USGS
science for a changing world

Prepared in cooperation with the U.S. Environmental Protection Agency

Estimating Metal Concentrations with Regression Analysis and Water-Quality Surrogates at Nine Sites on the Animas and San Juan Rivers, Colorado, New Mexico, and Utah

Scientific Investigations Report 2018–5116

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

Cover: Background photo of Animas River at Durango, Colorado, photograph by Kevin Murphy, U.S. Geological Survey.
Left photo of San Juan River at Bluff, Utah, photograph by Christopher Wilkowske, U.S. Geological Survey.
Upper right photo of Animas River below Silverton, Colorado, photograph by Kevin Murphy, U.S. Geological Survey.
Lower right photo of Cement Creek at Silverton, Colorado, photograph by Kevin Murphy, U.S. Geological Survey.

Estimating Metal Concentrations with Regression Analysis and Water-Quality Surrogates at Nine Sites on the Animas and San Juan Rivers, Colorado, New Mexico, and Utah

By M. Alisa Mast

Prepared in cooperation with the
U.S. Environmental Protection Agency

Scientific Investigations Report 2018–5116

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

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

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

U.S. Geological Survey, Reston, Virginia: 2018

Suggested citation:
Mast, M.A., 2018, Estimating metal concentrations with regression analysis and water-quality surrogates at nine sites on the Animas and San Juan Rivers, Colorado, New Mexico, and Utah: U.S. Geological Survey Scientific Investigations Report 2018–5116, 68 p., https://doi.org/10.3133/sir20185116.

ISSN 2328-0328 (online)

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Length	
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)

Datum

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD88).

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD83).

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

Supplemental Information

Specific conductance is given in microsiemens per centimeter at 25 degrees Celsius (µS/cm at 25 °C).

Concentrations of chemical constituents in water are given in either milligrams per liter (mg/L) or micrograms per liter (µg/L).

Abbreviations

EPA	U.S. Environmental Protection Agency
GKM	Gold King Mine
MAS	model archive summary
mi	mile
NRTWQ	National Real-Time Water Quality
R^2	coefficient of determination
SC	specific conductance
USGS	U.S. Geological Survey
WQP	Water Quality Portal
µg/L	microgram per liter

Estimating Metal Concentrations with Regression Analysis and Water-Quality Surrogates at Nine Sites on the Animas and San Juan Rivers, Colorado, New Mexico, and Utah

By M. Alisa Mast

Abstract

The purpose of this report is to evaluate the use of site-specific regression models to estimate metal concentrations at nine U.S. Geological Survey streamflow-gaging stations on the Animas and San Juan Rivers in Colorado, New Mexico, and Utah. Downstream users could use these regression models to determine if metal concentrations are elevated and pose a risk to water supplies, agriculture, and recreation. Multiple linear-regression models were developed by relating metal concentrations in discrete water-quality samples to continuously monitored streamflow and surrogate parameters (specific conductance, pH, turbidity, and water temperature) collected at the U.S. Geological Survey stations. Models were developed for dissolved and total concentrations of aluminum, arsenic, cadmium, copper, iron, lead, manganese, and zinc using water-quality samples collected from 2005 to 2017 by several Federal, State, Tribal, and local agencies using different collection methods and analytical laboratories. Model performance varied but, in general, models for dissolved metals did not perform as well as those for total metals. Dissolved metals generally were correlated to specific conductance or streamflow and total metals generally were better correlated with turbidity.

Explanatory variables in the models reflected hydrologic and geochemical processes within the basin. A larger number of regression models were statistically significant for the most upstream sites, where metal concentrations were elevated by drainage from abandoned mines and mineralized bedrock. Models generally did not perform as well at downstream sites, especially for dissolved metals, which occurred at lower concentrations than at the upstream sites. In the lower reaches of the rivers, the input of more alkaline water from tributaries and groundwater reduced metal solubility and diluted metal concentrations. The number and distribution of samples in the calibration datasets also may have been a factor in model development. At some sites on the San Juan River, calibration datasets were more limited and did not represent the full range of observed hydrologic and water-quality conditions, especially during storm events in summer and fall. Recommendations for model use are given based on estimates of model precision, biases, and adequacy of the calibration datasets in terms of the number of samples and representativeness of the observed range of streamflow and water-quality conditions.

Introduction

On August 5, 2015, about 3 million gallons of low pH and metal-rich water and sediment were released from the Gold King Mine (GKM) near Silverton, Colorado, into Cement Creek, a tributary of the Animas River (fig. 1). The plume of contaminated water and sediment quickly reached the Animas River, where it turned the river a yellow color that intensified as it traveled more than 100 miles (mi) downstream along the entire length of the Animas River (Sullivan and others, 2017). Concentrations of some metals in the Animas River particularly aluminum, arsenic, copper, and iron increased by more than two orders of magnitude as the plume passed the city of Durango, Colo. Within 3 days of the GKM release, the plume entered the San Juan River at Farmington, New Mexico, and continued westward until eventually discharging into Lake Powell in Utah. Within 2 weeks, the water quality of the Animas and San Juan Rivers returned to pre-event conditions (Sullivan and others, 2017).

Following the GKM release, there was heightened concern by downstream communities and Tribes in Colorado, New Mexico, and Utah about the risks elevated metals posed to public and private drinking-water supplies, water for irrigation and ranching, and recreational fishing and boating activities. Representatives from the concerned entities identified monitoring of basic water-quality parameters and metals in the Animas and San Juan Rivers as a priority for characterizing and understanding the water quality of the rivers. In response to this need, the U.S. Geological Survey (USGS), in cooperation with the U.S. Environmental Protection Agency

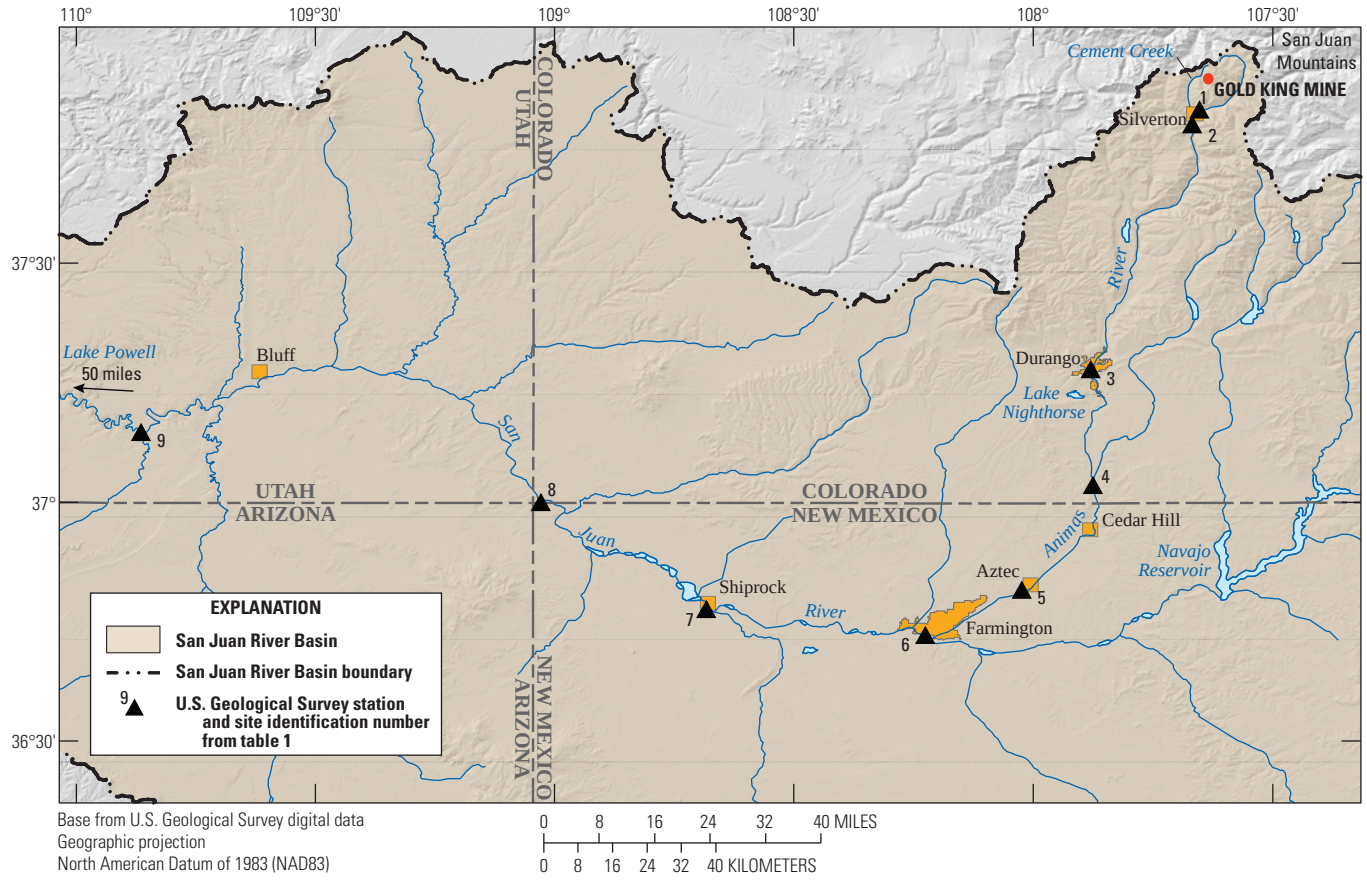

Figure 1. Locations of nine U.S. Geological Survey streamflow-gaging stations with continuous water-quality sensors on the Animas and San Juan Rivers in Colorado, New Mexico, and Utah.

(EPA), agreed on a scope of work to install and operate continuous water-quality sensors and use the data obtained from the sensors to develop regression models that correlate one or more parameters collected by the sensors (referred to as "surrogates") with dissolved (filtered) and total (unfiltered) concentrations of key metals of concern. Regression models would allow communities and tribes to evaluate the potential for elevated metal concentrations in the Animas and San Juan Rivers on a continuous and nearly real-time basis. Although numerous studies have used water-quality surrogates to predict suspended sediment concentrations, nutrients, and bacterial counts (Rasmussen and others, 2009; Schilling and others, 2017; Senior, 2017), relatively few studies have investigated their use for predicting metal concentrations in rivers (Nasrabadi and others, 2016; Aulenbach and others, 2017).

Purpose and Scope

The purpose of this report is to evaluate the use of site-specific regression models based on statistical relations between continuous and discrete water-quality and streamflow data collected at nine USGS stations along the Animas and San Juan Rivers to estimate concentrations of selected metals. The results of this study could be used to document regression equations deemed statistically valid to be reported on the USGS National Real-Time Water Quality (NRTWQ) website (http://nrtwq.usgs.gov/). Recommendations for model use are given based on estimates of model precision, biases, and adequacy of the calibration datasets in terms of the number of samples and representativeness of the observed range of streamflow and water-quality conditions. This study also identifies gaps in currently (2018) available datasets that need to be addressed to improve regression models at sites where adequate models could not be developed.

This study used data from nine USGS streamflow-gaging stations (herein referred to as "USGS stations"): (1) Cement Creek at Silverton, Colo. (USGS station 09358550); (2) Animas River below Silverton, Colo. (USGS station 09359020); (3) Animas River at Durango, Colo. (USGS station 09361500); (4) Animas River near Cedar Hill, N.Mex. (USGS station 09363500); (5) Animas River below Aztec, N. Mex. (USGS station 09364010); (6) San Juan River at Farmington, N. Mex. (USGS station 09365000), (7) San Juan River at Shiprock, N. Mex. (USGS station 09368000);

(8) San Juan River at Four Corners, Colo. (USGS station 09371010); and (9) San Juan River near Bluff, Utah (USGS station 09379500) (fig. 1). Although streamflow has been measured at all nine stations for many decades, continuous water-quality monitors that measure water temperature, specific conductance (SC), pH, and turbidity were not installed until April 2016, approximately 9 months after the GKM release. Metal concentration data for discrete water-quality samples collected at sites in the vicinity of each USGS station were compiled. Samples were collected by various agencies using different collection methods and analyzed by different laboratories. Regression models were evaluated for dissolved and total fractions of aluminum, arsenic, cadmium, copper, iron, lead, manganese, and zinc at each of the nine sites. Data used for the analyses described in this report were collected from 2005 through 2017.

Study Area Description

The area for this study includes sections of Cement Creek, the Animas River, and the San Juan River flowing through Colorado, New Mexico, and Utah (fig. 1). The GKM is located in the subalpine zone of the San Juan Mountains at an altitude of 11,400 feet and drains into Cement Creek, which is a tributary of the Animas River. Below the confluence with Cement Creek, the Animas River flows south for more than 100 mi through increasingly arid landscapes before joining the San Juan River near Farmington, N. Mex., at an altitude of 5,230 feet. The San Juan River then flows west-northwest through desert landscapes of the Colorado Plateau and eventually discharges into Lake Powell, approximately 50 mi downstream from Bluff, Utah. Annual streamflow in the Animas and San Juan Rivers is dominated by spring snowmelt and streamflows typically peak in May or June (fig. 2). Flows in the Animas River are largely unregulated whereas flows in the San Juan River are partially regulated by the Navajo Reservoir, 44 mi upstream from Farmington, N. Mex. Forest is the dominant land cover upstream from Durango, Colo., and desert scrubland, rangeland, and agricultural land dominate in the lower part of the study area (Miller and others, 2013).

The Animas River is used for water supply for the city of Durango, Colo., and numerous irrigation ditches serve the surrounding farmland along the river. Water also is pumped from the Animas River for storage in Lake Nighthorse. Major water uses of the San Juan River include municipal and private drinking water supplies, irrigation, and ranching. Some of the most intense agriculture is on tribal lands along the San Juan River near the community of Farmington, N. Mex., which includes nearly 70,000 acres of irrigated farmland (Nania and others, 2014).

The headwaters of Cement Creek and the Animas River upstream from Silverton, Colo., are underlain by highly mineralized bedrock and the area has a long history of metal mining (Church and others, 2007). Because of drainage from both abandoned mine lands and mineralized bedrock, surface water is characterized by elevated metal concentrations and low pH. Metal concentrations at the Cement Creek site (site 1), which receives drainage from the GKM, are the highest and pH values are the lowest (median pH=4.3) (Wright and others, 2007) of the nine sites included in this report. The Animas River below Silverton site (site 2) has a more circumneutral pH (median pH=6.8) (fig. 2) than Cement Creek because of the greater proportion of non-mineralized bedrock in its watershed and as a result, concentrations of most metals are 2 to 10 times more dilute than in Cement Creek (Wright and others, 2007).

Downstream from Silverton, bedrock is composed primarily of igneous and metamorphic rock types that are more resistant to weathering and contribute circumneutral surface water to the river with only a small amount of sediment and dissolved solids causing pH to increase in the Animas River at Durango (site 3) with relatively little change in SC and turbidity (fig. 2) (Church and others, 1997). Metal concentrations in the Animas River continue to decline downstream from Silverton with most metal concentrations being 2 to 10 times lower at Durango. Downstream from Durango, the bedrock changes to more easily weathered sedimentary rock types that contribute dissolved solids and suspended sediment to the river, which cause increases in pH, SC, and turbidity as the river flows downstream to Bluff, Utah (site 9) (fig. 2). Most metal concentrations continue to decline downstream from Durango; however, total aluminum, arsenic, and iron concentrations gradually increase as more sediment is contributed to the river.

