in model performance among the sites analyzed (table 4). However, the difference between median and MAD for PBIAS was less than 25 percent (median plus or minus MAD), so the MLFLOW model performed satisfactorily even for sites toward the tails of the distribution (table 4). The other three metrics were all greater than or equal to 0.5 (median minus MAD) toward the lower distributional tail, also indicating satisfactory model performance (table 4).

Model performance at the three sites with the most streamflow observations in this study (sites 25, 27, and 31, fig. 1) was acceptable (fig. 8A–C). All four goodness-of-fit metrics indicated site 31 (fig. 8C) had the best performance in general ($R^2=0.88$, $NSE=0.88$, $\logNSE=0.91$, and

$PBIAS=0.8$ percent). Site 25 (fig. 8A) had R^2 of 0.89 and NSE of 0.84, comparable with values for site 31, but had worse logNSE (0.77) and PBIAS (–3.6 percent) than for site 31. Site 27 (fig. 8B) had the best PBIAS (–0.6 percent) and had logNSE (0.75) similar to the value for site 25, but R^2 (0.62) and NSE (0.61) were more than 25 percent less than for sites 25 or 31. Qualitatively, the MLFLOW model tended to underpredict the highest flows in most years, whereas lower flows were overpredicted in some years but underpredicted in others (fig. 8A–C). On average for the three sites, NSE (0.78) was lower than logNSE (0.81) (fig. 8A–C), indicating performance was only about 3 percent better for low flows than for high flows, the same as for the MLFLOW model fitted to all data. The tendency to underpredict the highest flows probably resulted from fewer observations of high flows than of low flows available for use in fitting the MLFLOW model.

In comparison, Miller and others (2018) used a random forest model fitted to USGS streamgage data to predict monthly streamflow for the conterminous United States. The approach of Miller and others (2018) included temporal conditioning of dynamic climatic predictor variables, as in this study. Median NSE ranged from 0.5 to 0.9 and median

Table 4. Distribution of goodness-of-fit metrics from repeated cross-validation of the machine learning streamflow model, fitted to all streamflow observations, computed for each site independently.

[Metrics were analyzed for 85 sites with at least 8 streamflow observations, which was the median number of observations per site of all 125 sites sampled. R^2 , coefficient of determination; NSE, Nash-Sutcliffe efficiency; logNSE, NSE with log-transformed data; PBIAS, percent bias; MAD, median absolute deviation]

Number of sites analyzed	R^2		NSE		logNSE		PBIAS	
	Median	MAD	Median	MAD	Median	MAD	Median	MAD
85	0.81	0.18	0.74	0.24	0.77	0.21	2.3	16.8

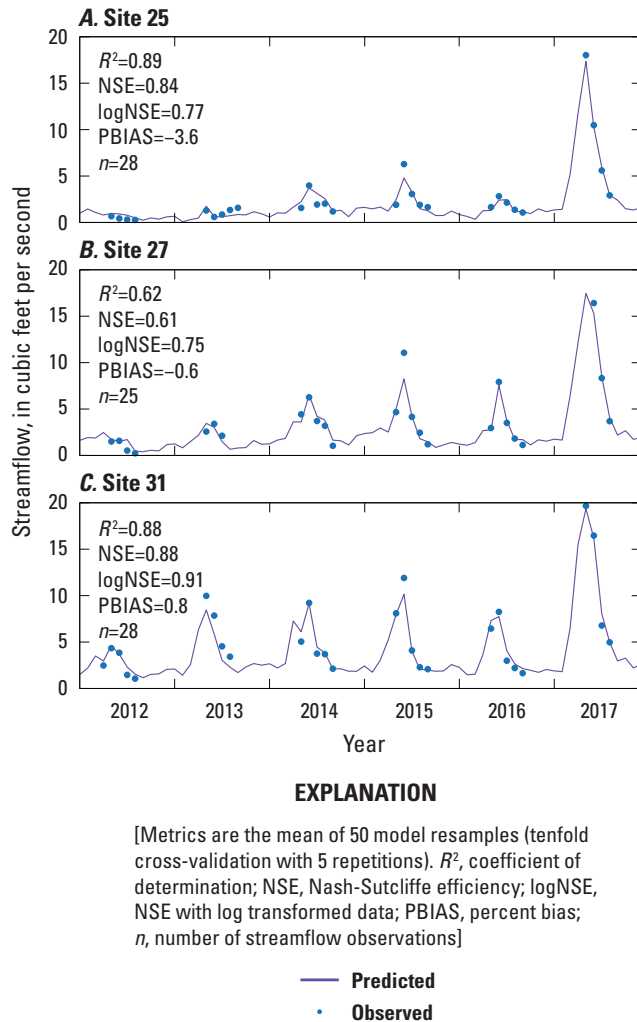


Figure 8. Observed and predicted streamflow at the three sites (fig. 1) with the most streamflow observations. A, site 25; B, site 27; and C, site 31.

PBIAS ranged from -15 to 5 percent (Miller and others, 2018), compared with median NSE of 0.74 and median PBIAS of 2.3 percent for every site with at least eight observations in this study (table 4). In addition, the model in Miller and others (2018) was fitted to few data from low-order streams, unlike the MLFLOW model in this study.

Additionally, the Precipitation-Runoff Modeling System, a physically based hydrologic model, was used to model monthly streamflow in seven watersheds (about HUC12 in size) in Montana (Chase and others, 2016). USGS streamgage data were used by Chase and others (2016) to calibrate and validate the model, whereas discrete streamflow observations were used in this study. The model performance in Chase and others (2016) ranged from negative values for NSE to about 0.75 , whereas in this study, median NSE was 0.74 at sites with at least eight observations (table 4) and NSE was 0.84 or more at the two sites with the most observations (fig. 8A, C).

Variable Importance

The most important variables (statistically important in the MLFLOW model) for explaining monthly streamflow were the 6-month moving average of precipitation and the 3-month moving average of snow water equivalent (fig. 9). The top 20 important variables were populated by the following 14 moving-average variants: 4 precipitation variants, 4 snow water equivalent variants, 3 evapotranspiration variants, and 3 temperature variants. Forest cover was the only static variable among the top five variables (fig. 9). Elevation, drainage area, depth to water table, number of diversions, and number of surficial geology contacts—also static variables—also were important variables for explaining monthly streamflow.

The 20 most important variables in the MLFLOW model had simple to more complex relations with monthly streamflow as interpreted with partial dependence plots. Many of the relations between the predictor variables and monthly streamflow were intuitive. Monthly streamflow increased with increasing drainage area (fig. 10O) and number of surficial geology contacts (fig. 10P) and decreased with increasing elevation (fig. 10R) and depth to water table (fig. 10T). Monthly streamflow also increased with increasing 6-, 9-, and 12-month moving averages of precipitation (fig. 10E–G, respectively) and 3- and 6-month moving averages of snow water equivalent (fig. 10J, K, respectively) and decreased with increasing 9-month moving averages of evapotranspiration (fig. 10C) and temperature (fig. 10N). Other relations between the predictor variables and monthly streamflow were counter-intuitive. Monthly streamflow decreased with increasing number of water diversions (fig. 10Q) and forest cover (fig. 10S). However, in the study area, most forest cover was at higher elevations where channels had smaller drainage areas, and

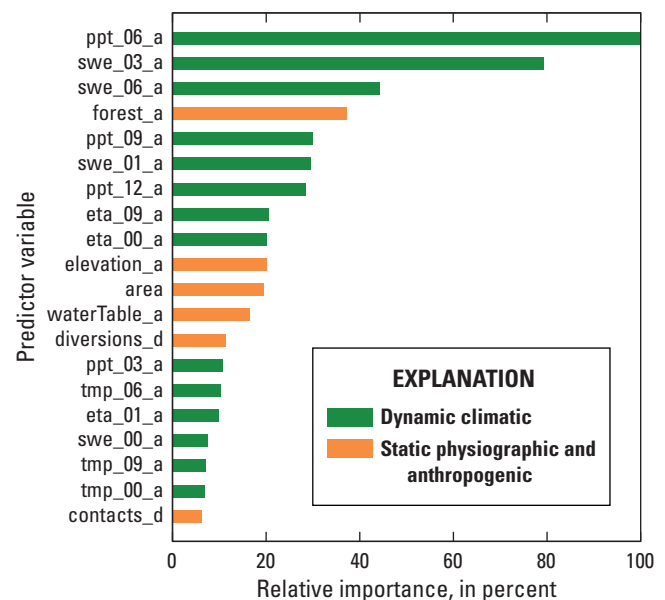


Figure 9. Relative importance of the top 20 variables (code, table 3) used in the model fitted to all data.