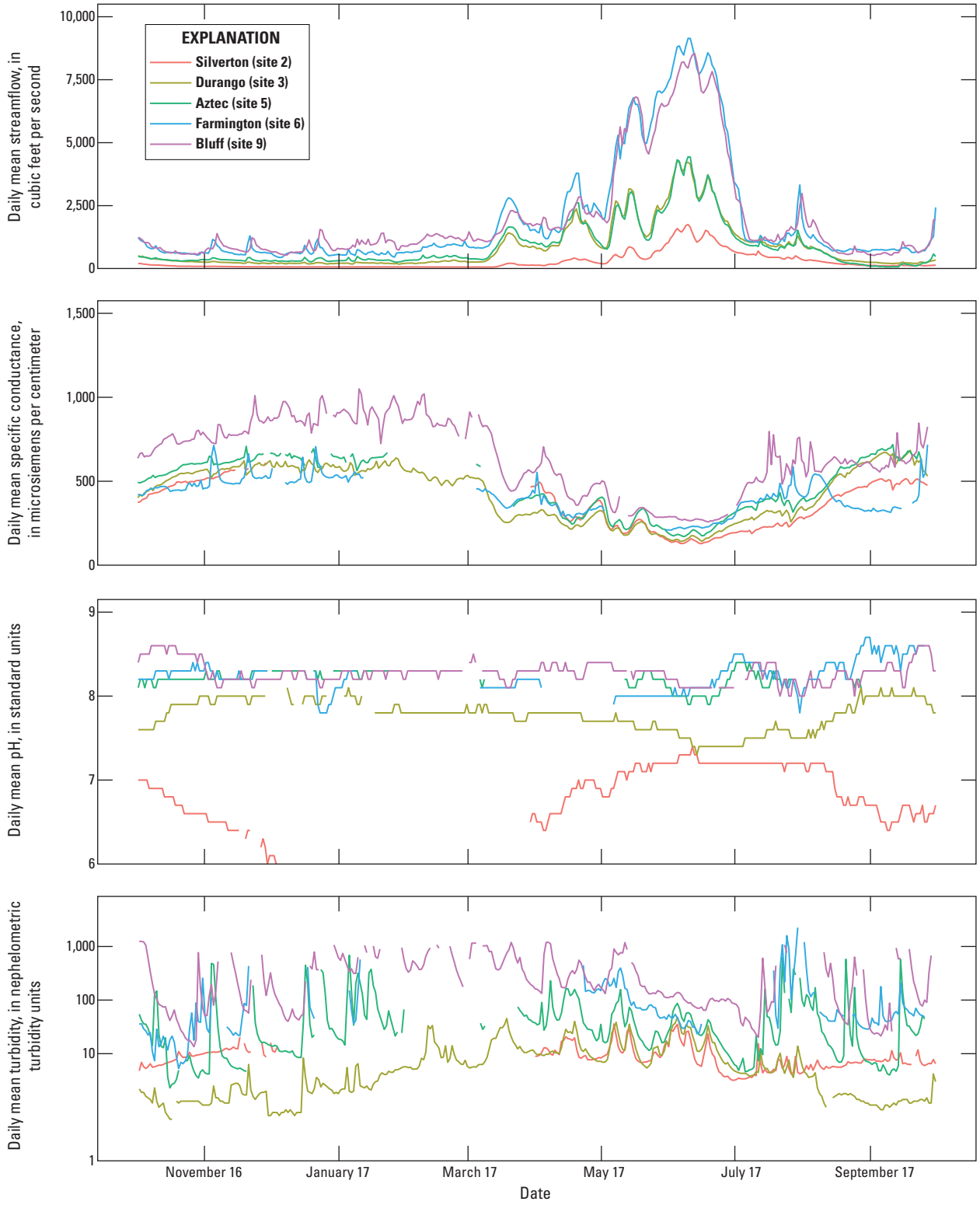

Figure 2. Daily mean streamflow, specific conductance, pH, and turbidity for the Cement Creek site, two sites on the Animas River, and two sites on the San Juan River for 2017. Data from U.S. Geological Survey National Water Information System, http://dx.doi.org/10.5066/F7P55KJN.

Approach and Methods

The general approach used in this study was to compile discrete water-quality samples with laboratory analyzed metal concentrations and concurrent measurements of surrogate parameters (streamflow, SC, pH, turbidity, and water temperature) and then to develop statistical models to relate the metal concentrations to the surrogate parameters. All of the nine sites had some water-quality concentration data for the metals of interest, which are dissolved and total fractions of aluminum, arsenic, cadmium, copper, iron, lead, manganese, and zinc. Surrogate data were collected manually at the time of sampling, obtained from the continuous water-quality sensors that began operation in April 2016, or were determined in the laboratory (SC and pH only). Linear-regression techniques were used to develop models that were statistically valid and physically meaningful for each site.

Compilation of Discrete and Continuous Water-Quality Data

Available data on water-quality samples that were collected through the end of 2017 in the vicinity of each USGS station were retrieved from the Water Quality Portal (WQP) (https://www.waterqualitydata.us/), which is a cooperative service that provides public access to water-quality data collected by local, State, Tribal, and Federal agencies. All USGS water-quality data retrieved through the WQP originate from the USGS National Water Information System (http://dx.doi.org/10.5066/F7P55KJN). Additional water-quality data for samples collected after the GKM spill were obtained from the EPA (Kate Sullivan, EPA, written commun., 2017). The data obtained from both sources (WQP and EPA) are available as a USGS data release of the model archive summaries (MASs; Mast, 2018). Data obtained from the WQP were from samples collected by various agencies using different collection methods and different analytical laboratories. A list of location identifiers and associated collection agencies for water-quality samples used in the regression analysis are presented in table 1, and maps showing sampling locations for each site are in the appendix. Data were only included in the analysis if the sample fraction (dissolved versus total), the parameter name, units of measurement, and remark codes were unambiguous.

For this study, the water-quality datasets were reviewed at a relatively basic level because of limited quality-assurance samples and the lack of other metadata, such as sampling and laboratory methods. A comparison of dissolved and total concentrations was made for samples that had both analytical results to ensure that the results made physical sense (for example, dissolved concentration not exceeding total concentration) and that the concentrations appeared to be reported using the correct units. All metal concentrations of 0 micrograms per liter (µg/L) were removed from the dataset because it was not clear if the 0 µg/L was due to missing data or was a nondetect with an unknown censoring level. Seasonal patterns in concentrations also were reviewed to help identify questionable or inconsistent data. Comparability of data from different sources was mainly evaluated by comparing plots of metal concentrations versus SC or streamflow among the different sampling locations in the vicinity of each USGS station. Because the quality-assurance review was limited, the datasets may contain errors that affect the precision and accuracy of the statistical results.

USGS-approved daily and unit-value (15-minute) streamflow, water temperature, pH, SC, and turbidity were obtained for the nine USGS stations from the USGS National Water Information System at http://dx.doi.org/10.5066/F7P55KJN. The operation of the water-quality sensors and streamflow gages and subsequent processing of the measured data followed USGS guidelines and standard operating procedures as documented in Wagner and others (2006) and Rantz and others (1982). Periods of missing data in the continuous record were not estimated or interpolated for this study. Because many of the discrete samples did not have the surrogate parameters (streamflow, water temperature, pH, SC, and turbidity) needed for the modeling, each water-quality sample was merged with the daily and unit-value streamflow and water-quality data from the USGS gages. Although streamflow has been measured at all nine stations for many decades, continuous water-quality monitors that measure water temperature, SC, pH and turbidity were not installed until April 2016, approximately 9 months after the GKM release. Therefore, samples collected prior to this time without surrogate parameters could not be used in the modeling, which greatly reduced the size of the calibration dataset at most sites.

A substantial fraction (up to 90 percent) of some dissolved metals were reported as censored values, which occurs when concentrations fall below the reporting limits of the analytical methods. Censored values were initially retained in the dataset along with the associated remark code. Samples collected during the 3-month period following the GKM spill were removed from the dataset because post-spill metal concentrations were elevated well above the normal concentration range and would strongly influence the regression model if included. Data retrievals from WQP and the National Water Information System and the merging of this data with the continuous data were accomplished using functions in the USGS smwrBase package in R (Lorenz, 2015).

Regression Analysis

Regression models were developed to relate dissolved and total aluminum, arsenic, cadmium, copper, iron, lead, manganese, and zinc to surrogate measurements from continuous monitors following guidelines from Helsel and Hirsch (2002) and Rasmussen and others (2009). The response variable was either concentration or log-transformed concentration. The explanatory variables included streamflow, log-transformed streamflow, SC, log-transformed SC, pH, and water temperature. Turbidity and log-transformed turbidity, which

Table 1. Agency location identifiers and location names from the Water Quality Portal (https://www.waterqualitydata.us/) for the nine U.S. Geological Survey (USGS) streamflow-gaging stations and other water-quality sampling sites included in this study.

[CO, Colorado; CDPHE, Colorado Department of Public Health and Environment; Ck, creek; blw, below; bfr, before; R, river; EPA8, U.S. Environmental Protection Agency, Region 8; St, street; @, at; abv, above; conf, confluence; CR, county road; NM, New Mexico; NNEPA, Navajo Nation Environmental Protection Agency; xing, crossing]

Site identification number for USGS site (figure 1)	Agency	Location identifier	Location name
1	USGS	USGS-09358550	Cement Creek at Silverton, CO
	CDPHE	21COL001_WQX-CEM49	Cement Ck blw Silverton just bfr Animas R
	EPA8	USEPA_REGION8-CC48	CC48
2	USGS	USGS-09359020	Animas River below Silverton, CO
	CDPHE	21COL001-WQX-000082	Animas River near Silverton
	EPA8	USEPA_REGION8-A72	Animas Gauge below Silverton
3	USGS	USGS-09361500	Animas River at Durango, CO
	CDPHE	21COL001_WQX-9423A	Animas River at Durango—9th St Bridge
	CDPHE	21COL001_WQX-9423	Animas River @ Durango, abv conf with Junction Ck
	EPA8	USEPA_REGION8-GKM05	GKM05
	EPA8	USEPA_REGION8-Rotary Park	Rotary Park
	EPA8	USEPA_REGION8-32nd St Bridge	32nd St Bridge
4	USGS	USGS-09363500	Animas River near Cedar Hill, NM
	Southern Ute Tribe	SOUTHUTE-AR 2-7	Animas River 2-7
	EPA8	USEPA_REGION8-AR2-7a	AR2-7A
	EPA8	USEPA_REGION8-AR7-2	AR7-2
5	USGS	USGS-09364010	Animas River below Aztec, NM
	National Park Service, Water Resources Division	11NPSWRD-AZRU_L1_ANIMASR	Animas River
	New Mexico Environmental Department	21NMEX_WQX-66Animas028.1	Animas River above Estes Arroyo—66ANIMAS028.1
	New Mexico Environmental Department	21NMEX_WQX-66Animas017.4	Animas @ CR 350 Bridge—66ANIMAS017.4
	City of Farmington, New Mexico	City of Farmington-APS #2	APS #2
	EPA8	USEPA_REGION8-ADW-010	ADW-010
	EPA8	USEPA_REGION8-FW-012	FW-012

Table 1. Agency location identifiers and location names from the Water Quality Portal (https://www.waterqualitydata.us/) for the nine U.S. Geological Survey (USGS) streamflow-gaging stations and other water-quality sampling sites included in this study.—Continued

[CO, Colorado; CDPHE, Colorado Department of Public Health and Environment; Ck, creek; blw, below; bfr, before; R, river; EPA8, U.S. Environmental Protection Agency, Region 8; St, street; @, at; abv, above; conf, confluence; CR, county road; NM, New Mexico; NNEPA, Navajo Nation Environmental Protection Agency; xing, crossing]

Site identification number for USGS site (figure 1)	Agency	Location identifier	Location name
6	USGS	USGS-09365000	San Juan River at Farmington, NM
	NNEPA	NNEPA-10SANJUANR37	10SANJUANR37
	EPA8	USEPA_REGION8-LVW-020	LVW-020
	EPA8	USEPA_REGION8-SJLP	SJLP
7	USGS	USGS-09368000	San Juan River at Shiprock, NM
	USGS	USGS-364446108321510	San Juan River above Hogback Diversion Channel
	NNEPA	NNEPA-10SANJUANR36	10SANJUANR36
	EPA8	USEPA_REGION8-SJSR	SJSR
8	USGS	USGS-09371010	San Juan River at Four Corners, CO
	Utah Division of Water Quality	UTAHDWQ_WQX-4954000	San Juan R at US160 xing in CO
	Ute Mountain Ute Tribe	UTEMTN-SJ-4C	San Juan @ USGS gaging station
	NNEPA	NNEPA-02SANJUANR06	02SANJUANR06
	EPA8	USEPA_REGION8-SJ4C	SJ4C
9	USGS	USGS-09379500	San Juan River near Bluff, UT
	NNEPA	NNEPA-29SANJUANR05	29SANJUANR05
	EPA8	USEPA_REGION8-SJMH	SJMH

have been shown to be reliable surrogates for total metals (Nasrabadi and others, 2016), were only included in the models for total metals. Because samples without turbidity could still be modeled for dissolved metals, the number of samples in the calibration datasets was generally different for dissolved than for total metals. Seasonal terms were also included as explanatory variables to account for seasonality not captured by variations in streamflow (Helsel and Hirsh, 2002). Two pairs of seasonal terms were calculated, $\sin(2\pi T*k)$ and $\cos(2\pi T*k)$, where T is the decimal portion of the year starting on January 1 and k equals 1 or 2 to account for annual or 6-month cycles, respectively (Helsel and Hirsh, 2002). These seasonal terms provide a mathematical way to fit the magnitude and phase shift of a sine function in a multiple linear-regression equation, as opposed to fitting a non-linear model to fit the phase shift (Stolwijk and others, 1999).

To keep the models fairly simple and avoid overparameterization, models were limited to no more than two measured explanatory variables and one pair of seasonal terms. The best regression model among all possible combinations of the explanatory variables was selected using the R package AICcmodavg, which ranks models based on the small-sample corrected Akaike information criterion (Mazerolle, 2017). During model development, some outliers were removed from

the nine datasets based on residual plots and outlier statistics (leverage and Cook's D) including 1, 0.1, 7, 6, 1, 4, 1, 0, and 2 percent of concentrations for sites 1–9, respectively. Details of outlier removal are provided in the MAS for each site (Mast, 2018). Regression models were not computed for datasets with fewer than 20 samples to avoid poorly representative models, especially because sampling was not targeted for model development (Aulenbach, 2013).

Statistical measures of fit included the coefficient of determination (R^2), adjusted R^2, residual standard error, and model standard percentage error. Models were considered suitable to use for constituent concentration computations if the amount of variability explained by the model (adjusted R^2) was greater than 50 percent. Each model was also evaluated using diagnostic plots including residuals versus predicted, measured versus predicted, and normal-probability plots of residuals. For regression equations using log-transformed metal concentrations, the predicted concentrations were corrected for transformation bias using the Duan Smearing factor (Duan, 1983) in the R package smwrStats. For metals with a substantial fraction of censored observations (greater than 20 percent), comparisons were made between models developed using a censored regression approach that uses the maximum likelihood method to estimate the regression parameters (R package cenReg) and linear models developed with the censored values removed. The censored regression models did not substantially change the model form or improve the fit, so it was decided to remove all censored values from the calibration datasets and use the standard linear-regression approach. Regressions were run without the censored values for all metals except those with greater than 70 percent censored values or fewer than 20 samples, which were not modeled. An R-script developed by the USGS Kansas Water Science Center was used to generate a MAS for each metal (https://nrtwq.usgs.gov/co/methods/). The MAS provides information about the regression model selected on the basis of best-fit metrics and review of residual distributions as calculated using the R-script, as well as the dataset used to develop the regression model (Mast, 2018).