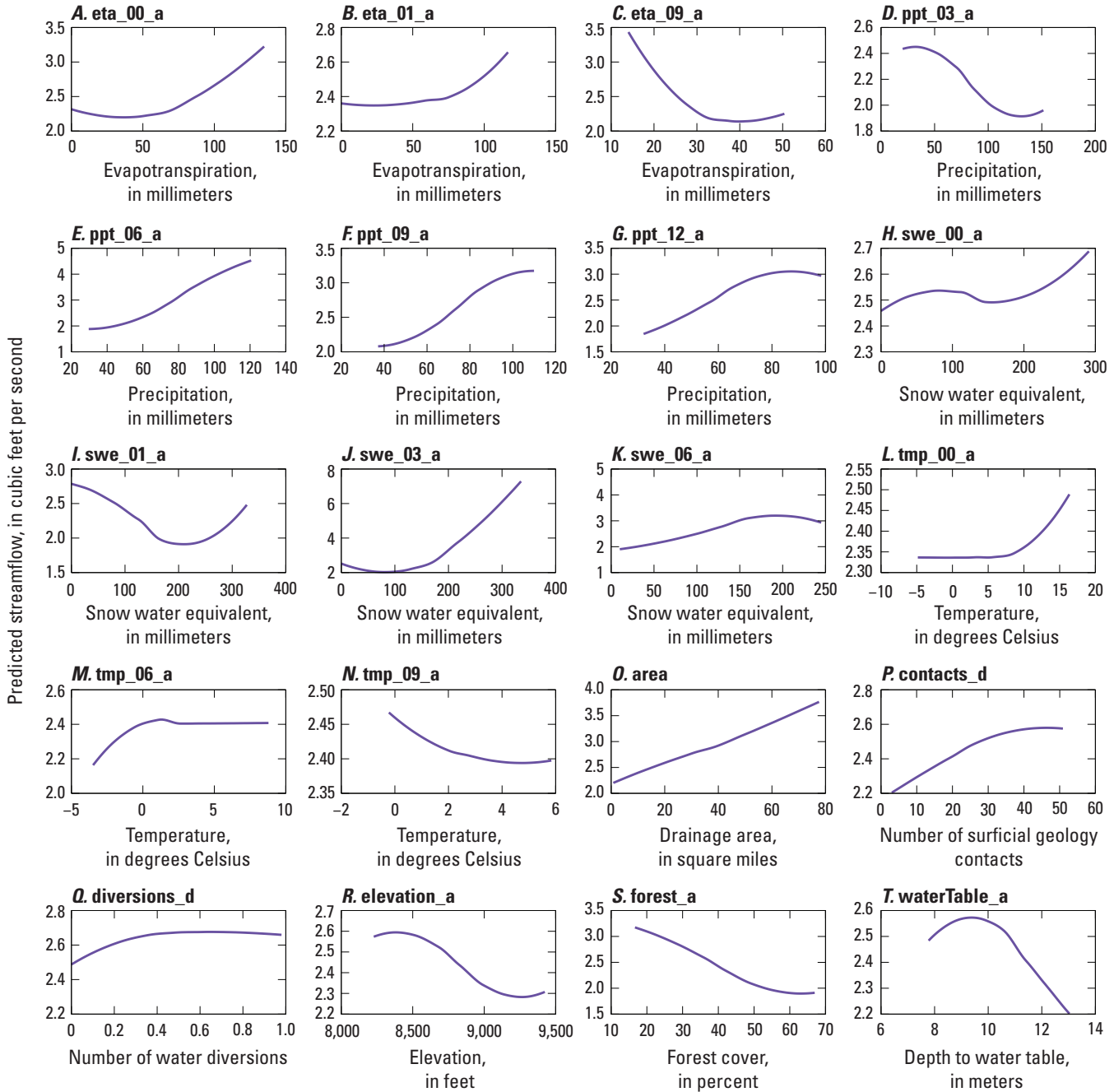


Figure 10. Partial dependence of streamflow on the top 20 most important variables (fig. 9) used in the model fitted to all data. A–N, dynamic climatic variables; O–T, static physiographic and anthropogenic variables (code, table 3).

most water diversions were on channels with larger drainage areas at lower elevations. Additionally, each climate variable had relations with monthly streamflow that reversed when proceeding from shorter to longer moving-average variants. For example, monthly streamflow increased with increasing current month and 1-month moving average of evapotranspiration (fig. 10A, B, respectively) and current month and 6-month moving average of temperature (fig. 10L, M, respectively), but monthly streamflow decreased with 9-month moving averages of evapotranspiration (fig. 10C) and temperature (fig. 10N). In addition, intermediate values of some predictor variables

produced the highest monthly streamflow (such as depth to water table, fig. 10T) or the lowest (such as 1-month moving average of snow water equivalent, fig. 10I), whereas for other predictor variables, the relation with monthly streamflow had more than one inflection. For example, as current month of snow water equivalent increased (fig. 10H), monthly streamflow increased, then decreased, and then increased again.

The most important variables in models fitted to data from each watershed did not differ substantially among the watersheds or from the MLFLOW model fitted to all data. For each watershed, moving-average variants of precipitation and

snow water equivalent were still the most important variables (fig. 11A–C), suggesting insubstantial spatial variability in the relation between monthly streamflow and the predictor variables. The effect of precipitation and snow water equivalent on monthly streamflow was apparently as important across small or large areas (each watershed modeled separately; fig. 11A–C) as across the study area (all watersheds modeled together; fig. 9). However, forest cover, contrary to its ranking as third most important variable in the model fitted to all data (fig. 9), was the fifth most important variable for watershed 1 (fig. 11A) and was not among the top five important variables for the other two watersheds (fig. 11B, C). In contrast, the 9-month moving average of snow water equivalent was not among the top 20 important variables in the model fitted to all data (fig. 9) but was the third most important variable for watershed 3 (fig. 11C). Additionally, drainage area was among the top five important variables for watersheds 2 and 3 (fig. 11B, C, respectively), but drainage area was only the 11th most important variable in the model fitted to all data (fig. 9). The increased importance of drainage area was understandable for watershed 3, which has the largest area (55 mi²) of the three watersheds, but the increased importance of drainage area was surprising for watershed 2, which has the smallest area (26 mi²) (U.S. Geological Survey, 2019a).

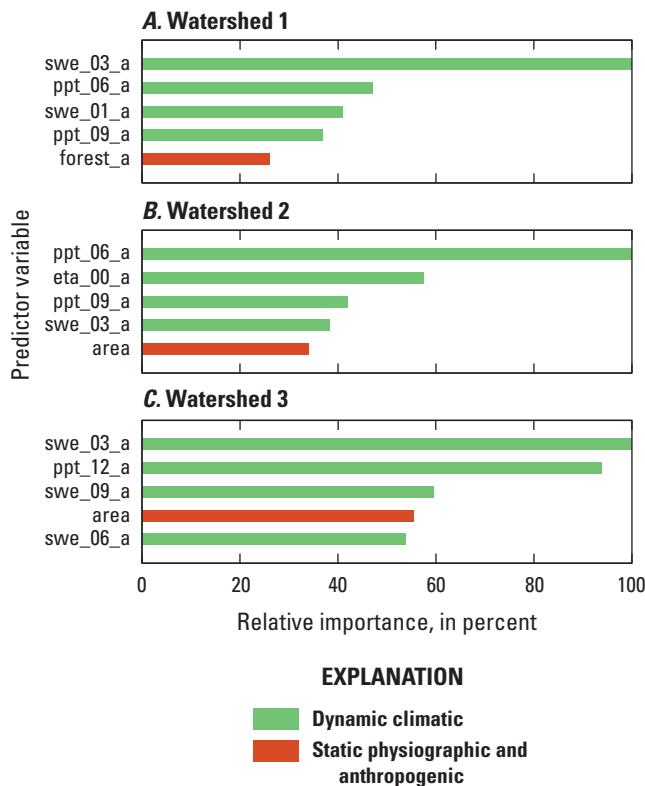


Figure 11. Relative importance of the top five variables (code, table 3) used in models fitted to data only from each watershed. A, watershed 1; B, watershed 2; and C, watershed 3.

The most important variables in models fitted to data from each year differed considerably among the years and from the MLFLOW model fitted to all data. Six static variables (cropland cover, number of surficial geology contacts, number of water diversions, forest cover, depth to water table, and drainage area)—from one to three variables for any year—were among the top five most important variables (fig. 12A–F), which suggests considerable temporal variability

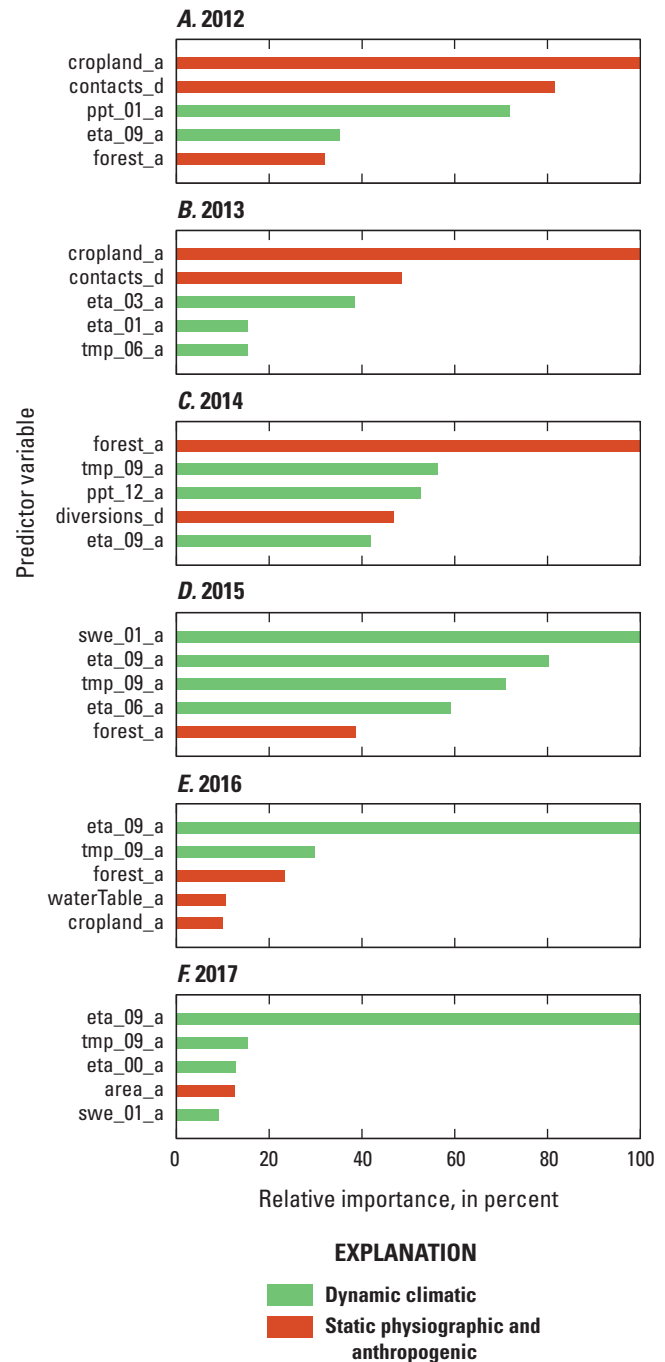


Figure 12. Relative importance of the top five variables (code, table 3) used in models fitted to data only from each year. A, 2012; B, 2013; C, 2014; D, 2015; E, 2016; and F, 2017.