Concentrations and the 90-percent prediction intervals were computed using the best regression model for each metal applied to the continuous water-quality and streamflow records for the period April 2016 through December 2017. Concentrations were not computed below the lowest measured value (to prevent prediction of negative values) or more than 1.2 times the maximum concentration in the calibration dataset to avoid extrapolating too far above the range of concentrations in the calibration dataset. The prediction interval is a range in which there is a 90-percent certainty that the true metal concentration occurs. The larger the prediction interval, the more uncertainty there is associated with the computed metal concentration. A median prediction interval that exceeded the median predicted concentration by more than 100 percent was set as a secondary condition for evaluating model reliability, which is sometimes referred to as the "prediction interval" in the tables and report. In this report, no attempt was made to account for error in the surrogate measurements, which if included and propagated through the computations, could substantially increase the uncertainty of these estimates.

Estimating Metal Concentrations with Regression Analysis and Water-Quality Surrogates

Regression models were developed for nine USGS stations to relate dissolved and total metals to continuously monitored surrogates. Model performance varied but, in general, models for total metals performed better than the models for dissolved metals (table 2). Specific conductance was the most common explanatory variable in dissolved metal models and turbidity was the most common explanatory variable in total metal models (table 2). Because SC is a measure of total dissolved solids in the water, it makes sense that SC is the best surrogate for dissolved metal concentrations. Similarly, because turbidity reflects increased particulates in water, it provides the best surrogate for total metal concentrations. Streamflow, pH, and water temperature were less important but occurred at a similar frequency in both the dissolved and total models. The majority of models contained seasonal terms indicating streamflow and temperature were not always sufficient for capturing seasonality in metal concentrations.

The regression models did not include a time term mainly because of the short duration of the calibration datasets. If stream chemistry is changing over time, the regression models could produce biased estimates when applied to future datasets. Continued water sampling would be needed to verify that the regression model remains valid beyond the period of record of the calibration dataset (Rasmussen and others, 2009).

Table 2. Summary of explanatory variables (including log transformed variables) included in statistically significant regression models for dissolved and total metals at the nine study sites.

	Dissolved metals	Total metals
Percent of statistically significant models (total of 72)	35	92
Percent of models with turbidity variables	—[a]	80
Percent of models with specific conductance variables	72	26
Percent of models with streamflow variables	32	42
Percent of models with pH variables	20	20
Percent of models with water temperature variables	28	18
Percent of models with seasonal variables	88	70

[a]Turbidity was not included as an explanatory variable in models for dissolved metals.

The results presented herein could be used to document regression equations deemed adequate to be reported on the USGS NRTWQ website (http://nrtwq.usgs.gov/). The NRTWQ website provides hourly computed concentrations and loads for salinity, selenium, sediment, and other constituents at selected stream sites across the United States. The information on the NRTWQ site is used by State and Federal agencies to make decisions applicable to drinking water, water treatment, regulatory programs, recreation, and public safety.

The results of the regression analysis for the nine USGS stations are discussed separately in the following sections of this report. Each section includes a table of summary statistics for the final calibration dataset and resulting regression equations and statistics (see table 3 for example) and two sets of graphs. In the summary table, the explanatory variables are listed in order of increasing *p*-value, although the seasonal terms are always kept together regardless of the *p*-value of the less significant term. The *p*-value for each term tests the null hypothesis that the coefficient is equal to zero or has no effect. A low *p*-value (less than 0.05) indicates that the null hypothesis can be rejected. The first set of graphs shows predicted versus measured concentrations for samples in the calibration dataset (see fig. 3 for example). For completeness, predicted concentrations for models with R^2 less than 0.50 are included in the graphs but the symbols are shown in a lighter shade of color to indicate the models that are less reliable. The second set of graphs shows daily mean concentrations predicted using the corresponding regression equations applied to the continuous streamflow and water-quality records from the USGS stations (see fig. 4 for example). In these graphs, the modeled concentrations are shown by the blue symbols, the associated 90-percent prediction intervals are indicated by the gray bands, measured concentrations in discrete samples are shown by the red symbols, and the symbols are in a lighter shade of color for models with R^2 less than 0.50. The last column on the summary tables is the "prediction interval," which is the percent of the median estimated concentration computed from the data displayed in the second set of graphs. Model archive summaries, including diagnostic plots and complete calibration datasets for models, are in Mast (2018) and sampling site location maps are in the report appendix.

Cement Creek at Silverton, Colorado (Station 09358550)

The dataset compiled for this site contained 636 water-quality samples collected in the vicinity of the USGS station at Cement Creek at Silverton (site 1 in fig. 1) with dissolved or total metal analyses for 1971–2017. Because of the lack of field data in the early part of the record and before installation of the continuous water-quality sensors, the calibration dataset included 36 samples collected from 2016–17 with associated surrogate parameters (streamflow, SC, pH, water temperature, and turbidity). Concentrations were not substantially censored and only 6 percent of dissolved arsenic and 3 percent of dissolved copper concentrations were censored. Summary statistics for the final calibration dataset and resulting regression equations and statistics are presented in table 3 and graphs of predicted versus measured concentrations for dissolved and total metals are shown in figure 3.

Suitable models (R^2 greater than 0.50) were found for most dissolved and total metals except for dissolved arsenic (R^2=0.45), dissolved copper (R^2=0.42), and total arsenic (R^2=0.44). The regression models were strongest for metals with relatively high concentrations, such as dissolved and total aluminum, iron, manganese, and zinc. Metal concentrations are elevated at this site because of drainage from abandoned mine lands and highly mineralized bedrock in the watershed (Church and others, 2007). Most regression models for dissolved metals at this site included SC (aluminum, manganese, and zinc) or streamflow (iron and lead) as the main explanatory variable and seasonal terms, pH, and water temperature as secondary variables. In the models for total metals, SC was the main explanatory variable for many metals (aluminum, cadmium, copper, manganese, and zinc) and turbidity was the main explanatory variable only for iron and lead. The lack of a strong correlation between total metals and turbidity at this site reflects the low pH of stream water in Cement Creek (median pH=4.3) (Wright and others, 2007) associated with high metals solubility, which resulted in a high proportion of total metals occurring in the dissolved phase.

To further evaluate model performance, daily mean concentrations were estimated from the regression equations in table 3 using the daily streamflow and water-quality data from the USGS station Cement Creek at Silverton, Colo., for the period April 2016 through December 2017 (fig. 4). Concentrations were not estimated for the winter months when the water-quality sensor was not operated because of ice formation at the gage. The prediction interval for estimated concentrations increased at higher concentrations for some dissolved metals, especially cadmium, iron, and lead. This was likely because there was greater variability in concentrations and fewer samples collected during low-flow conditions when concentrations are highest. The models appeared to perform best for constituents that showed substantial dilution during snowmelt (May and June) such as aluminum, manganese, and zinc. Model performance was weaker for metals that had a narrower range of concentrations (dissolved arsenic and copper) although metals with little variability may not necessitate a predictive model. Models also were weaker for metals that showed a flushing or source-limited behavior during early snowmelt such as total copper and lead, perhaps indicating the seasonal terms were not sufficient to capture this seasonal pattern.

In summary, the regression models are deemed adequate for predicting concentrations of most dissolved and total metals over a wide range of concentrations and flow conditions at this site. Although the dataset contained only 2 years of data, the residuals plotted over time (see Mast, 2018) did suggest an apparent downward trend for some metals, most notably dissolved cadmium and total copper.

Table 3. Summary statistics, regression model equations, and evaluation statistics for dissolved and total metal concentrations for Cement Creek at Silverton, Colorado (U.S. Geological Survey station 09358550), 2016–17.

[Models with R^2 of less than 0.5 are shown in gray. Model terms listed in order of increasing *p*-value. Min., minimum concentration; µg/L, microgram per liter; Med., median concentration; Max., maximum concentration; N, number of samples in calibration dataset; R^2, coefficient of determination adjusted for serial correlation; Std. error, standard error of estimate; Bias corr., bias correction factor (Duan, 1983); Pred. int., median prediction interval as a percent of the median estimated concentration; *C*, metal concentration; *SC*, specific conductance in microsiemens per centimeter at 25 degrees Celsius; *sin*_2, sin(4π*T*); *cos*_2, cos(4π*T*), where *T* is the decimal portion of the year starting on January 1; —, not computed; *Temp*, water temperature in degrees Celsius; *log*, log base 10; *Q*, streamflow in cubic feet per second; *pH*, pH in standard units; *sin*, sin(2π*T*); *cos*, cos(2π*T*); *Turb*, turbidity in nephelometric turbidity units]

Metal, fraction	Percent censored	Min. (µg/L)	Med. (µg/L)	Max. (µg/L)	N	Model equation and coefficients	R^2	Std. error	Bias corr.	Pred. int.
Aluminum, dissolved	0	120	2,600	7,200	36	$C = 6.895SC - 364.0sin\_2 - 512.2cos\_2 - 980.3$	0.96	424	—	20
Arsenic, dissolved	6	0.18	0.34	1.0	34	$C = -0.03585Temp - 0.007792sin\_2 + 0.1858cos\_2 - 0.4112logQ + 1.365$	0.45	—	—	—
Cadmium, dissolved	0	1.4	2.8	4.7	36	$logC = -0.08414pH + 0.2514logSC + 0.1140$	0.77	0.069	1.01	30
Copper, dissolved	3	19	44	83	34	$C = -4.647sin\_2 - 8.350cos\_2 - 34.85logQ - 0.04422SC + 133.0$	0.42	—	—	—
Iron, dissolved	0	1,200	4,450	13,000	36	$logC = -0.002784Q - 0.03469Temp - 0.06560sin\_2 - 0.04077cos\_2 + 4.107$	0.85	0.10	1.02	50
Lead, dissolved	0	0.35	7.6	18	36	$C = -6.118logQ - 2.119pH - 1.492sin - 0.4141cos + 28.08$	0.75	2.2	—	30
Manganese, dissolved	0	640	1,900	4,700	36	$logC = 0.6101logSC - 0.2860logQ + 0.008238sin\_2 - 0.02704cos\_2 + 2.130$	0.98	0.035	1.00	20
Zinc, dissolved	0	380	1,050	2,100	36	$C = 0.9963SC - 198.6pH + 1412$	0.91	144	—	20
Aluminum, total	0	1,200	3,300	7,400	32	$C = 4.868SC - 99.46Temp - 237.4sin\_2 - 435.7cos\_2 + 1837$	0.91	516	—	20
Arsenic, total	0	0.70	3.4	6.2	33	$C = 0.02046Q + 0.005990SC + 1.241sin + 0.3178cos - 1.807$	0.44	—	—	—
Cadmium, total	0	1.6	2.8	4.6	33	$C = 2.303logSC + 0.01072Turb - 0.1058sin\_2 - 0.2779cos\_2 - 3.596$	0.79	0.32	—	20
Copper, total	0	32	58	120	32	$logC = -0.0002826SC + 0.001789Turb - 0.02920sin\_2 - 0.05793cos\_2 + 1.887$	0.57	0.092	1.02	50
Iron, total	0	3,000	9,800	21,000	33	$C = 68.31Turb - 613.5Temp + 1332sin + 3120cos + 14080$	0.68	2,460	—	50
Lead, total	0	9.0	16.5	59	32	$logC = 0.3798logTurb + 0.0004769Q + 0.7820$	0.56	0.14	1.05	80
Manganese, total	0	780	1,800	5,100	33	$C = 4.090SC + 5.315Turb + 312.5sin + 84.59cos - 191.4$	0.98	164	—	10
Zinc, total	0	430	1,000	2,100	33	$C = 1927logSC + 5.307Turb - 113.7sin\_2 - 103.4cos\_2 - 4180$	0.94	107	—	10

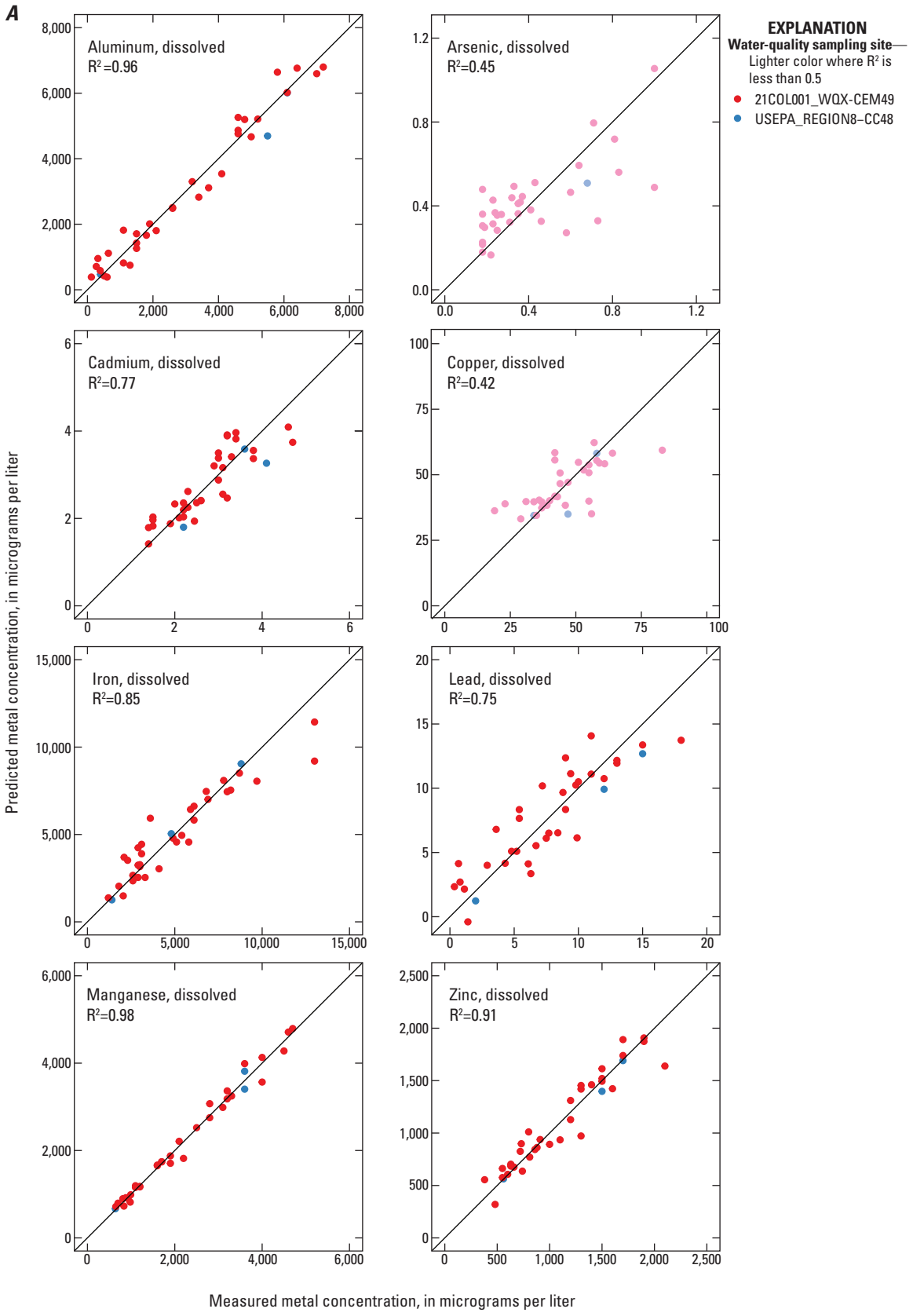

Figure 3. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for samples collected at or near Cement Creek at Silverton, Colorado (U.S. Geological Survey station 09358550), 2016–17.