in the relation between the predictor variables and monthly streamflow. In addition, moving-average variants of precipitation (fig. 12A, C) and snow water equivalent (fig. 12D, F) were among the top five important variables in only 2 years, whereas moving-average variants of evapotranspiration and temperature were among the top five important variables in all but 1 year (fig. 12B–F). During a shorter period (each year modeled separately; fig. 12A–F), the effect of evapotranspiration and temperature seemed more important, whereas the effect of precipitation and snow water equivalent seemed more important during the longer study period (all years modeled together; fig. 9). For 3 years, a dynamic variable was the most important—the 1-month moving average of snow water equivalent for 2015 (fig. 12D) and the 9-month moving average of evapotranspiration for 2016 (fig. 12E) and 2017 (fig. 12F). However, for the other 3 years, a static variable was the most important—cropland cover for 2012 (fig. 12A) and 2013 (fig. 12B) and forest cover for 2014 (fig. 12C). Although the varying importance of static variables from 1 year to another year may seem illogical, several reasons are possible. Static variables can interact differently with dynamic variables from year to year, resulting in the increased importance of a static variable. Dynamic variables alone can decrease in importance from year to year, which relatively increases the importance of a static variable. Dynamic variables may vary less within 1 year than during several years, resulting in the decreased importance of dynamic variables for a single year. Lastly, one static variable may substitute, in relative importance, for many correlated dynamic variables.

Model Sensitivity

Temporal and spatial conditioning intensified the relation of many predictor variables with monthly streamflow, resulting in more information that the MLFLOW model could use for predicting streamflow. For many of the variables, such as forest cover or surficial geology contacts, unconditioned values of many cells on the channel networks were zero (figs. 4A, 5A), but with spatial conditioning, values of most cells on the channel networks were nonzero (figs. 4B, 5B), providing the MLFLOW model with more diversely valued variables that might better explain variation in streamflow. Temporal conditioning increased r for the dynamic variables by as much as 0.34 compared with the equivalent variable without conditioning (table 5). For example, current month of precipitation had r of -0.02 (ppt_00, table 5), whereas the 6-month moving average of precipitation had r of 0.36 (ppt_06, table 5). Spatial conditioning increased r for the moving-average variants of the dynamic variables by as much as 0.25 (table 5). For example, the 1-month moving average of snow water equivalent without conditioning had r of 0.10 (swe_01, table 5), but the 1-month moving average of snow water equivalent with conditioning had r of 0.35 (swe_01_a, table 5). For the static variables, spatial conditioning increased r by as much as 0.15, such as for surficial geology contacts (contacts, table 5). Also,

for some static variables, correlation between the unconditioned variable and monthly streamflow was not possible to compute because all values of the variable were zero for any streamflow observation, such as for water diversions (diversions, table 5). However, spatial conditioning produced a correlation of the variable with streamflow; for example, r of 0.09 for water diversions (diversions_d, table 5).

Models fitted to different combinations of predictor variables were used to qualitatively assess sensitivity of the MLFLOW model to selection and conditioning of the predictor variables—static versus dynamic; conditioned versus unconditioned. Performance improved progressively from model 1 to model 5 (table 6). All models had acceptable performance, although model 1 was marginal for three goodness-of-fit metrics ($R^2=0.54$, $NSE=0.52$, and $\logNSE=0.49$, table 6). However, model 5 (the model fitted to all data) had the best performance ($R^2=0.80$, $NSE=0.79$, $\logNSE=0.82$, and $PBIAS=0.7$ percent, table 6).

Related to this study, Jaeger and others (2019) used a machine learning approach to modeling streamflow permanence (presence or absence of streamflow) in the Columbia River Basin. Jaeger and others (2019) also applied spatial conditioning by area-averaged accumulation to the predictor variables, as in this study. Model performance was about 80 percent (out-of-bag error rate of 20 percent; Jaeger and others, 2019), which was comparable with this study ($R^2=0.80$, $NSE=0.79$, and $\logNSE=0.82$, table 6). Updated streamflow permanence modeling has been progressing in the upper Missouri River Basin (not shown), including temporal conditioning of dynamic predictor variables and spatial conditioning by distance-decayed accumulation (Roy Sando, U.S. Geological Survey, oral commun., 2020), similar to this study.

Different combinations of static or dynamic variables and unconditioned or temporally or spatially conditioned variables, or both, were used to fit the MLFLOW model, which resulted in considerable improvements in performance. Temporal conditioning of dynamic variables in model 2 increased R^2 , NSE , and \logNSE by as much as 0.09 compared with the unconditioned dynamic variables in model 1 (table 6). Adding unconditioned physiographic and anthropogenic variables in model 3 increased R^2 , NSE , and \logNSE by as much as 0.15 compared with just the climatic variables in model 1 (table 6). Spatial conditioning of all variables in model 4 increased R^2 , NSE , and \logNSE by as much as 0.05 compared with the unconditioned dynamic and static variables in model 3 (table 6). Spatial conditioning of all variables together with temporal conditioning of dynamic variables in model 5 increased R^2 , NSE , and \logNSE by as much as 0.13 compared with just the spatial conditioning in model 4 (table 6). $PBIAS$ varied less than 1 percent among the five models (table 6), indicating the MLFLOW model did not overpredict or underpredict monthly streamflow in general within the study area during the study period. In addition, \logNSE was worse than NSE by 0.03 in model 1 but was better by 0.03 in model 5 (table 6). This 0.06 improvement of \logNSE relative to NSE

Table 5. Relation of streamflow to some dynamic and static variables before and after temporal or spatial conditioning.[*r*, Pearson correlation coefficient; --, not possible to compute]

Variable without conditioning		Equivalent variable with conditioning	
Code (table 3)	<i>r</i>	Code (table 3)	<i>r</i>
Temporal conditioning of climatic (dynamic) variables			
eta_00	0.26	eta_01	0.16
ppt_00	−0.02	ppt_06	0.36
swe_00	−0.06	swe_03	0.33
tmp_00	0.01	tmp_06	−0.26
Spatial conditioning of dynamic and physiographic and anthropogenic (static) variables			
eta_12	0.13	eta_12_a	0.20
ppt_09	0.31	ppt_09_a	0.51
swe_01	0.10	swe_01_a	0.35
tmp_01	−0.10	tmp_01_a	−0.12
contacts	−0.02	contacts_d	0.17
diversions	--	diversions_d	0.09
forest	−0.07	forest_a	−0.20
waterTable	0.04	waterTable_a	−0.06

Table 6. Goodness-of-fit metrics for models fitted to different combinations of static and dynamic variables before or after spatial or temporal conditioning, indicating qualitative model sensitivity to selection and conditioning of the predictor variables.[Metrics are the mean of 50 resamples (tenfold cross-validation with 5 repetitions) per model. R^2 , coefficient of determination; NSE, Nash-Sutcliffe efficiency; logNSE, NSE with log-transformed data; PBIAS, percent bias]

Model number	Selection and conditioning of variables used to fit model	Number of variables	R^2	NSE	logNSE	PBIAS
1	Unconditioned climatic (dynamic) variables	4	0.54	0.52	0.49	0.6
2	Dynamic variables with temporal conditioning	20	0.60	0.58	0.58	1.2
3	Unconditioned dynamic and physiographic and anthropogenic (static) variables	24	0.66	0.65	0.64	1.2
4	Dynamic and static variables with spatial conditioning	24	0.69	0.67	0.69	0.6
5	Static variables with spatial conditioning and dynamic variables with spatial and temporal conditioning	44	0.80	0.79	0.82	0.7

indicates selection of all variables with all temporal and spatial conditioning improved prediction of low flows relative to prediction of high flows.

The MLFLOW model was most sensitive to selection of dynamic climatic variables. Unconditioned dynamic climatic variables (model 1) alone explained 54 percent of the variance ($R^2=0.54$, table 6) in monthly streamflow, whereas adding static physiographic and anthropogenic variables (model 3) only explained 12 percent more of the variance ($R^2=0.66$, table 6), indicating the greater importance of the time series data. Also, spatial conditioning of all variables together with temporal conditioning of dynamic variables (model 5) increased the variance explained in the MLFLOW model by another 14 percent ($R^2=0.80$, table 6). Kratzert and others (2019) used another machine learning approach (long

short-term memory networks) with different combinations of dynamic and static variables that performed similarly to this study. Models developed with only dynamic climatic variables performed well (NSE=0.63) but performed moderately better with the inclusion of static physiographic variables (NSE=0.74) (Kratzert and others, 2019). However, model performance indicating that the dynamic variables had greater explanatory power than the static variables in the MLFLOW model was not surprising because time series data intrinsically have more variation, temporal and spatial, to use in fitting the model to the streamflow observations, which also are innately variable, temporally and spatially.

Sensitivity of the MLFLOW model to using data from a different area (watershed) or period (year) was qualitatively assessed by differentially grouping streamflow observations

by watershed or year for use in model training and testing. For models trained with all watersheds and years, performance was better in testing on observations from each watershed than from each year separately (table 7), indicating the MLFLOW model had greater sensitivity to temporal than to spatial differences in the data. Means of the goodness-of-fit metrics were 0.79 for R^2 , 0.79 for NSE, 0.82 for logNSE, and 1.4 percent for absolute value of PBIAS for models trained with all watersheds but were 0.68 for R^2 , 0.67 for NSE, 0.67 for logNSE, and 3.5 percent for absolute value of PBIAS for models trained with all years (table 7). However, in contrast, models trained with all except 1 year left out sequentially and tested on the left-out year performed better than models trained with all except one watershed left out sequentially and tested on the left-out watershed. Means of the goodness-of-fit metrics were 0.51 for R^2 , 0.49 for NSE, 0.50 for logNSE, and 11 percent for absolute value of PBIAS for models trained with a watershed left out but were 0.56 for R^2 , 0.53 for NSE, 0.56 for logNSE, and 12 percent for absolute value of PBIAS for models trained with 1 year left out (table 7).

In general, performance for models trained with all except one watershed or 1 year left out sequentially was satisfactory for all the test watersheds and years (table 7). Model performance was only unsatisfactory (for one goodness-of-fit metric) for watershed 3 (logNSE=0.37) and 2013 (logNSE=0.37) (table 7), indicating the relations between the predictor variables and monthly streamflow at low flows in watershed 3 and in 2013 were difficult to explain with data from the other watersheds or years. Less similarity of data among the three watersheds than of data among the 6 years may explain the better performance of the model in testing on 1 year left out from model training than on one watershed left out from model training. However, another explanation may be fewer data were excluded from model training by leaving out 1 year than by leaving out one watershed, and consequently, more data (with more variance) were included in model testing on the left-out watershed than on the left-out year—on average, each watershed and year had 324 and 162 observations, respectively.

Sensitivity of the MLFLOW model to the quantity of data used was qualitatively assessed by progressively reducing streamflow observations by percentage of sites or months

Table 7. Goodness-of-fit metrics for models trained with observations from all watersheds and years and tested on observations from each watershed or year separately or for models trained with observations from all except one watershed or 1 year left out sequentially and tested on observations from the left-out watershed or year, indicating qualitative model sensitivity to data used from a different area or period.

[Metrics are the mean of 50 resamples (tenfold cross-validation with 5 repetitions) per model. R^2 , coefficient of determination; NSE, Nash-Sutcliffe efficiency; logNSE, NSE with log-transformed data; PBIAS, percent bias]

Groups used to train model	Group used to test model	R^2	NSE	logNSE	PBIAS
Data grouped by watershed					
All watersheds	Watershed 1	0.82	0.82	0.84	-1.3
	Watershed 2	0.79	0.79	0.84	3.0
	Watershed 3	0.77	0.77	0.77	0.0
All except the test watershed	Watershed 1	0.49	0.48	0.56	-8.1
	Watershed 2	0.54	0.52	0.56	5.1
	Watershed 3	0.50	0.46	0.37	19.9
Data grouped by year					
All years	2012	0.68	0.63	0.58	9.3
	2013	0.70	0.69	0.65	2.5
	2014	0.56	0.56	0.63	2.8
	2015	0.62	0.62	0.65	-1.4
	2016	0.78	0.77	0.74	2.5
	2017	0.75	0.75	0.75	-2.2
All except the test year	2012	0.61	0.59	0.55	9.9
	2013	0.54	0.51	0.37	18.6
	2014	0.59	0.58	0.63	6.9
	2015	0.57	0.55	0.63	-10.3
	2016	0.53	0.50	0.57	9.4
	2017	0.49	0.46	0.62	-14.6

available for use in model fitting. Performance was better for models fitted to fewer sites than to fewer months of observations (table 8), which indicated the MLFLOW model was more sensitive to temporal than to spatial differences in the data. The goodness-of-fit metrics for models fitted to 50 percent fewer months were worse than for models fitted to 50 percent fewer months by 0.12 for R^2 , 0.14 for NSE, 0.17 for logNSE, and 6.9 percent for absolute value of PBIAS (table 8). In addition, further reduction in sites and months from 50 to 15 percent affected the goodness-of-fit metrics more for models fitted to fewer months than to fewer sites. On average, for three goodness-of-fit metrics (R^2 , NSE, and logNSE), values decreased by 0.21 for models fitted to 85 percent fewer sites but decreased by 0.27 for models fitted to 85 percent fewer months (table 8). Performance was still satisfactory for models fitted to 50 percent of sites and was marginally satisfactory for models fitted to 50 percent of months but was no longer satisfactory for models fitted to only 15 percent of sites or months (table 8).

The MLFLOW model fitted to all data performed very well ($R^2=0.80$, NSE=0.79, logNSE=0.82, and PBIAS=0.7 percent, table 8), but performance in general progressively decreased for models fitted to progressively reduced percentages of sites or months (table 8). However, the reductions in sites and months were applied using random samples of the data, which may explain the poor performance of models fitted to only 15 percent of sites or months. The 15-percent samples of sites or months used to train the MLFLOW model may not have been adequately representative of the variance in the remaining 85 percent of the data used to test the model. If sites could be sampled more uniformly across a watershed, focusing on proportional representation of small, medium, and large streams (by drainage area), or if months could be sampled more intentionally throughout the year, focusing on key moments of the annual hydrograph for streams at different elevations (low, medium, and high), then data for fewer sites

and months might be sufficient for the MLFLOW model to perform as well as the model performed with 971 streamflow observations.

Streamflow Predictions

The MLFLOW model predicted spatially and temporally continuous monthly streamflow for the study area (17,518 cells; fig. 1) and study period (72 months; 2012–17), using 971 discrete streamflow observations (table 1). Spatial and temporal variations in the streamflow predictions are discussed in this section using, as an example, a subset of the study area (part of the channel network of watershed 2; figure area, fig. 1) and a subset of the study period (every month of 2017 and every year in August).

Intra-annual variation in streamflow was simulated realistically by the MLFLOW model—seasonality of streamflow was well characterized. Predicted flows in 2017 were much lower from January through March (fig. 13A–C) and from August through December (fig. 13H–L), months before or after the snowmelt high flows during April through June (fig. 13D–F). Using February, May, August, and November to represent winter, spring, summer, and fall, respectively, mean of cells on the channel network was 8.9 ft³/s in May (fig. 13E) but only 1.0, 1.8, and 0.51 ft³/s in February, August, and November, respectively (fig. 13B, H, K, respectively). Furthermore, in May, some cells had predicted flows of almost 17 ft³/s (fig. 13E), whereas in February and November, many cells on the channel networks had predicted flows of 0 ft³/s (fig. 13B, K, respectively). However, because 2017 was a wetter year, no cells on the channel network in August had predicted flows of 0 ft³/s (zero flow; fig. 14F), whereas in August of 2012, which was a drier year, about 45 percent of cells on smaller channels had zero-flow predictions (fig. 14A). In February, predicted flows were higher in smaller channels than in larger channels (fig. 13B), but in May, larger channels had considerably higher predicted flows (fig. 13E), and in August

Table 8. Goodness-of-fit metrics for the model fitted to all observations or for models fitted to observations progressively reduced by percentage of sites or months available, indicating qualitative model sensitivity to quantity of the data used.