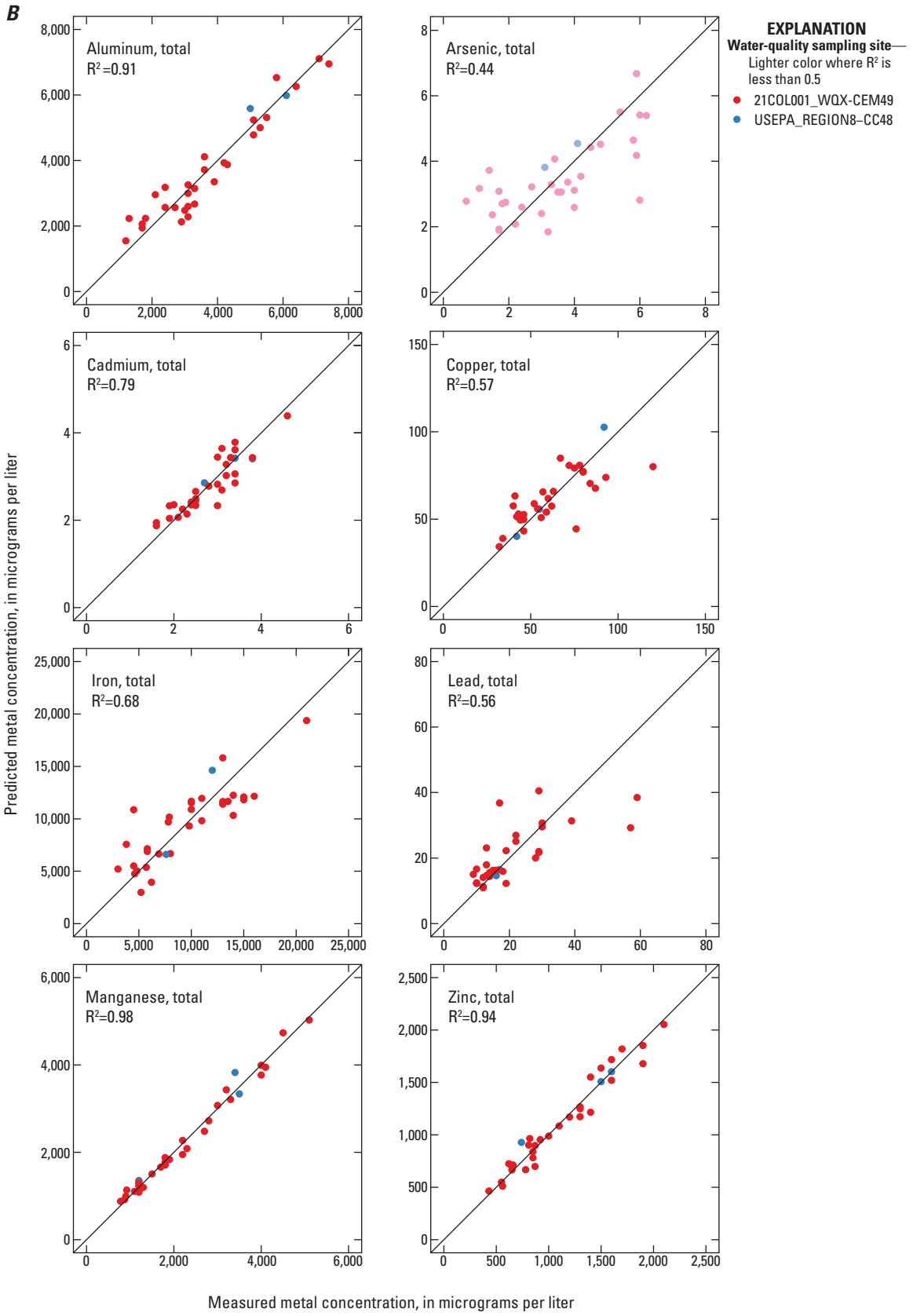

Figure 3. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for samples collected at or near Cement Creek at Silverton, Colorado (U.S. Geological Survey station 09358550), 2016–17.—Continued

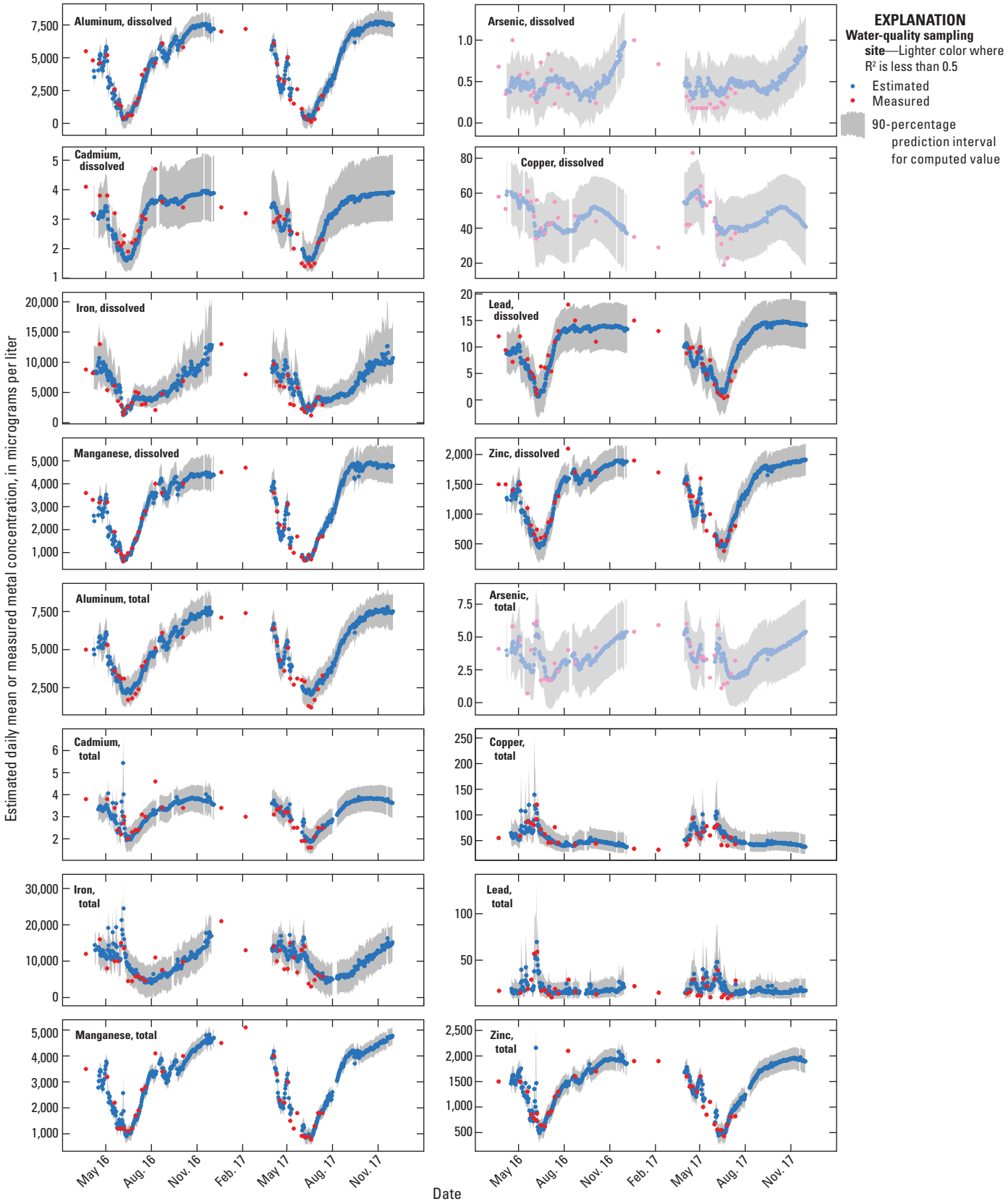

Figure 4. Time series plots of estimated daily mean and measured metal concentrations for Cement Creek at Silverton, Colorado (U.S. Geological Survey station 09358550), for 2016–17. Gaps in estimated concentrations indicate periods of missing surrogate data.

Animas River below Silverton, Colorado (Station 09359020)

The dataset compiled for this site included 958 water-quality samples collected in the vicinity of the USGS station on the Animas River below Silverton (site 2 in fig. 1) with dissolved or total metal analyses for 1969–2017. Because of the lack of field data in the early part of the record and before installation of the continuous water-quality sensors, the calibration dataset included 68 samples collected from 2012–17 with associated surrogate parameters (streamflow, SC, pH, water temperature, and turbidity). A higher proportion of concentrations were censored compared to the Cement Creek site, especially for dissolved and total arsenic (30 and 11 percent, respectively) and dissolved lead (13 percent) (table 4). Summary statistics for the final dataset and resulting regression equations and statistics are presented in table 4 and graphs of predicted versus measured concentrations for all eight metals are shown in figure 5.

Adequate models (R^2 greater than 0.50) were found for most dissolved and total metals except for dissolved arsenic (R^2=0.35), dissolved copper (R^2=0.39), dissolved lead (R^2=0.31), and total copper (R^2=0.20). The regression models for dissolved metals at this site had SC as the main explanatory variable and pH, water temperature, and seasonal terms as secondary variables. Three of the models for total metals (arsenic, iron, and lead) had turbidity as the main explanatory variable and four models (aluminum, cadmium, manganese, and zinc) had SC as the main explanatory variable. The strong relation with SC rather than turbidity for some total metals (cadmium, manganese, and zinc) is because of a relatively high concentration of these metals from upstream mining sources and mineralized bedrock (Church and others, 2007) and their occurrence primarily in the dissolved phase even at circumneutral pH (median stream pH=6.8) (Wright and others, 2007). Total metals that had models with a turbidity term (aluminum, arsenic, iron, and lead) had a larger fraction present in the particulate phase caused by decreasing solubility of these metals with increasing pH (Schemel and others, 2000). For example, total lead was correlated with turbidity and had a median concentration of 7.2 µg/L, which is 20 times higher than the median dissolved concentration of 0.34 µg/L, indicating that most lead was in particulate forms. In contrast, zinc was correlated with SC and had nearly identical total (352 µg/L) and dissolved (342 µg/L) concentrations, which indicates that most zinc was in dissolved forms (table 4).

To further evaluate model performance, daily mean concentrations were estimated from the corresponding regression equations in table 4 using the daily streamflow and water-quality data from the Animas River below Silverton site for the period April 2016 through December 2017 (fig. 6). Concentrations were not estimated for the winter months when the water-quality sensor was not operated because of ice formation at the gage.

Similar to the models for the Cement Creek site, the models appeared to perform best for constituents that showed substantial dilution during snowmelt (May and June) such as cadmium, iron, manganese, and zinc. Model performance was weaker for metals that showed a narrower range of concentrations (dissolved arsenic and copper) or metals that tended to increase in concentration during early snowmelt (total arsenic and lead).

The model for dissolved aluminum, which has an R^2 of 0.55, estimated a relatively sudden increase in concentrations during late fall and early winter. However, the model is poorly constrained during these months because of a lack of samples representing low-flow conditions. In addition, the median prediction interval for dissolved aluminum exceeded the median concentration by 240 percent (table 4) indicating a high degree of uncertainty in the predicted concentrations. In summary, the regression models are deemed adequate for predicting concentrations of most dissolved and total metals over a wide range of concentrations and flow conditions.

Table 4. Summary statistics, regression model equations, and evaluation statistics for dissolved and total metal concentrations for Animas River below Silverton, Colorado (U.S. Geological Survey station 09359020), 2012–17.

[Models with R^2 of less than 0.5 are shown in gray. Model terms listed in order of increasing *p*-value. Min., minimum concentration; µg/L, microgram per liter; Med., median concentration; Max., maximum concentration; N, number of samples in calibration dataset; R^2, coefficient of determination adjusted for serial correlation; Std. error, standard error of estimate; Bias corr., bias correction factor (Duan, 1983); Pred. int., median prediction interval as a percent of the median estimated concentration; *C*, metal concentration; *SC*, specific conductance in microsiemens per centimeter at 25 degrees Celsius; *sin*_2, sin(4π*T*); *cos*_2, cos(4π*T*), where *T* is the decimal portion of the year starting on January 1; —, not computed; *Temp*, water temperature in degrees Celsius; *log*, log base 10; *Q*, streamflow in cubic feet per second; *pH*, pH in standard units; *sin*, sin(2π*T*); *cos*, cos(2π*T*); *Turb*, turbidity in nephelometric turbidity units]

Metal, fraction	Percent censored	Min. (µg/L)	Med. (µg/L)	Max. (µg/L)	N	Model equation and coefficients	R^2	Std. error	Bias corr.	Pred. int.
Aluminum, dissolved	1	17	55	2,050	67	$C = 1.520SC - 115.6sin\_2 + 201.0cos\_2 - 159.4pH + 748.4$	0.55	209	—	240
Arsenic, dissolved	30	0.084	0.18	0.42	32	$logC = 0.1404sin + 0.02580cos - 0.1283pH + 0.1201$	0.35	—	—	—
Cadmium, dissolved	0	0.46	0.99	2.5	68	$logC = 0.7240logSC + 0.09115sin + 0.06085cos - 1.730$	0.84	0.074	1.01	30
Copper, dissolved	3	2	5.3	17	66	$logC = 0.1489sin + 0.02723cos + 0.001849SC + 0.4824logQ - 1.115$	0.39	—	—	—
Iron, dissolved	0	80.3	660	3,650	68	$C = 4.834SC - 268.6sin\_2 + 57.46cos\_2 - 51.60Temp - 249.1$	0.84	314	—	50
Lead, dissolved	13	0.02	0.34	6.6	54	$logC = 1.081logQ - 0.6241pH + 0.9385$	0.31	—	—	—
Manganese, dissolved	0	230	730	2,200	68	$logC = 1.314logSC + 0.05191sin + 0.02847cos - 0.3554$	0.97	0.049	1.01	20
Zinc, dissolved	0	100	342	926	68	$C = 1.081SC + 85.28sin + 75.60cos + 62.17$	0.92	54	—	20
Aluminum, total	0	220	1,100	3,600	47	$C = 4.229SC + 52.73Turb - 470.8$	0.80	307	—	40
Arsenic, total	11	0.12	1.0	3.8	39	$C = 0.04852Turb + 0.0005967Q + 0.2847sin + 0.3372cos + 0.4045$	0.60	0.45	—	130
Cadmium, total	0	0.73	1.1	2.2	41	$C = 0.00358SC + 0.339sin - 0.238cos + 0.227pH - 1.65$	0.83	0.16	—	20
Copper, total	0	8.7	21	80	47	$logC = 0.1202sin + 0.05913cos - 0.4390logSC + 2.385$	0.20	—	—	—
Iron, total	0	710	2,380	5,550	50	$C = 104.5Turb + 416.6sin + 1107cos + 1637logSC - 2045$	0.72	647	—	50
Lead, total	0	0.078	7.2	60	41	$C = 0.9687Turb + 0.007145Q - 4.329$	0.76	5.0	—	220
Manganese, total	0	360	670	2,200	47	$C = 2.950SC + 163.8sin + 147.2cos + 58.99$	0.89	156	—	30
Zinc, total	0	190	352	780	41	$C = 560.7logSC + 90.14sin + 92.51cos - 915.9$	0.87	59	—	20