[Metrics are the mean of 50 resamples (tenfold cross-validation with 5 repetitions) per model. R^2 , coefficient of determination; NSE, Nash-Sutcliffe efficiency; logNSE, NSE with log-transformed data; PBIAS, percent bias]

Sampling of data used to fit model	Median number of sites per sample	Median number of months per sample	Median number of observations per sample	R^2	NSE	logNSE	PBIAS
All sites and months	125	35	971	0.80	0.79	0.82	0.7
Data reduced by site							
50 percent of sites	63	34	492	0.58	0.56	0.62	1.5
15 percent of sites	19	32	148	0.38	0.36	0.39	1.9
Data reduced by month							
50 percent of months	125	18	510	0.46	0.42	0.45	8.6
15 percent of months	97	6	155	0.23	0.16	0.14	–1.8

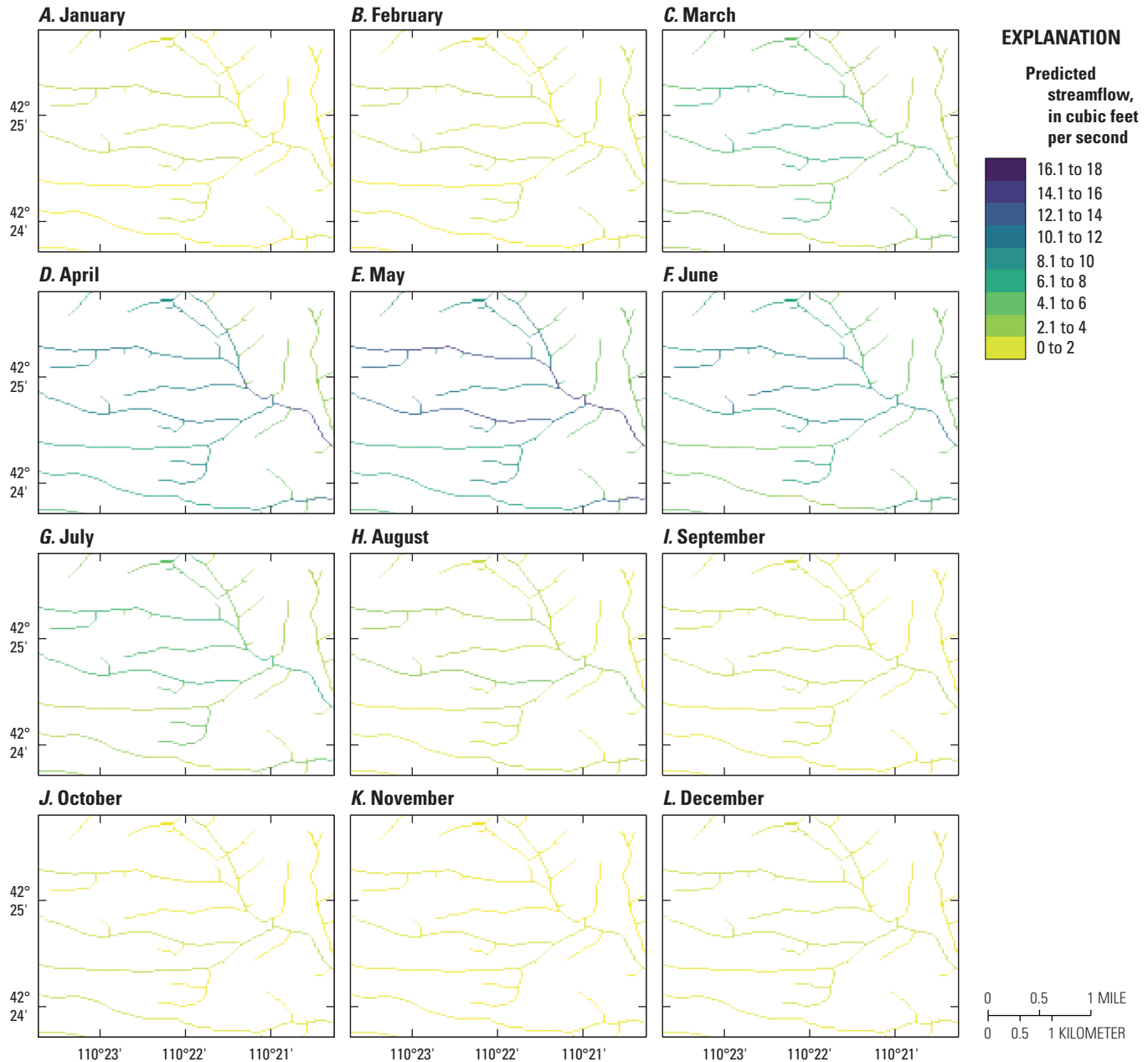


Figure 13. Intra-annual variation in streamflow predictions on part of a channel network (figure area, [fig. 1](#)) for each month of 2017. A, January; B, February; C, March; D, April; E, May; F, June; G, July; H, August; I, September; J, October; K, November; and L, December.

and November, streamflow receded, remaining higher in more intermediate channels ([fig. 13H, K](#), respectively). Higher flows were observed at sites with smaller drainage areas and higher snow water equivalent in February, at sites with larger drainage areas in May, and at sites with larger drainage areas or higher precipitation in August and November (McShane and Eddy-Miller, 2021).

The MLFLOW model also realistically simulated interannual variation in streamflow, well representing yearly hydroclimatic conditions. Predicted flows in August were low in 2012 and 2013 ([fig. 14A, B](#), respectively), which were drier years, but warmer in 2012 and cooler in 2013 (McShane

and Eddy-Miller, 2021), whereas predicted flows were high in 2014 and 2017 ([fig. 14C, F](#), respectively), which were years with more normal temperature but wetter, with more rainfall in 2014 and more snowpack in 2017 (McShane and Eddy-Miller, 2021). Although 2015 was a drier, warmer year similar to 2012, predicted flows were higher in 2015 ([fig. 14D](#)) than in 2012 ([fig. 14A](#)), most likely because 2015 followed 2014, a wetter year. Similarly, although 2016 was a wetter year similar to 2014 and 2017, predicted flows in 2016 were more intermediate ([fig. 14E](#)) similar to 2015 ([fig. 14D](#)), most likely because the preceding year, 2015, was drier. Using

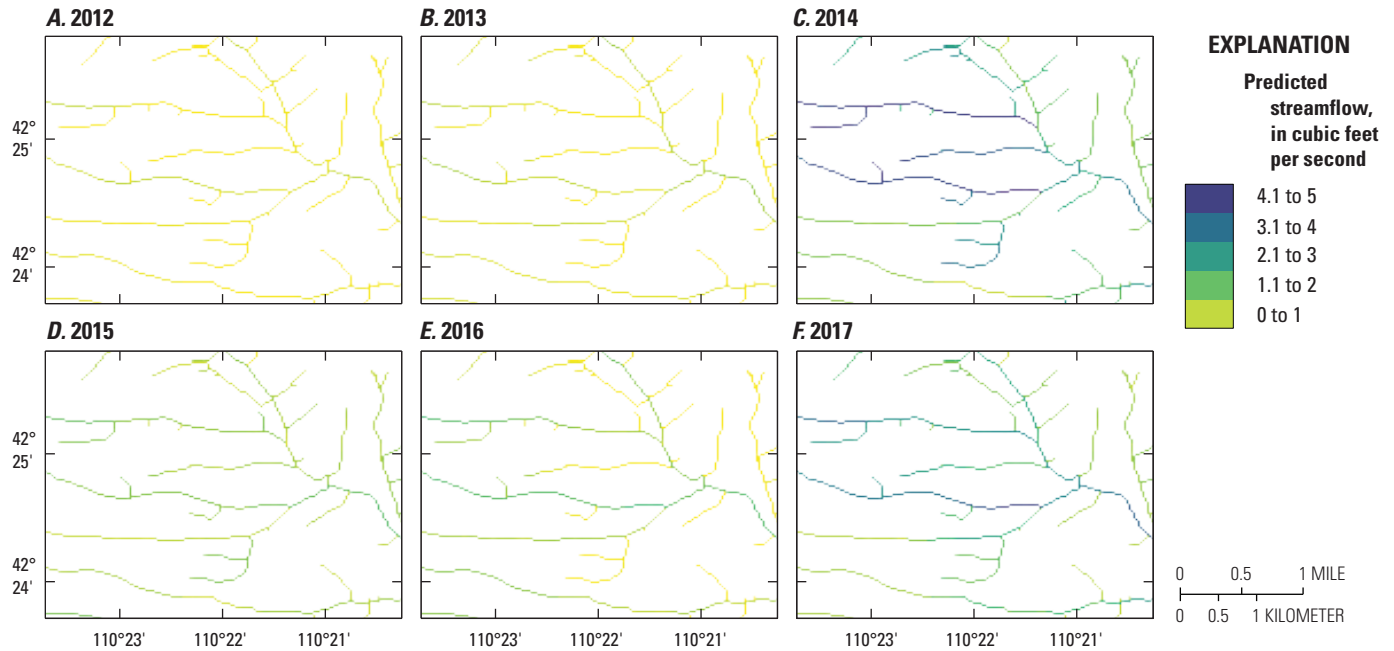


Figure 14. Interannual variation in streamflow predictions on part of a channel network (figure area, [fig. 1](#)) for each year in August. A, 2012; B, 2013; C, 2014; D, 2015; E, 2016; and F, 2017.

2012 and 2017 to represent a range of hydroclimatic conditions from warmer and drier to cooler and wetter, respectively, mean predicted flow was only 0.12 ft³/s in 2012 but 1.8 ft³/s in 2017. Moreover, maximum predicted flow in 2012 was only 0.61 ft³/s, whereas minimum predicted flow in 2017 was 0.19 ft³/s. In 2017, a year with more snowpack, mean predicted flow in August was 1.8 ft³/s, whereas in 2014, a year with more rainfall, mean predicted flow in August was 2.4 ft³/s, indicating the responsiveness of the model to recent-month precipitation versus earlier-year snow water equivalent (McShane and Eddy-Miller, 2021). In August, observed flows at sites on larger channels were lower in 2012 than in 2013 because these sites had lower snow water equivalent in 2012 than in 2013. Observed flows were higher at sites on larger channels in 2017 but were higher at sites on more intermediate channels in 2014. This difference resulted from higher snow water equivalent at sites on larger channels in 2017 than in 2014 and higher precipitation at sites on more intermediate channels in 2014 than in 2017 (McShane and Eddy-Miller, 2021).

Utility and Limitations of Modeling Approach

The greatest utility of the modeling approach is the ease of use and the speed of processing input data, running the model, and interpreting the model output, whereas the greatest limitation is the need for spatially and temporally representative streamflow observations to drive the model. Not only is the quantity of observations an important limitation but the quality of observations also is an important limitation.