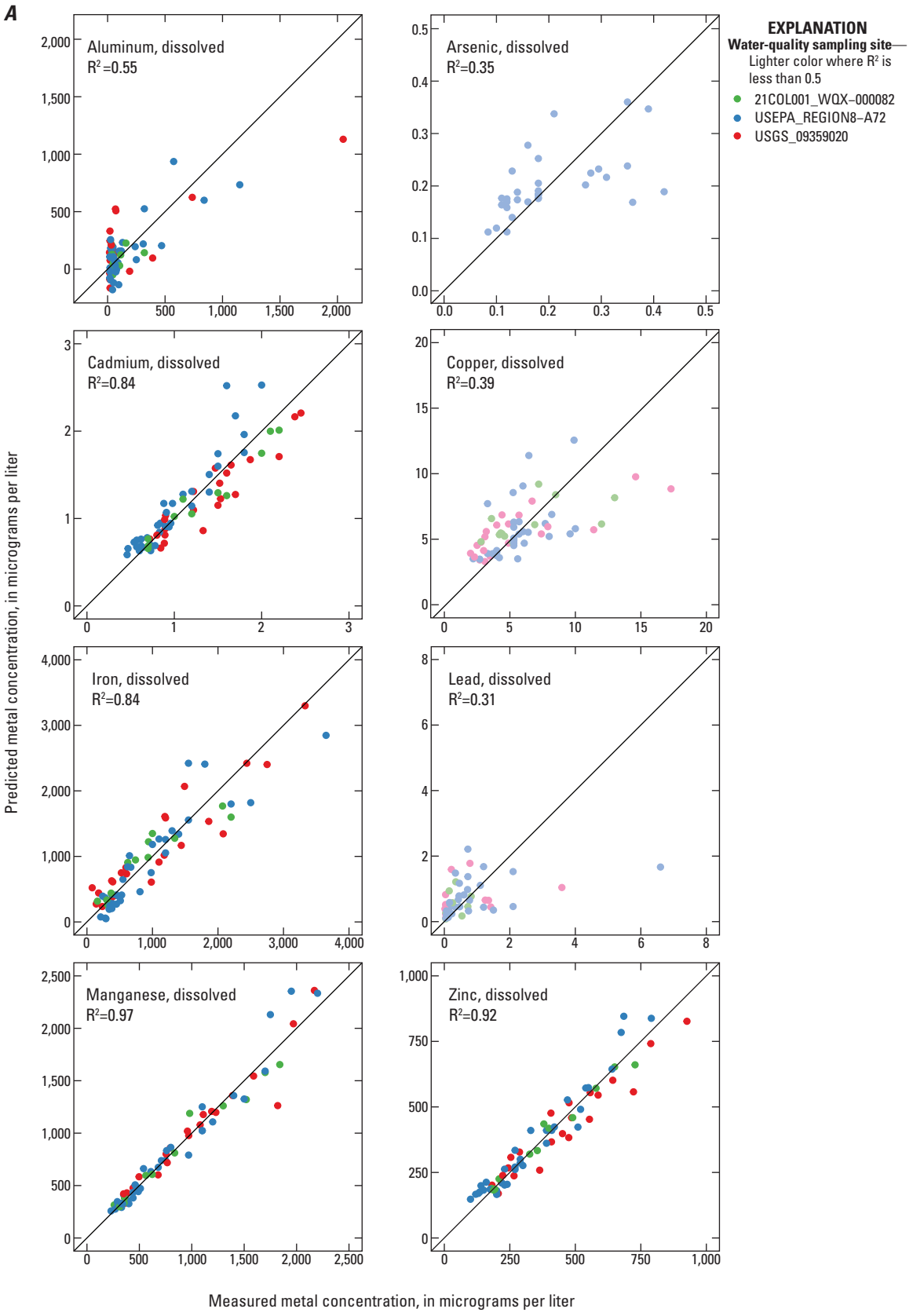

Figure 5. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for samples collected at or near Animas River below Silverton, Colorado (U.S. Geological Survey station 09359020), 2012–17.

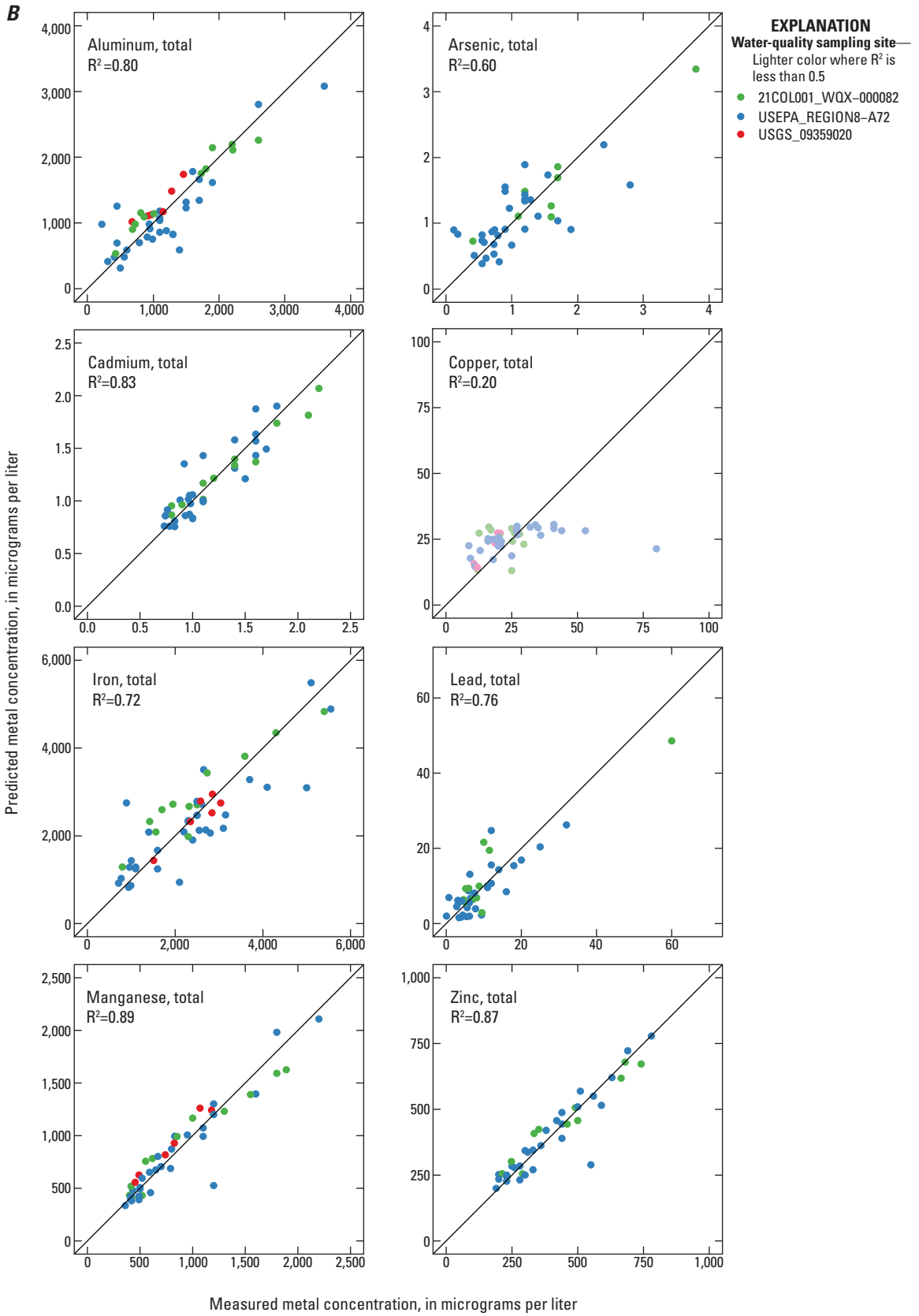

Figure 5. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for samples collected at or near Animas River below Silverton, Colorado (U.S. Geological Survey station 09359020), 2012–17.—Continued

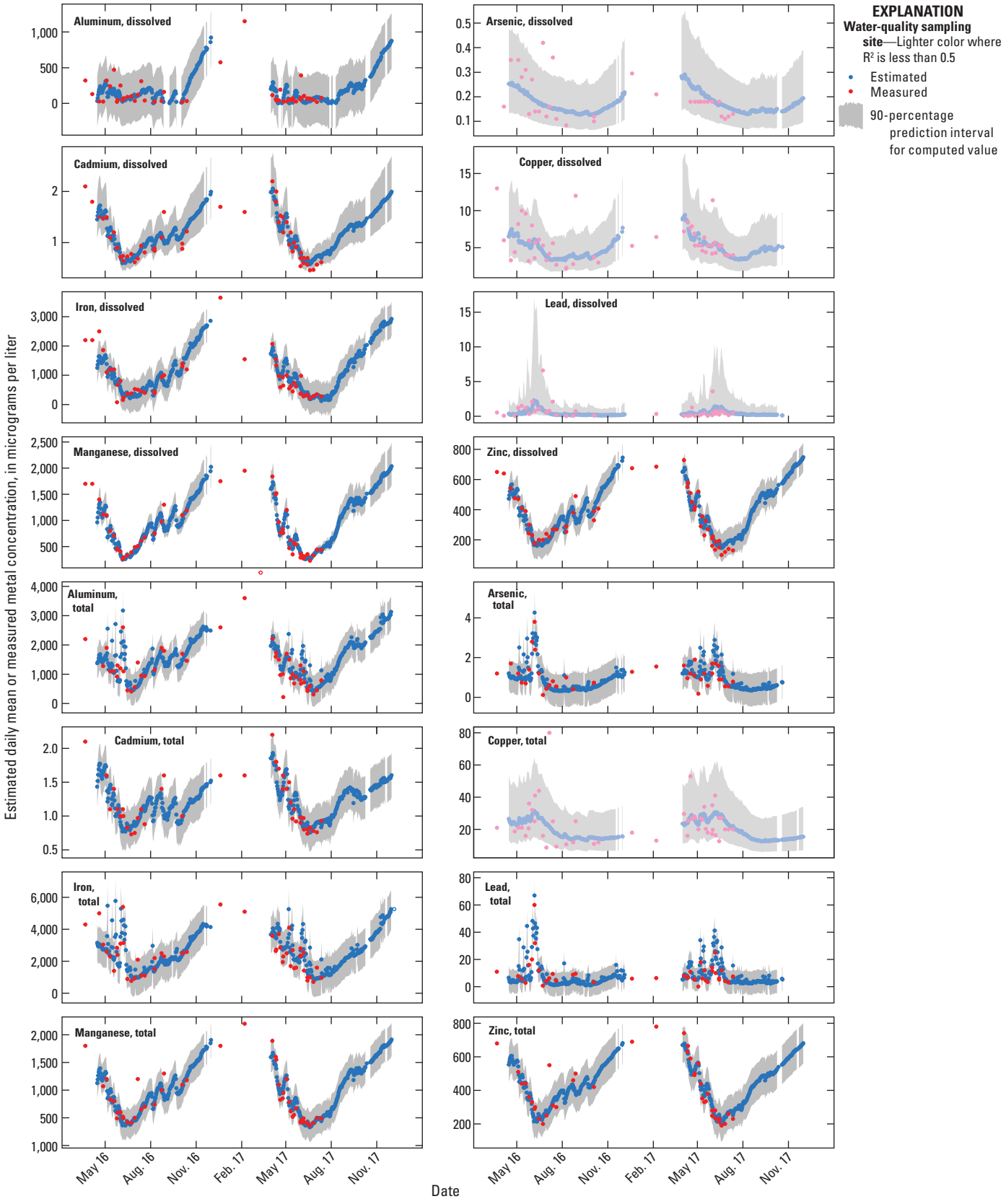

Figure 6. Time series plots of estimated daily mean and measured metal concentrations for Animas River below Silverton, Colorado (U.S. Geological Survey station 09359020), for 2016–17. Gaps in estimated concentrations indicate periods of missing surrogate data.

Animas River at Durango, Colorado (Station 09361500)

The dataset compiled for this site included 739 water-quality samples collected in the vicinity of the USGS station on the Animas River at Durango (site 3 in fig. 1) with dissolved or total metal analyses for 1958–2017. Because of the lack of field data in the early part of the record and before installation of the continuous water-quality sensors, the calibration dataset included 107 samples collected from 2015–17 with associated surrogate parameters (streamflow, SC, pH, water temperature, and turbidity). The proportion of censored concentrations was higher than at the Animas River below Silverton site, with more than one-half of dissolved aluminum, arsenic, and iron concentrations being censored. Of the total metal concentrations, only total arsenic (27 percent) was censored at greater than 3 percent (table 5). Summary statistics for the final dataset and resulting regression equations and statistics are presented in table 5 and graphs of predicted versus measured concentrations for the eight metals are shown in figure 7.

Adequate models (R^2 greater than 0.50) were found for only three dissolved metals (iron, lead, and manganese) and all of the total metals, although the R^2 for total arsenic was only 0.52. All of the regression models for dissolved metals at this site had different main explanatory variables, with SC for iron, seasonal terms for lead, and streamflow for manganese. All the models for total metals except arsenic had turbidity as the main explanatory variable with SC, streamflow, and seasonal terms as secondary variables. Although the R^2 values were more than 0.5 for all the total metals, there was greater scatter at the high end of the concentration range especially for aluminum, iron, and lead, which likely makes the model predictions less precise for these metals.

To further evaluate model performance, daily mean concentrations were estimated from the corresponding regression equations in table 5 using the daily streamflow and water-quality data from the USGS station at Animas River at Durango for the period April 2016 through December 2017 (fig. 8). Model performance for dissolved metals was much weaker than at the upstream Animas River below Silverton site. This partially reflects the range of concentrations, which are nearly an order of magnitude lower than at the Silverton site. Lower concentrations result from inputs of higher pH water downstream from Silverton that more than double streamflow and dilute metals and lower their solubility. Although the iron and lead models had an R^2 of more than 0.5, the median prediction interval for both metals exceeded the median concentration by at least 120 percent (table 5), indicating a high degree of uncertainty in the predictions.

Model performance for total metals was better than for dissolved metals. Although total metal concentrations were lower than at Silverton, they were not reduced by the same degree as dissolved metals. This was likely caused by some of the dissolved load being converted and transported in particulate phase. All the total metals had turbidity as the main explanatory variable, which indicates metal transport at this site is controlled by movement of particulate matter. In contrast to the dilution pattern observed for some total metals at Silverton during snowmelt, especially aluminum, manganese, and zinc, total metals at Durango increased during snowmelt as higher flows mobilize particulate matter that was deposited on the streambed during lower flow conditions. Based on the prediction interval, some total metals, especially aluminum and iron, had a higher level of uncertainty especially during snowmelt when concentrations were highest.

In summary, the regression models were deemed adequate for predicting total metal concentrations at this site over a wide range of concentrations and flow conditions; however, some metals may have higher degree of uncertainty. Use of regression models for predicting dissolved concentrations is much less reliable, possibly because of lower concentrations, and generally are not recommended for most dissolved metals at this site.

Table 5. Summary statistics, regression model equations, and evaluation statistics for dissolved and total metal concentrations for Animas River at Durango, Colorado (U.S. Geological Survey station 09361500), 2015–17.