Streamflow measurements might be made by poorly trained persons and errors could be proportionally larger at low flows than at high flows; however, in this study, all measurements were made by well trained personnel following standard USGS techniques and methods (Rantz and others, 1982; Sauer and Turnipseed, 2010). In addition, the MLFLOW model performed well with the available streamflow measurements, which were made at a spatially representative number of sites and in a temporally representative number of months.

Streamflow predictions seemed to well represent the annual hydrograph within the study area during the study period. However, streamflow measurements were not made in January, March, October, or December, so model performance was not computable for those 4 months. In addition, for the other 8 months, streamflow measurements were not made at every site in every month of all 6 years, so the evaluation of predictive performance was more limited than what would be possible for a model developed using time series data from a USGS streamgauge. Also, model predictions in this study were assumed representative of natural streamflow. Although three anthropogenic predictor variables (water diversions, roads, and water wells) were used in the MLFLOW model, the variables were merely informative of potential hydrologic alteration in a qualitative sense. Therefore, the MLFLOW model did not quantitatively characterize any actual streamflow regulation that was due to diversion, release, or return of water.

These observations, although single measurements in a month, were treated as representative of mean monthly streamflow for use in the MLFLOW model. Because the annual hydrograph of the study area is characterized by a relatively short period (about 3 months) of higher flows and a

much longer period (about 9 months) of lower flows, a streamflow measurement made on any day of the month is likely representative of the mean monthly streamflow for most months of the year. However, during the rise and fall of the high-flow period, measurements made earlier and later in a month may be quite different from each other. Measurements may be lower in early May than in late May, and measurements may be higher in early June than in late June. Therefore, the magnitude of the highest flow may not be accurately represented with the streamflow observations, but the magnitude of the shorter high-flow period (2 months) relative to the magnitude of the longer low-flow period (8 months) was evident with the available streamflow observations.

The MLFLOW model, fitted as is with the data used in this study, is not transferable to another watershed unless the other watershed has similar statistical distributions of the physiographic, anthropogenic, and climatic variables. Transfer of the model to another watershed with values of these variables at the margins of the variables' statistical distributions would require judgment on the reliability of the model's application. However, because of the data-driven nature of the machine learning approach in this study, new streamflow observations in addition to the current observations can be used to refit the model. Nonetheless, transferring the MLFLOW model, fitted as is, to a watershed with a different hydroclimatic regime would not be as reasonable as fitting a new model to data for the new regime.

The modeling approach in this study was not a process-based model with mathematical functions derived from first principles or empirical research. Instead, temporal and spatial conditioning processes were applied to the predictor variables as a means to substitute simple properties of the data—temporally and spatially averaged or decayed values of climatic, physiographic, and anthropogenic variables—for some of the complex process-based functions (requiring much parameter optimization) for climate and land surface elements in a physically based hydrologic model.

Temporal conditioning was an effective and efficient way of increasing the information content of the dynamic climatic variables. A moving average was computationally simple and provided a simplified means for simulating short to long periods of the water storage change in a runoff model. For monthly streamflow as modeled in this study, the effect of climate is less immediate than in an event-based model. Generating a progression of moving average values of the time series data provided the model with more dimensionality for exploring the relation of climate to streamflow. However, the multiplicity of moving-average variants of the dynamic variables may have increased the use of these variables in the MLFLOW model because of the greater abundance of the dynamic variables relative to the static variables. In addition, too many moving-average variants of the climate variables can complicate evaluation of variable importance, including interpretation of partial dependence plots.

Spatial conditioning also had an important effect on the information content of the dynamic and static variables. Spatial conditioning was especially important to the static variables with binary data because no or few nonzero values of water diversions, springs, roads, water wells, surficial geology contacts, or bedrock geology faults corresponded to any sampling site, meaning these six variables (without spatial conditioning) had no variance for the MLFLOW model to explain. However, spatial conditioning of the unconditioned variables altered them from discrete (0 or 1) to continuous (ranging from 0 to 1) data for each cell (with only nonzero values on the channel network). Consequently, all sampling sites had a nonzero value of the six spatially conditioned variables—variance was generated in these variables for the MLFLOW model to explain. Area-averaged or distance-decayed accumulation of the spatial data not only increased explanatory power of the variables but also modified incompatible spatially discrete variables into applicable spatially continuous variables.

However, conditioning of the variables was less important than selection of dynamic variables for use in the MLFLOW model. The variance of the time series data was more attuned to the variance of the streamflow observations because every monthly streamflow observation had a corresponding monthly value of evapotranspiration, precipitation, snow water equivalent, and temperature. Therefore, the selection of additional dynamic variables might increase performance more than any static variable already selected for use in the MLFLOW model.

Gradient boosting machines used in this study are just one of many available machine learning models, and other circumstances might make a different model better. However, most machine learning models can be readily applied in R, such as neural networks, support vector machines, and random forests. The “caret” package in R has more than 100 different models (Kuhn, 2020). Carlisle and others (2016) evaluated five machine learning models, including random forests and gradient boosting machines, and determined that random forests performed more ably than the other four machine learning models. However, in this study, the MLFLOW model also was developed with a random forest, but performance was not as good as for the MLFLOW model developed with the gradient boosting machine.

The modeling approach in this study is relatively simple to apply. Other predictor variables can be used in the MLFLOW model, and variables that were most important in the model are available for larger areas and longer periods than were applicable to this study. The R scripts are adaptable—other machine learning models can be substituted for the gradient boosting machine used in the MLFLOW model. Although familiarity with R is necessary, only a working knowledge of hydrology (for selecting appropriate predictor variables and evaluating the quality of streamflow observations) and a rudimentary understanding of machine learning models are needed. Therefore, this modeling approach is practicable for other scientists who work with water but who are not hydrologists.

Summary

Streamflow is a necessary ecological resource for many animal and plant species. However, scant availability of streamflow data can impede the utility of streamflow as a variable in ecological models of aquatic and terrestrial species, especially when studying small streams in watersheds that lack streamgages. Much ecological research on small streams is concentrated on species-habitat relations throughout the channel network of a watershed, but streamflow data, needed at fine resolution and broad extent for this research, are typically sparse. For instance, collaborators with the Wyoming Cooperative Fish and Wildlife Research Unit needed more detailed streamflow information for researching the effects of multiple stressors on fish and invertebrates in several ungaged watersheds in the upper Green River Basin in southwestern Wyoming, where streamflow data are sparse.

Several approaches to modeling streamflow at various scales have been developed that may improve the utility of sparse streamflow data to ecological research in small streams, including machine learning models that can fit potentially complex relations between streamflow observations and environmental predictor variables. We developed a machine learning approach in R for modeling spatially and temporally continuous monthly streamflow from 2012 through 2017 in three semiarid montane-steppe watersheds (with drainage areas of 26–55 square miles and mean elevations of 8,031–8,455 feet) on the Wyoming Range.

A machine learning streamflow (MLFLOW) model was developed to predict monthly streamflow using static and dynamic variables derived from geospatial and time series data and to explain the multidimensional relation of monthly streamflow to the predictor variables. Streamflow measurements were made at 125 sites in 35 months during 2012–17, totaling 971 discrete observations; only a single measurement was made at any site in any month. Based on the potential to affect streamflow, 24 variables describing physiographic, anthropogenic, and climatic characteristics were chosen as predictors in the MLFLOW model—20 variables described static physiographic and anthropogenic conditions, and 4 variables described dynamic climatic conditions.

The dynamic and static variables were temporally and spatially conditioned to amplify the relation of predictor variables to monthly streamflow. To generate a diverse lagged effect of climatic conditions on streamflow, a temporal conditioning process was applied to the climate variables. The process consisted of moving averages of the time series data that ranged from the prior month to the prior year. Two geospatial processes were used to account for area and distance effects on streamflow. An area-averaged accumulation of the spatial data generated upstream effects of a variable. A distance-decayed accumulation of the spatial data generated more localized effects of a variable.

The MLFLOW model used a gradient boosting machine that is one of many machine learning models that are used in various hydrologic modeling, including modeling of

streamflow. A gradient boosting machine is a machine learning model used to solve regression or classification problems. The gradient boosting machine produces a predictive model that is an ensemble of many weaker models, which are typically implemented as decision trees. Development of the MLFLOW model proceeded in an iterative process of model calibration and validation. Streamflow observations used to fit the model were split into training and testing samples using repeated k -fold cross-validation.

The MLFLOW model was initially fitted to all data to explain monthly streamflow in relation to the predictor variables for all the study area and study period. Additional models were fitted, one model for each watershed and year to explain spatial or temporal variation in the relation of monthly streamflow to the predictor variables. Multiple models were fitted to qualitatively assess sensitivity of the MLFLOW model to variations of the predictor variables and streamflow observations. Models fitted to different combinations of predictor variables were used to assess sensitivity to selection and conditioning of the predictor variables. Model sensitivity to fitting data from a different area or period was assessed by differentially grouping streamflow observations by watershed or year for use in model training and testing. Model sensitivity to the quantity of data fitted was assessed by progressively reducing streamflow observations by percentage of sites or months available for use in model fitting.