[Models with R^2 of less than 0.5 are shown in gray. Model terms listed in order of increasing *p*-value. Min., minimum concentration; µg/L, microgram per liter; Med., median concentration; Max., maximum concentration; N, number of samples in calibration dataset; R^2, coefficient of determination adjusted for serial correlation; Std. error, standard error of estimate; Bias corr., bias correction factor (Duan, 1983); Pred. int., median prediction interval as a percent of the median estimated concentration; *log*, log base 10; *C*, metal concentration; *SC*, specific conductance in microsiemens per centimeter at 25 degrees Celsius; *sin*_2, sin(4π*T*), where *T* is the decimal portion of the year starting on January 1; *cos*_2, cos(4π*T*); *pH*, pH in standard units; —, not computed; *Temp*, water temperature in degrees Celsius; *Q*, streamflow in cubic feet per second; *sin*, sin(2π*T*); *cos*, cos(2π*T*); *Turb*, turbidity in nephelometric turbidity units]

Metal, fraction	Percent censored	Min. (µg/L)	Med. (µg/L)	Max. (µg/L)	N	Model equation and coefficients	R^2	Std. error	Bias corr.	Pred. int.
Aluminum, dissolved	53	23	58	170	48	$logC = -0.0008447SC - 0.1246sin\_2 - 0.08146cos\_2 + 0.1258pH + 1.027$	0.47	—	—	—
Arsenic, dissolved	54	0.18	0.24	0.68	38	$logC = 0.0005080SC + 0.01250Temp - 0.8772$	0.46	—	—	—
Cadmium, dissolved	8	0.076	0.12	0.21	47	$logC = -0.3541logQ + 0.04290sin\_2 + 0.07453cos\_2 - 0.0006371SC + 0.4136$	0.41	—	—	—
Copper, dissolved	7	0.70	1.3	5.6	65	$logC = 0.3023logQ - 0.01427Temp - 0.6054$	0.49	—	—	—
Iron, dissolved	52	5.8	66.5	230	50	$logC = -0.002279SC - 0.3385sin\_2 + 0.08451cos\_2 - 0.0001473Q + 2.505$	0.62	0.22	1.12	160
Lead, dissolved	10	0.062	0.54	2.5	38	$logC = 0.1475sin - 0.3852cos + 0.00008027Q - 0.7745$	0.69	0.19	1.15	120
Manganese, dissolved	0	13	64	240	87	$logC = -1.084logQ - 1.654logSC + 0.1178sin + 0.04703cos + 9.175$	0.63	0.10	1.03	50
Zinc, dissolved	0	2.8	39	64	107	$C = -0.009641Q - 3.772sin\_2 + 13.64cos\_2 - 1.563Temp + 68.00$	0.36	—	—	—
Aluminum, total	1	62	455	5,340	98	$logC = 0.9568logTurb + 0.0008199SC + 1.608$	0.73	0.248	1.44	160
Arsenic, total	27	0.34	1	6.5	59	$C = 0.001126Q + 3.652logSC - 9.374$	0.52	0.861	—	150
Cadmium, total	3	0.12	0.36	1	39	$C = 0.009996Turb - 0.0007006SC - 0.1395sin + 0.1685cos + 0.6228$	0.72	0.126	—	60
Copper, total	0	1.2	5.6	83.5	69	$logC = 0.01873Turb + 0.2393sin - 0.08733cos + 0.4243$	0.75	0.213	1.23	140
Iron, total	0	109	710	8,800	101	$logC = 0.01771Turb + 0.2205sin + 0.02787cos + .00007809Q + 2.536$	0.73	0.251	1.35	170
Lead, total	0	0.6	6.1	140	99	$logC = 0.01823Turb + 0.1840sin - 0.1108cos + 0.0001377Q + 0.2500$	0.83	0.221	1.14	140
Manganese, total	0	64.5	157	1,100	79	$logC = 0.01066Turb + 0.00007038Q + 0.06592sin + 0.1221cos + 2.048$	0.80	0.124	1.04	70
Zinc, total	0	19	86.1	390	99	$logC = +0.008940Turb + 0.06373sin + 0.1418cos - 0.0005755SC + 2.081$	0.72	0.125	1.04	70

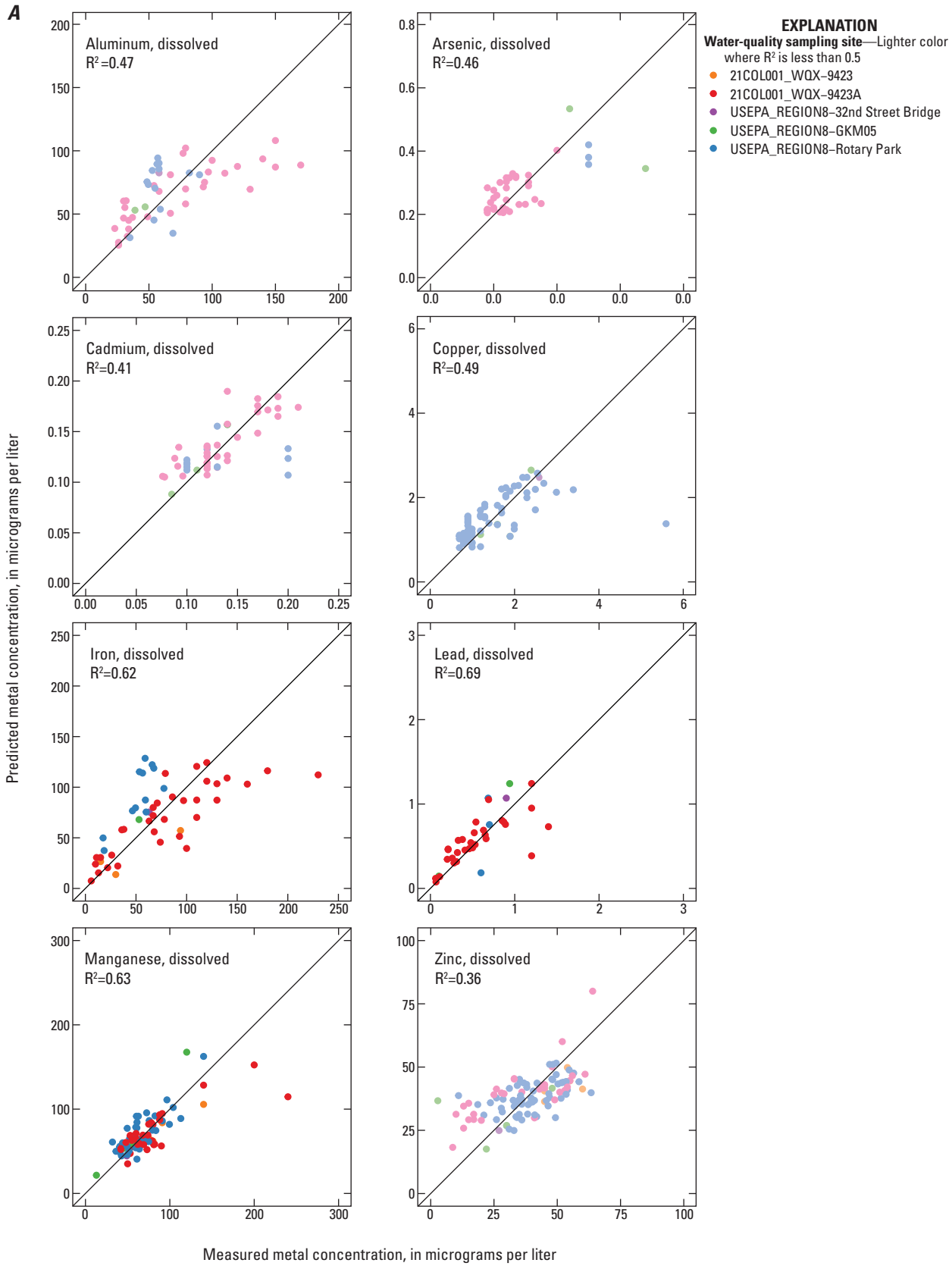

Figure 7. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for samples collected at or near Animas River at Durango, Colorado (U.S. Geological Survey station 09361500), 2015–17.

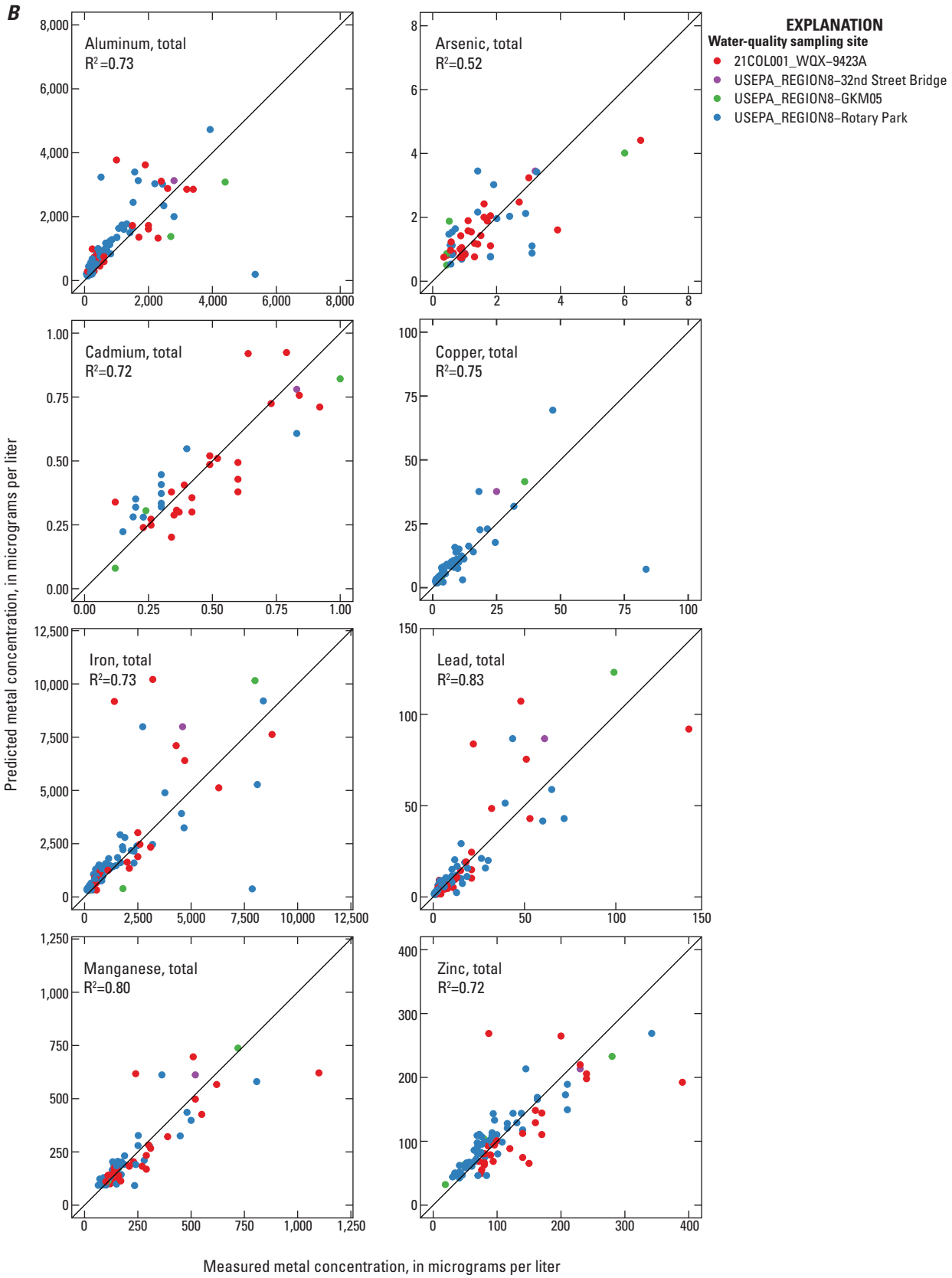

Figure 7. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for samples collected at or near Animas River at Durango, Colorado (U.S. Geological Survey station 09361500), 2015–17.—Continued

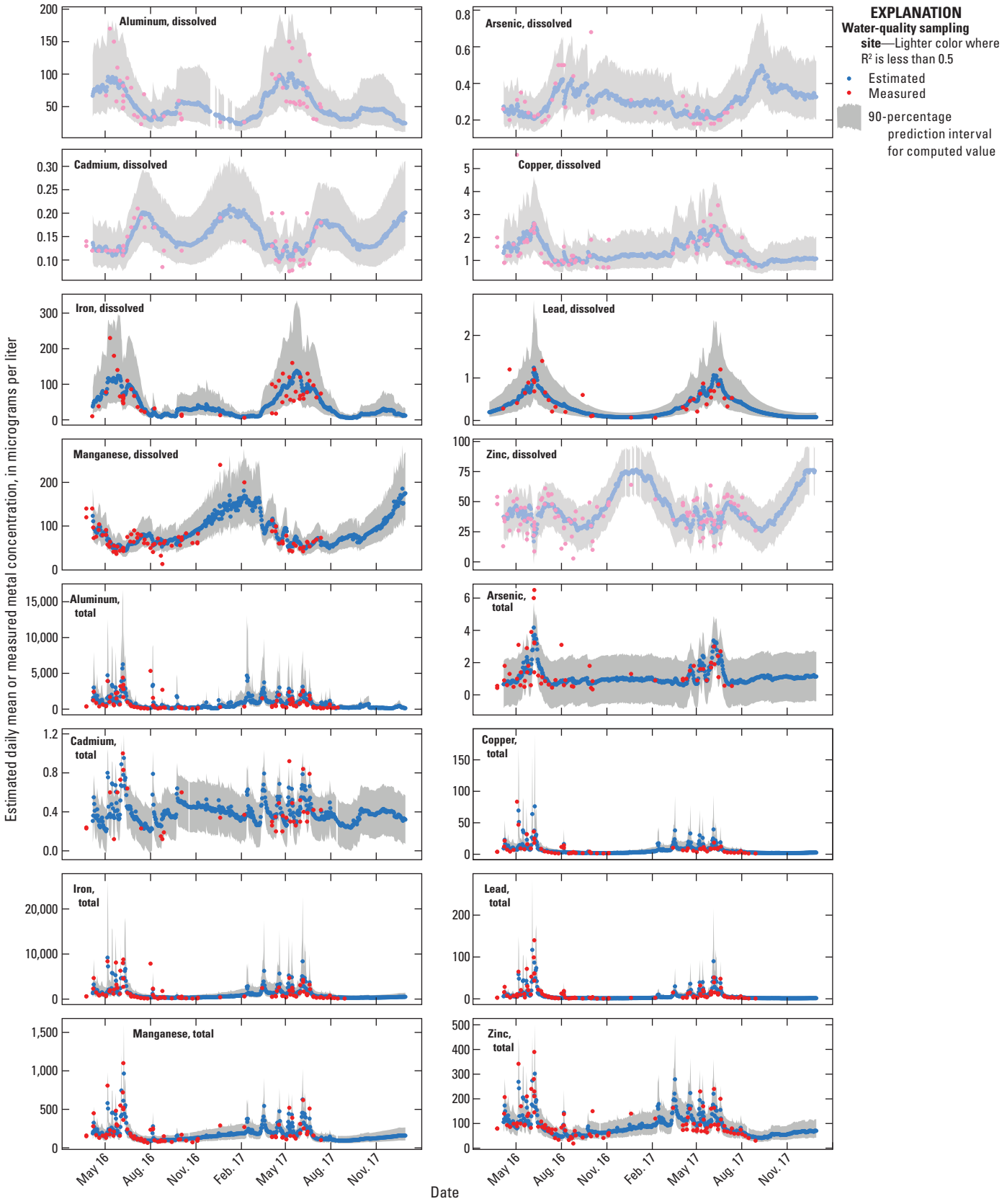

Figure 8. Time series plots of estimated daily mean and measured metal concentrations for Animas River near Durango, Colorado (U.S. Geological Survey station 09361500), for 2016–17. Gaps in estimated concentrations indicate periods of missing surrogate data.