The MLFLOW model fitted to all data had satisfactory agreement between observed and predicted streamflow (coefficient of determination [R^2]=0.80, Nash-Sutcliffe efficiency [NSE]=0.79, NSE with log-transformed data [logNSE]=0.82, and percent bias [PBIAS]=0.7 percent). The equivalence between NSE (0.79) and logNSE (0.82) indicated the MLFLOW model performed equally well for high and low flows. PBIAS (0.7 percent) indicated the MLFLOW model did not overpredict or underpredict monthly streamflow in general. The MLFLOW model performed equally well for all months with streamflow observations.

The most important variables (statistically important in the MLFLOW model) for explaining monthly streamflow were moving averages of precipitation and snow water equivalent. Importance of the static and dynamic variables did not differ substantially among the three watersheds but differed considerably among the 6 years. The 20 most important variables in the MLFLOW model had simple to more complex relations with monthly streamflow as interpreted with partial dependence plots. Monthly streamflow increased with increasing moving averages of precipitation and snow water equivalent and decreased with increasing moving averages of evapotranspiration and temperature. Monthly streamflow also increased with increasing drainage area and decreased with increasing forest cover and elevation.

Temporal and spatial conditioning intensified the relation of many predictor variables with monthly streamflow, resulting in more information that the MLFLOW model could use

for predicting streamflow. However, conditioning of the variables was less important than selection of dynamic variables for use in the MLFLOW model.

The MLFLOW model was most sensitive to selection of dynamic climatic variables. Unconditioned dynamic climatic variables alone explained 54 percent of the variance ($R^2=0.54$) in monthly streamflow, whereas adding static physiographic and anthropogenic variables only explained 12 percent more of the variance ($R^2=0.66$), indicating the greater importance of the time series data. Also, spatial conditioning of all variables together with temporal conditioning of dynamic variables increased the variance explained in the MLFLOW model by another 14 percent ($R^2=0.80$).

For models trained with all watersheds and years, performance was better in testing on observations from each watershed than from each year separately, indicating the MLFLOW model had greater sensitivity to temporal than to spatial differences in the data. However, in contrast, models trained with all except 1 year left out sequentially and tested on the left-out year performed better than models trained with all except one watershed left out sequentially and tested on the left-out watershed.

Performance was better for models fitted to fewer sites than to fewer months of observations, which indicated the MLFLOW model was more sensitive to temporal than to spatial differences in the data. The MLFLOW model fitted to all data performed very well, but performance in general progressively decreased for models fitted to progressively reduced percentages of sites or months.

Streamflow predictions seemed to well represent the annual hydrograph within the study area during the study period. Intra-annual variation in streamflow was simulated realistically by the MLFLOW model—seasonality of streamflow was well characterized. The MLFLOW model also realistically simulated interannual variation in streamflow, well representing yearly hydroclimatic conditions.

The greatest utility of the modeling approach is the ease of use and the speed of processing input data, running the model, and interpreting the model output, whereas the greatest limitation is the need for spatially and temporally representative streamflow observations to drive the model. Although familiarity with R is necessary, only a working knowledge of hydrology (for selecting appropriate predictor variables and evaluating the quality of streamflow observations) and a rudimentary understanding of machine learning models are needed. Therefore, this modeling approach is practicable for other scientists who work with water but who are not hydrologists.

References Cited

- Barnhart, T.B., Sando, R., Siefken, S.A., McCarthy, P.M., and Rea, A.H., 2020, Flow-conditioned parameter grid tools: U.S. Geological Survey software release, accessed May 6, 2021, at <https://doi.org/10.5066/P9W8UZ47>.
- Barrett, A.P., 2003, National operational hydrologic remote sensing center snow data assimilation system (SNODAS) products at NSIDC: National Snow and Ice Data Center Special Report 11, p. 19, accessed March 11, 2018, at https://nsidc.org/sites/nsidc.org/files/technical-references/nsidc_special_report_11.pdf.
- Bartolino, J.R., Konrad, C.P., Sando, R., and Hockman-Wert, D.P., 2019, RockType to permeability crosswalk table, Northwest U.S.: U.S. Geological Survey data release, accessed March 12, 2019, at <https://doi.org/10.5066/P95DIXXT>.
- Bellos, V., and Carbajal, J.P., 2020, Machine learning applied to hydraulic and hydrological modelling: special issue of *Water*, accessed March 13, 2020, at https://www.mdpi.com/journal/water/special_issues/Machine_Learning_Hydraulic_Hydrological.
- Bowen, Z.H., Aldridge, C.L., Anderson, P.J., Assal, T.J., Bern, C.R., Biewick, L.R.H., Boughton, G.K., Carr, N.B., Chalfoun, A.D., Chong, G.W., Clark, M.L., Fedy, B.C., Foster, K., Garman, S.L., Germaine, S., Hethcoat, M.G., Homer, C., Kauffman, M.J., Keinath, D., Latysh, N., Manier, D., McDougal, R.R., Melcher, C.P., Miller, K.A., Montag, J., Potter, C.J., Schell, S., Shafer, S.L., Smith, D.B., Sweat, M.J., and Wilson, A.B., 2014, U.S. Geological Survey science for the Wyoming Landscape Conservation Initiative—2012 annual report: U.S. Geological Survey Open-File Report 2014–1093, 71 p. [Also available at <https://doi.org/10.3133/ofr20141093>.]
- Carlisle, D.M., Wolock, D.M., Howard, J.K., Grantham, T.E., Fesenmyer, K., and Wiczorek, M., 2016, Estimating natural monthly stream flows in California and the likelihood of anthropogenic modification: U.S. Geological Survey Open File Report 2016–1189, 27 p. [Also available at <https://doi.org/10.3133/ofr20161189>.]
- Chaney, N.W., Minasny, B., Herman, J.D., Nauman, T.W., Brungard, C.W., Morgan, C.L.S., McBratney, A.B., Wood, E.F., and Yimam, Y., 2019, POLARIS soil properties—30-m probabilistic maps of soil properties over the contiguous United States: *Water Resources Research*, v. 55, no. 4, p. 2916–2938. [Also available at <https://doi.org/10.1029/2018WR022797>.]

- Chase, K.J., Haj, A.E., Regan, R.S., and Viger, R.J., 2016, Potential effects of climate change on streamflow for seven watersheds in eastern and central Montana: *Journal of Hydrology. Regional Studies*, v. 7, p. 69–81. [Also available at <https://doi.org/10.1016/j.ejrh.2016.06.001>.]
- Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., and Pasteris, P.P., 2008, Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States: *International Journal of Climatology*, v. 28, no. 15, p. 2031–2064. [Also available at <https://doi.org/10.1002/joc.1688>.]
- Esri, 2019, ArcGIS Desktop: Redlands, Calif., Environmental Systems Research Institute, ver. 10.6.1.
- Friedman, J.H., 2001, Greedy function approximation—A gradient boosting machine: *Annals of Statistics*, v. 29, no. 5, p. 1189–1232. [Also available at <https://doi.org/10.1214/aos/1013203451>.]
- Friedman, J.H., 2002, Stochastic gradient boosting: *Computational Statistics & Data Analysis*, v. 38, no. 4, p. 367–378. [Also available at [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).]
- Girard, C.E., and Walters, A.W., 2018, Evaluating relationships between native fishes and habitat in streams affected by oil and natural gas development: *Fisheries Management and Ecology*, v. 25, no. 5, p. 366–379. [Also available at <https://doi.org/10.1111/fme.12303>.]
- Greenwell, B., Boehmke, B., and Cunningham, J., 2020, gbm—Generalized boosted regression models: R package, ver. 2.1.8, accessed August 18, 2020, at <https://cran.r-project.org/web/packages/gbm>.
- 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.
- Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., and Megown, K., 2015, Completion of the 2011 national land cover database for the conterminous United States—Representing a decade of land cover change information: *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, p. 345–354. [Also available at <https://www.ingentaconnect.com/content/asprs/pers/2015/00000081/00000005/art00002>.]
- Jaeger, K.L., Sando, R., McShane, R.R., Dunham, J.B., Hockman-Wert, D.P., Kaiser, K.E., Hafen, K., Risley, J.C., and Blasch, K.W., 2019, Probability of streamflow permanence model (PROSPER)—A spatially continuous model of annual streamflow permanence throughout the Pacific Northwest: *Journal of Hydrology: X*, v. 2, p. 100005. [Also available at <https://doi.org/10.1016/j.hydroa.2018.100005>.]
- Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A.K., Hochreiter, S., and Nearing, G.S., 2019, Toward improved predictions in ungauged basins—Exploiting the power of machine learning: *Water Resources Research*, v. 55, no. 12, p. 11344–11354. [Also available at <https://doi.org/10.1029/2019WR026065>.]
- Kuhn, M., 2008, Building predictive models in R using the caret package: *Journal of Statistical Software*, v. 28, no. 5, p. 1–26. [Also available at <https://doi.org/10.18637/jss.v028.i05>.]
- Kuhn, M., 2020, caret—Classification and regression training: R package, ver. 6.0–86, accessed April 14, 2020, at <https://cran.r-project.org/web/packages/caret>.
- Kuhn, M., and Johnson, K., 2013, *Applied predictive modeling*: New York, Springer, 600 p.
- Leavesley, G.H., Lichty, R.W., Troutman, B.M., and Saindon, L.G., 1983, *Precipitation-runoff modeling system—User’s manual*: U.S. Geological Survey Water Resources Investigation Report 83–4238, 207 p. [Also available at <https://doi.org/10.3133/wri834238>.]
- 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>.]
- Markstrom, S.L., Regan, R.S., Hay, L.E., Viger, R.J., Webb, R.M.T., Payn, R.A., and LaFontaine, J.H., 2015, PRMS—IV, the precipitation-runoff modeling system, version 4: U.S. Geological Survey Techniques and Methods, book 6, chap. B7, 158 p. [Also available at <https://doi.org/10.3133/tm6B7>.]
- Massey, R., Sankey, T.T., Yadav, K., Congalton, R.G., Tilton, J.C., and Thenkabail, P.S., 2017, *Making Earth system data records for use in research environments (MEaSURES) global food security-support analysis data (GFSAD) cropland extent 2010 North America 30 m*: National Aeronautics and Space Administration, accessed May 14, 2019, at <https://doi.org/10.5067/MEaSURES/GFSAD/GFSAD30NACE.001>.
- McCabe, G.J., and Markstrom, S.L., 2007, A monthly water-balance model driven by a graphical user interface: U.S. Geological Survey Open-File Report 2007–1088, 6 p. [Also available at <https://doi.org/10.3133/ofr20071088>.]
- McShane, R.R., and Eddy-Miller, C.A., 2021, Input data, model output, and R scripts for a machine learning streamflow model on the Wyoming Range, Wyoming, 2012–17: U.S. Geological Survey data release, <https://doi.org/10.5066/P9XCP1AE>.

- Miller, M.P., Carlisle, D.M., Wolock, D.M., and Wieczorek, M., 2018, A database of natural monthly streamflow estimates from 1950 to 2015 for the conterminous United States: *Journal of the American Water Resources Association*, v. 54, no. 6, p. 1258–1269. [Also available at <https://doi.org/10.1111/1752-1688.12685>.]
- 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>.]
- Multi-Resolution Land Characteristics Consortium, 2017, National Land Cover Database (NLCD) 2011: Multi-Resolution Land Characteristics Consortium, accessed July 14, 2017, at <https://www.mrlc.gov>.
- Nash, J.E., and Sutcliffe, J.V., 1970, River flow forecasting through conceptual models, part 1—A discussion of principles: *Journal of Hydrology (Amsterdam)*, v. 10, no. 3, p. 282–290. [Also available at [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).]
- National Operational Hydrologic Remote Sensing Center, 2004, Snow data assimilation system (SNODAS) data: National Snow and Ice Data Center, accessed March 11, 2018, at <https://doi.org/10.7265/N5TB14TC>.
- O'Donnell, M.S., Fancher, T.S., Freeman, A.T., Ziegler, A.E., Bowen, Z.H., and Aldridge, C.L., 2014, Large scale Wyoming transportation data—A resource planning tool: U.S. Geological Survey Data Series 821, 21 p., accessed March 13, 2020, at <https://doi.org/10.3133/ds821>.
- Oriel, S.S., and Platt, L.B., 1980, Geologic map of the Preston 1 degree x 2 degree quadrangle, southeastern Idaho and western Wyoming: U.S. Geological Survey Miscellaneous Investigations Series Map I-1127, 1 sheet. [Also available at https://ngmdb.usgs.gov/Prodesc/proddesc_8978.htm.]
- POLARIS, 2019, 30-m probabilistic maps of soil properties over the contiguous United States: Durham, N.C., Duke University, accessed June 15, 2019, at www.polaris.earth.
- PRISM Climate Group, 2018, Parameter-elevation regressions on independent slopes model (PRISM) climate data: Corvallis, Oreg., Oregon State University, accessed March 11, 2018, at <https://prism.oregonstate.edu>.
- R Core Team, 2021, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, ver. 4.0.5., accessed March 4, 2021, at <https://www.R-project.org>.
- Rantz, S.E., and others, 1982, Measurement and computation of streamflow: U.S. Geological Survey Water-Supply Paper 2175, 631 p. [Also available at <https://doi.org/10.3133/wsp2175>.]
- Ridgeway, G., 2020, Generalized boosted models—A guide to the gbm package: R vignette, accessed August 18, 2020, at <https://cran.r-project.org/web/packages/gbm/vignettes>.
- Sauer, V.B., and Turnipseed, D.P., 2010, Stage measurement at gaging stations: U.S. Geological Survey Techniques and Methods book 3, chap. A7, 45 p. [Also available at <https://doi.org/10.3133/tm3A7>.]
- Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., and Verdin, J.P., 2013, Operational evapotranspiration mapping using remote sensing and weather datasets—A new parameterization for the SSEB approach: *Journal of the American Water Resources Association*, v. 49, no. 3, p. 577–591. [Also available at <https://doi.org/10.1111/jawr.12057>.]
- Shen, C., Elshorbagy, A., Gupta, H., and Nearing, G., 2018, Big data and machine learning in water sciences—Recent progress and their use in advancing science: special section of *Water Resources Research*, accessed March 13, 2020, at [https://agupubs.onlinelibrary.wiley.com/doi/toc/10.1002/\(ISSN\)1944-7973.MACHINELEARN](https://agupubs.onlinelibrary.wiley.com/doi/toc/10.1002/(ISSN)1944-7973.MACHINELEARN).
- Soil Survey Staff, 2016, U.S. general soil map (STATSGO2): Natural Resources Conservation Service, accessed February 11, 2019, at <https://gdg.sc.egov.usda.gov>.
- Soil Survey Staff, 2017, Gridded soil survey geographic (gSSURGO) database for the conterminous United States: Natural Resources Conservation Service, accessed May 14, 2019, at <https://gdg.sc.egov.usda.gov>.
- Stoeser, D.B., Green, G.N., Morath, L.C., Heran, W.D., Wilson, A.B., Moore, D.W., and Van Gosen, B.S., 2005, Preliminary integrated geologic map databases for the United States—Central States—Montana, Wyoming, Colorado, New Mexico, Kansas, Oklahoma, Texas, Missouri, Arkansas, and Louisiana: U.S. Geological Survey Open-File Report 2005–1351, digital data, accessed March 10, 2017, at <https://doi.org/10.3133/ofr20051351>.
- Tarboton, D.G., 2016, Terrain analysis using Digital Elevation Models (TauDEM): Logan, Utah, Utah State University, ver. 5.3.7, accessed December 19, 2017, at <https://hydrology.usu.edu/taudem>.
- U.S. Geological Survey, 2016, Watershed boundary dataset: U.S. Geological Survey digital data, accessed September 15, 2016, at <https://nrcs.app.box.com/v/gateway/folder/18546994164>.

- U.S. Geological Survey, 2018, Simplified surface energy balance (SSEBop) actual evapotranspiration data for the conterminous U.S.: U.S. Geological Survey data release, accessed April 12, 2018, at <https://www.sciencebase.gov/catalog/item/54dd5d21e4b08de9379b38b6>.
- U.S. Geological Survey, 2019a, National Hydrography Dataset Plus: U.S. Geological Survey digital data, accessed February 11, 2019, at <https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus>.
- U.S. Geological Survey, 2019b, National Hydrography Dataset Plus High Resolution: U.S. Geological Survey digital data, accessed May 14, 2019, at <https://www.usgs.gov/core-science-systems/ngp/national-hydrography/access-national-hydrography-products>.
- U.S. Geological Survey, 2021, USGS surface-water data for Wyoming, *in* USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed June 16, 2020, at <https://doi.org/10.5066/F7P55KJN>. [State information directly accessible at <https://waterdata.usgs.gov/wy/nwis/sw>.]
- Walker, R.H., Girard, C.E., Alford, S.L., and Walters, A.W., 2020, Anthropogenic land-use change intensifies the effect of low flow on stream fishes: *Journal of Applied Ecology*, v. 57, no. 1, p. 149–159. [Also available at <https://doi.org/10.1111/1365-2664.13517>.]
- Walters, A.W., Girard, C.E., Walker, R.H., Farag, A.M., and Alvarez, D.A., 2019, Multiple approaches to surface water quality assessment provide insight for small streams experiencing oil and natural gas development: *Integrated Environmental Assessment and Management*, v. 15, no. 3, p. 385–397. [Also available at <https://doi.org/10.1002/ieam.4118>.]
- Wolock, D.M., 2003, Base-flow index grid for the conterminous United States: U.S. Geological Survey Open-File Report 2003–263, digital data, accessed February 10, 2018, at <https://doi.org/10.3133/ofr03263>.
- Wyoming State Engineer's Office, 2016, SEO Wells: Wyoming State Engineer's Office, accessed July 15, 2018, at <https://sites.google.com/a/wyo.gov/seo/seo-files>.
- Wyoming State Geological Survey, 2014, Wyoming bedrock geology: Wyoming State Geological Survey digital data, accessed May 14, 2019, at <https://www.wsgs.wyo.gov/pubs-maps/gis.aspx>.
- Wyoming State Geological Survey, 2015, Wyoming surficial geology: Wyoming State Geological Survey digital data, accessed May 14, 2019, at <https://www.wsgs.wyo.gov/pubs-maps/gis.aspx>.
- Wyoming Water Development Office, 2007, Points of diversions: Wyoming Water Development Office digital data, accessed July 15, 2018, at <https://waterplan.state.wy.us/plan/statewide/2007/gis>.
- Zielinski, G.W., DeCoursey, G.M., Drahovzal, J.A., and Ruperto, J.M., 1985, Hydrothermics in the Wyoming Overthrust Belt—United States: *The American Association of Petroleum Geologists Bulletin*, v. 69, no. 5, p. 699–709.

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
Lafayette and Rolla Publishing Service Centers