Animas River near Cedar Hill, New Mexico (Station 09363500)

The dataset compiled for this site included 361 water-quality samples collected in the vicinity of the USGS station at Animas River near Cedar Hill (site 4 in fig. 1) with dissolved or total metal analyses for 1942–2017. Because of the lack of field data in the early part of the record and before installation of the continuous water-quality sensors, the calibration dataset included 36 samples collected from 2016–17 with associated surrogate parameters (streamflow, SC, pH, water temperature, and turbidity). All the dissolved metals except manganese and zinc had a large percentage (14 to 56 percent) of concentrations reported as censored. Total metals had censored concentrations in 20 percent or less of samples with the exception of arsenic, which had 29 percent censored (table 6). Summary statistics for the final dataset and resulting regression equations and statistics are presented in table 6 and graphs of predicted versus measured concentrations for all eight metals are shown in figure 9.

Adequate models (R^2 greater than 0.50) were found for four dissolved metals (copper, lead, manganese, and zinc) and all eight of the total metal fractions. Dissolved arsenic and cadmium had insufficient data to compute a regression model. The regression models for dissolved metals at this site included SC, temperature, or seasonal terms as the main explanatory variable and there was no consistent pattern in model form among the metals. All the models for total metals had streamflow as the main explanatory variable. Six of the total models included temperature and seasonal terms as secondary variables and only cadmium and copper included turbidity.

To further evaluate model performance, daily mean concentrations were estimated from the corresponding regression equations in table 6 using the daily streamflow and water-quality data from the USGS station at Animas River near Cedar Hill for the period April 2016 through December 2017 (fig. 10). Although the water-quality sensor was operated through the winter months at this station, there are several periods of missing SC record in early 2017 for which concentrations could not be estimated. Median prediction intervals for dissolved lead and zinc exceeded the median predicted concentration by 130 percent or more, indicating a large degree of uncertainty in the estimated concentrations. The large increase in dissolved manganese during winter months appears to be driven by the temperature term; however, there are no samples during base-flow conditions to confirm this pattern. Overall, the seasonal patterns in estimated total metal concentrations showed more consistency among the metals and the median prediction intervals, except for lead, were 100 percent or less (table 6).

Estimated concentrations for most total metals showed a strong increase during the 2016 snowmelt period with a weaker response during snowmelt in 2017. This can partially be explained by different runoff patterns in the 2 years, with snowmelt in 2016 having a shorter duration and one large peak compared to 2017, which had a more extended melt period with multiple lower peaks. The pattern for total cadmium was slightly different than other metals showing a second increase during the fall and winter months. This pattern may partially be an artifact of two higher concentration samples collected in October 2016, which causes the seasonal terms in the model to predict higher concentrations in fall and winter.

Most total metal models at this site had streamflow as the main explanatory variable in contrast to the Animas River at Durango site, where turbidity was the main explanatory variable. This result seemed unexpected given that turbidity increases substantially downstream from Durango (fig. 2) from inputs of sediment to the river from the surrounding sedimentary bedrock. Closer inspection of the data showed that the distribution of samples in the calibration dataset may have been a factor. The calibration dataset was dominated by samples collected during snowmelt, a period when turbidity tends to be lower because of the dominance of streamflow from higher elevation areas. The dataset contained few samples that were collected during short-term runoff events at other times of the year when turbidity is often elevated.

In summary, the regression models are generally adequate for predicting total metal concentrations during snowmelt but will have greater uncertainty during winter months or short-term runoff events because of a lack of samples collected under these flow conditions. Use of regression models for predicting dissolved concentrations is less reliable for most metals because of the low concentrations of dissolved metals at this site and lack of samples collected during base-flow conditions.

Table 6. Summary statistics, regression model equations, and evaluation statistics for dissolved and total metal concentrations for Animas River near Cedar Hill, New Mexico (U.S. Geological Survey station 09363500), 2016–17.

[Models with R^2 of less than 0.5 are shown in gray. Model terms listed in order of increasing *p*-value. Min., minimum concentration; µg/L, microgram per liter; Med., median concentration; Max., maximum concentration; N, number of samples in calibration dataset; R^2, coefficient of determination adjusted for serial correlation; Std. error, standard error of estimate; Bias corr., bias correction factor (Duan, 1983); Pred. int., median prediction interval as a percent of the median estimated concentration; *log*, log base 10; *C*, metal concentration; *sin*, sin(2π*T*), where *T* is the decimal portion of the year starting on January 1; *cos*, cos(2π*T*); —, not computed; *Q*, streamflow in cubic feet per second; *SC*, specific conductance in microsiemens per centimeter at 25 degrees Celsius; *Temp*, water temperature in degrees Celsius; *sin_2*, sin(4π*T*); *cos_2*, cos(4π*T*); *Turb*, turbidity in nephelometric turbidity units]

Metal, fraction	Percent censored	Min. (µg/L)	Med. (µg/L)	Max. (µg/L)	N	Model equation and coefficients	R^2	Std. error	Bias corr.	Pred. int.
Aluminum, dissolved	19	20	70	140	29	*logC* = 0.2250*sin* - 0.1329*cos* + 1.689	0.49	—	—	—
Arsenic, dissolved	56	0.20	0.30	0.98	16	—	—	—	—	—
Cadmium, dissolved	51	0.030	0.10	2.0	17	—	—	—	—	—
Copper, dissolved	14	0.60	1.7	4.0	31	*logC* = 0.2186*sin* - 0.1682*cos* + 0.00008780*Q* + 0.0003101*SC* - 0.1740	0.81	0.097	1.02	50
Iron, dissolved	33	5.5	71	160	24	*logC* = - 0.001592*SC* + 2.229	0.23	—	—	—
Lead, dissolved	34	0.070	0.50	2.6	23	*logC* = - 2.646*logSC* - 0.4308*sin* + 0.629*cos* + 6.807	0.62	0.30	1.20	270
Manganese, dissolved	0	4.6	19	67	36	*C* = - 4.355*Temp* + 16.29*sin_2* + 14.77*cos_2* + 79.49	0.61	9.2	—	80
Zinc, dissolved	6	2.0	17	61	33	*logC* = 0.3711*sin* - 0.4791*cos* + 0.7657	0.59	0.20	1.10	130
Aluminum, total	0	110	800	4,600	35	*C* = 1.139*Q* + 905.8*sin* + 2656*cos* + 235.3*Temp* - 2055	0.67	622	—	100
Arsenic, total	29	0.40	1.3	5.5	25	*logC* = 0.0002539*Q* + 0.2151*sin* + 0.4329*cos* + 0.03897*Temp* - 0.6701	0.80	0.13	1.03	100
Cadmium, total	20	0.080	0.40	1.3	28	*C* = 0.0002039*Q* + 0.003042*Turb* - 0.3193*sin* + 0.5544*cos* + 0.4475	0.77	0.15	—	50
Copper, total	0	1.2	6.9	42	35	*logC* = 0.0002170*Q* + 0.2510*sin* + 0.1650*cos* + 0.001661*Turb* + 0.3991	0.91	0.12	1.03	80
Iron, total	0	110	1,080	8,500	35	*C* = 2.123*Q* + 1384*sin* + 4785*cos* + 403.7*Temp* - 4035	0.82	791	—	90
Lead, total	0	0.50	6.3	110	35	*C* = 0.02664*Q* + 9.484*sin* + 55.20*cos* + 4.449*Temp* - 52.11	0.83	9.4	—	120
Manganese, total	0	25	150	900	35	*C* = 0.2089*Q* + 118.7*sin* + 447.0*cos* + 35.72*Temp* - 337.4	0.86	68	—	70
Zinc, total	0	8.0	79	350	35	*C* = 0.07910*Q* + 43.47*sin* + 180.1*cos* + 11.95*Temp* - 85.11	0.79	33	—	70

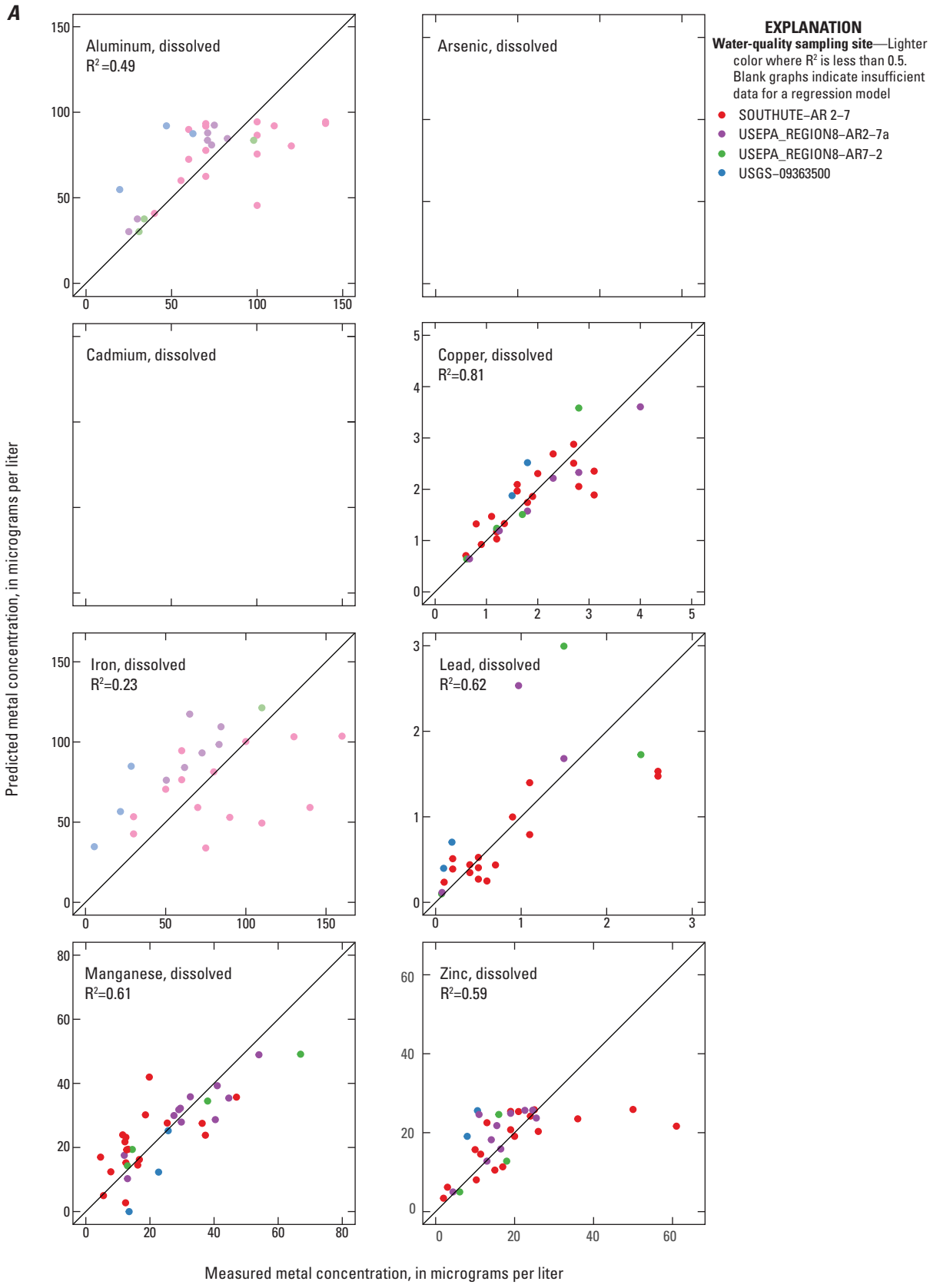

Figure 9. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for Animas River near Cedar Hill, New Mexico (U.S. Geological Survey station 09363500), 2016–17. Blank graphs indicate insufficient data for a regression model.

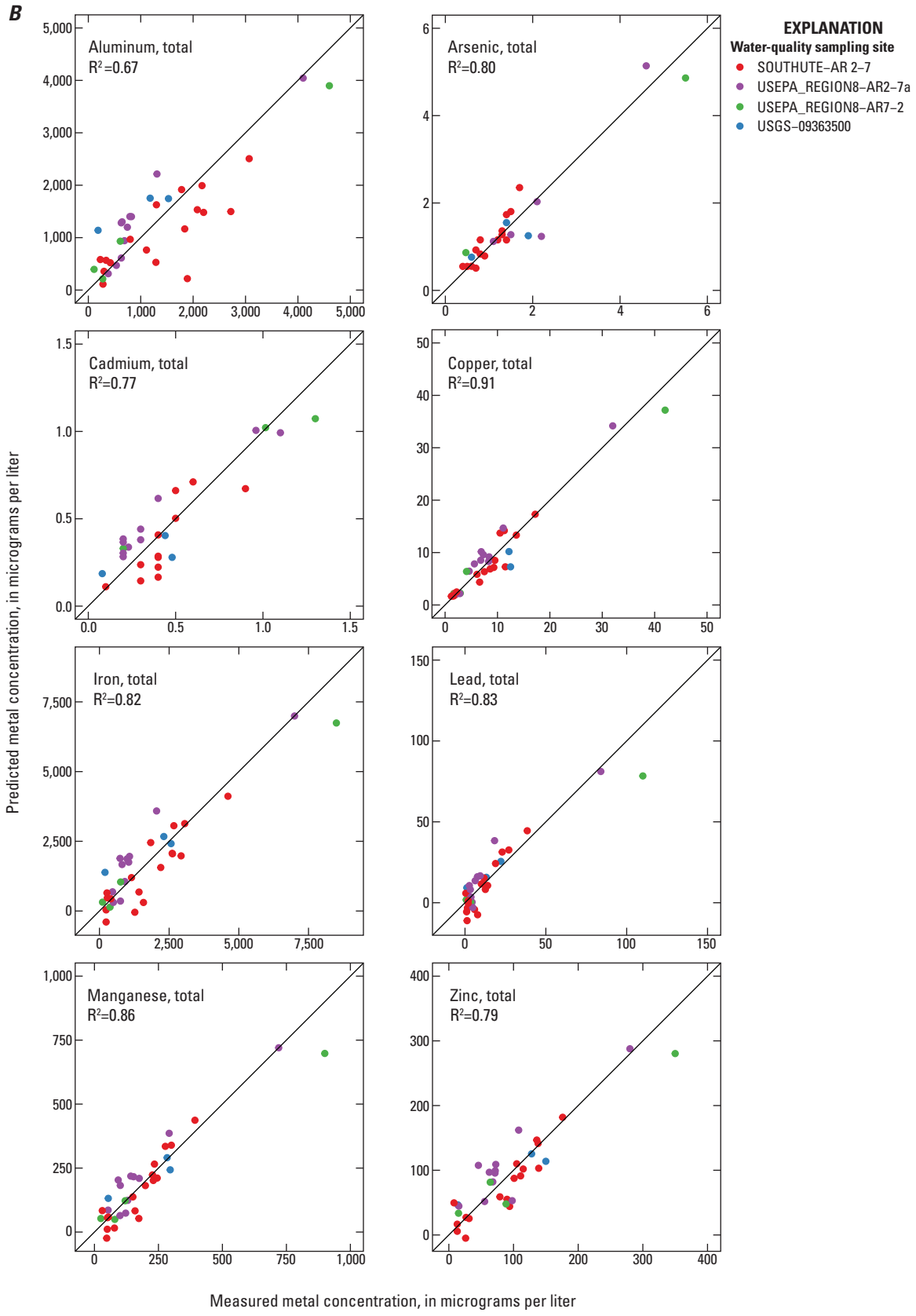

Figure 9. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for Animas River near Cedar Hill, New Mexico (U.S. Geological Survey station 09363500), 2016–17. Blank graphs indicate insufficient data for a regression model.—Continued

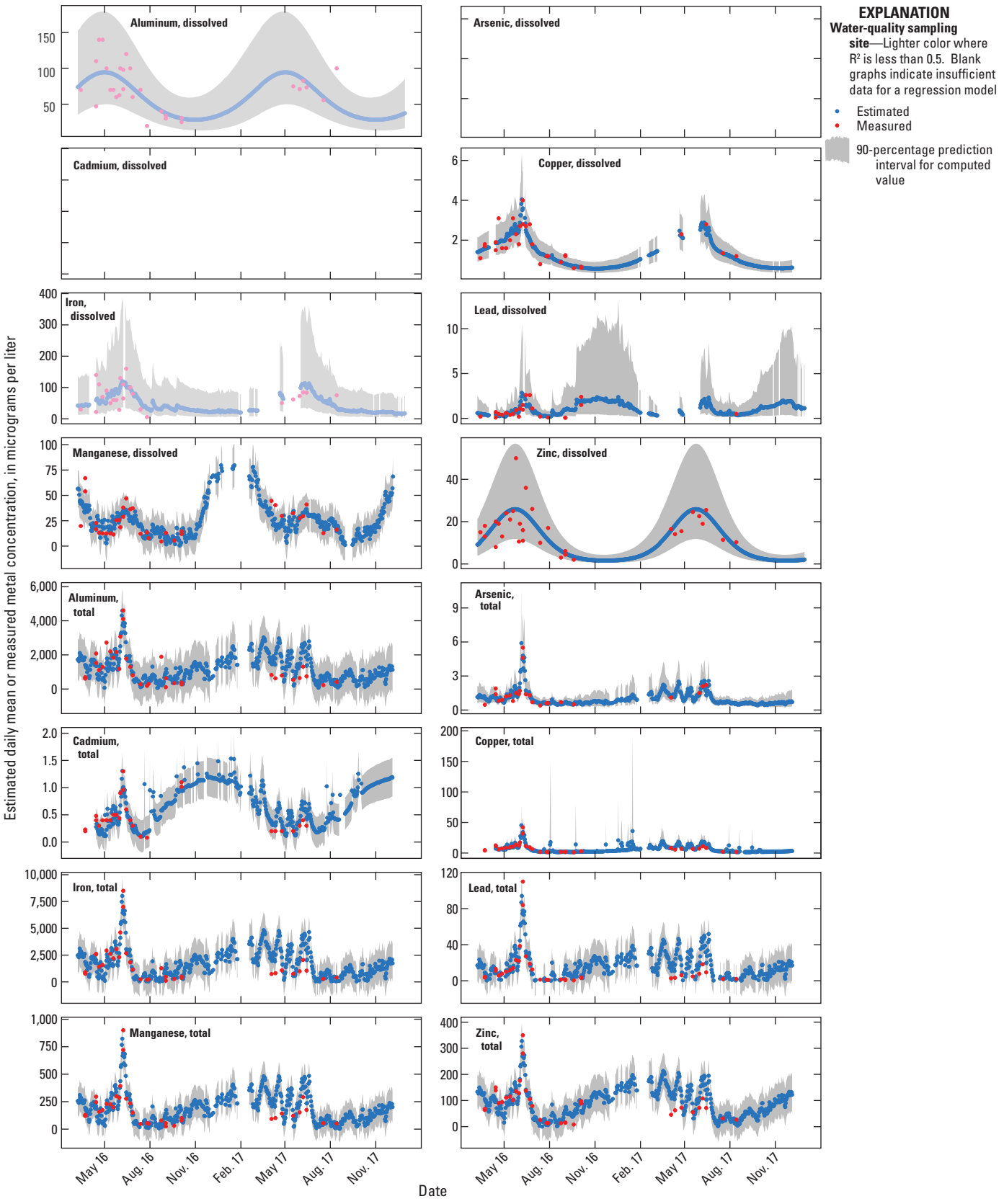

Figure 10. Time series plots of estimated daily mean and measured metal concentrations for Animas River near Cedar Hill, New Mexico (U.S. Geological Survey station 09363500), for 2016–17. Gaps in estimated concentrations indicate periods of missing surrogate data. Blank graphs indicate insufficient data for a regression model.

Animas River below Aztec, New Mexico (Station 09364010)

The dataset compiled for this site included 237 water-quality samples collected in the vicinity of the USGS station at Animas River below Aztec (site 5 in fig. 1) with dissolved or total metal analyses for 2005–17. Because of the lack of field data in the early part of the record and before installation of the continuous water-quality sensors, the calibration dataset included 46 samples collected from 2005–17 with associated surrogate parameters (streamflow, SC, pH, water temperature, and turbidity). All the dissolved metals except manganese and zinc had a substantial percentage (21 to 63 percent) of concentrations reported as censored. Of the total metals, only arsenic, cadmium, and zinc had censored concentrations, but for less than 10 percent of samples. Summary statistics for the final dataset and resulting regression equations and statistics are presented in table 7 and graphs of predicted versus measured concentrations for all eight metals are shown in figure 11.

Adequate models (R^2 greater than 0.50) were found for dissolved aluminum, copper, iron, and zinc and all eight of the total metals. Dissolved arsenic, cadmium, and lead had insufficient data to compute regression models. There was no consistent pattern in model form among dissolved metals. For example, aluminum and zinc included SC as the main explanatory variable but the main explanatory variables for copper and iron were temperature and streamflow, respectively.

All models for total metals included turbidity as the main explanatory variable and models for five of the total metals also contained streamflow and seasonal terms as secondary variables. This is in contrast to the Animas River near Cedar Hill site where streamflow was the main explanatory variable in total metal models. A possible explanation for the greater importance of turbidity at the Aztec site is that the calibration dataset included data for a few storm events in fall of 2016 with elevated metals and turbidity.

To further evaluate the models, daily mean concentrations were estimated from the corresponding regression equations in table 7 using the daily streamflow and water-quality data from the USGS station at Animas River below Aztec for the period April 2016 through December 2017 (fig. 12). Although the water-quality sensor was operated year round at this station, there were several extended periods of missing record for which concentrations could not be estimated. Despite an R^2 higher than the 0.5 level for four dissolved metals, wide prediction intervals for aluminum, iron, and zinc (greater than 100 percent of predicted concentrations) indicate that estimates of these metals may be less reliable (table 7). In addition, there were few samples available to verify concentrations outside of the snowmelt period.

The models for total metals generally performed better than for dissolved metals and the prediction intervals were below 100 percent except for aluminum (110 percent) and lead (120 percent). The predicted total metal concentrations showed a distinct increase during snowmelt as well as spikes during short-term runoff events at other times of the year. Although several metals reached the highest concentrations during snowmelt, aluminum, iron, and manganese had higher concentrations during short-term runoff events because their models were more sensitive to increases in turbidity. Unfortunately, this pattern is difficult to verify because few samples were collected during short-term runoff events in summer and fall. Samples also were lacking to verify total metal concentrations during the winter base-flow period.

In summary, use of regression models for predicting dissolved concentrations is unreliable for most metals because of the low concentrations of dissolved metals and a high proportion of censored concentrations, which reduced the number of samples in the calibration dataset. The regression models for predicting total metal concentrations also have a high degree of uncertainty, especially during short-term runoff events or winter months, because of a lack of samples collected under these flow conditions.

Table 7. Summary statistics, regression model equations, and evaluation statistics for dissolved and total metal concentrations for Animas River below Aztec, New Mexico (U.S. Geological Survey station 09364010), 2005–17.

[Models with R^2 of less than 0.5 are shown in gray. Model terms listed in order of increasing *p*-value. Min., minimum concentration; µg/L, microgram per liter; Med., median concentration; Max., maximum concentration; N, number of samples in calibration dataset; R^2, coefficient of determination adjusted for serial correlation; Std. error, standard error of estimate; Bias corr., bias correction factor (Duan, 1983); Pred. int., median prediction interval as a percent of the median estimated concentration; *log*, log base 10; *C*, metal concentration; *SC*, specific conductance in microsiemens per centimeter at 25 degrees Celsius; *sin*_2, sin(4π*T*), where *T* is the decimal portion of the year starting on January 1; *cos*_2, cos(4π*T*); *Temp*, water temperature in degrees Celsius; —, not computed; *Q*, streamflow in cubic feet per second; *sin*, sin(2π*T*); *cos*, cos(2π*T*); *Turb*, turbidity in nephelometric turbidity units]

Metal, fraction	Percent censored	Min. (µg/L)	Med. (µg/L)	Max. (µg/L)	N	Model equation and coefficients	R^2	Std. error	Bias corr.	Pred. int.
Aluminum, dissolved	33	4.7	56	86	22	$logC = -0.001896SC - 0.1974sin\_2 - 0.06818cos\_2 + 0.01324Temp + 1.974$	0.86	0.14	1.05	100
Arsenic, dissolved	63	0.26	0.47	2.1	13	—	—	—	—	—
Cadmium, dissolved	43	0.012	0.041	0.11	16	—	—	—	—	—
Copper, dissolved	21	1.1	1.8	4.2	29	$logC = -0.01849Temp - 0.3872logSC + 0.1056sin\_2 + 0.09212cos\_2 + 1.507$	0.56	0.11	1.02	50
Iron, dissolved	34	7.0	47	64	23	$logC = 0.6294logQ - 0.1736sin\_2 - 0.051876cos\_2 - 0.5621$	0.92	0.10	1.02	240
Lead, dissolved	59	0.05	0.45	1.5	14	—	—	—	—	—
Manganese, dissolved	0	4.7	23.6	100	42	$C = 0.1429SC + 0.01627Q + 5.970sin + 18.36cos - 49.25$	0.49	—	—	—
Zinc, dissolved	13	2.4	7.5	40	28	$logC = -1.032logSC - 0.03012Temp + 3.849$	0.65	0.18	1.09	130
Aluminum, total	0	143	1,150	16,800	41	$C = 39.01Turb - 1049pH + 8474$	0.84	1,510	—	110
Arsenic, total	9	0.52	1.6	5.6	42	$logC = 0.002265Turb - 0.8339logSC + 2.140$	0.77	0.12	1.04	100
Cadmium, total	2	0.040	0.30	1.4	41	$logC = 0.001808Turb + 0.0001420Q - 0.1913sin\_2 - 0.1582cos\_2 - 0.9976$	0.63	0.20	1.09	50
Copper, total	0	1.6	9.9	44	46	$logC = 0.003177Turb + 0.4204logQ - 0.2123sin\_2 - 0.05203cos\_2 - 0.6378$	0.79	0.16	1.06	80
Iron, total	0	140	1,870	17,500	39	$C = 44.88Turb - 1924sin\_2 + 279.9cos\_2 + 110.8Temp - 1816$	0.83	1,650	—	90
Lead, total	0	0.67	11	112	46	$logC = 0.003344Turb + 0.0002003Q - 0.3368sin\_2 - 0.06803cos\_2 + 0.2806$	0.78	0.23	1.14	120
Manganese, total	0	43	182	932	40	$C = 1.963Turb - 1521.5sin\_2 + 54.57cos\_2 + 8.924Temp - 28.22$	0.70	106	—	70
Zinc, total	3	6.0	73	211	39	$logC = 0.002212Turb - 0.2769sin\_2 - 0.1002cos\_2 + 0.4675logQ + 0.1481$	0.65	0.21	1.11	70

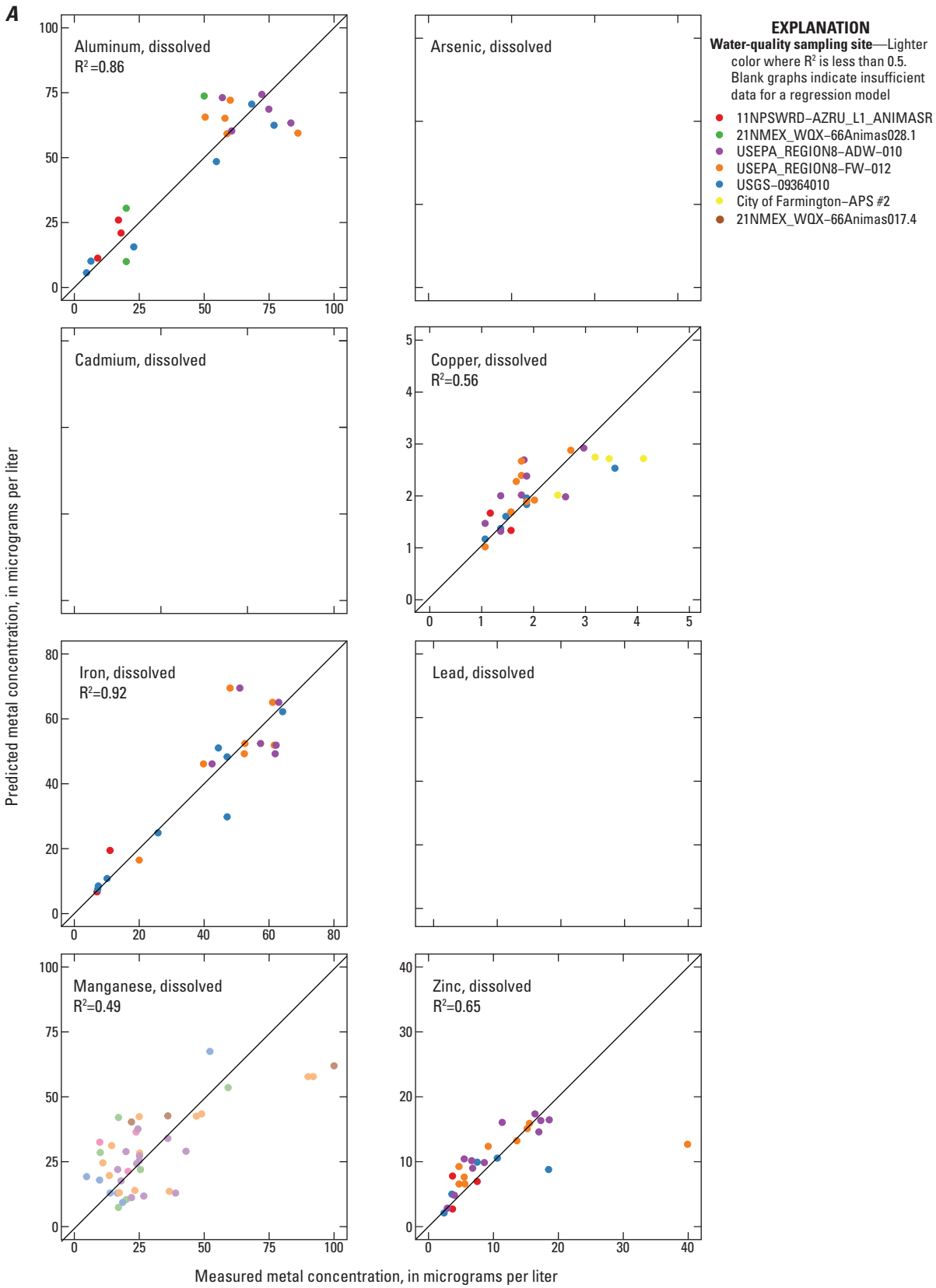

Figure 11. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for Animas River below Aztec, New Mexico (U.S. Geological Survey station 09364010), 2005–17. Blank graphs indicate insufficient data for a regression model.

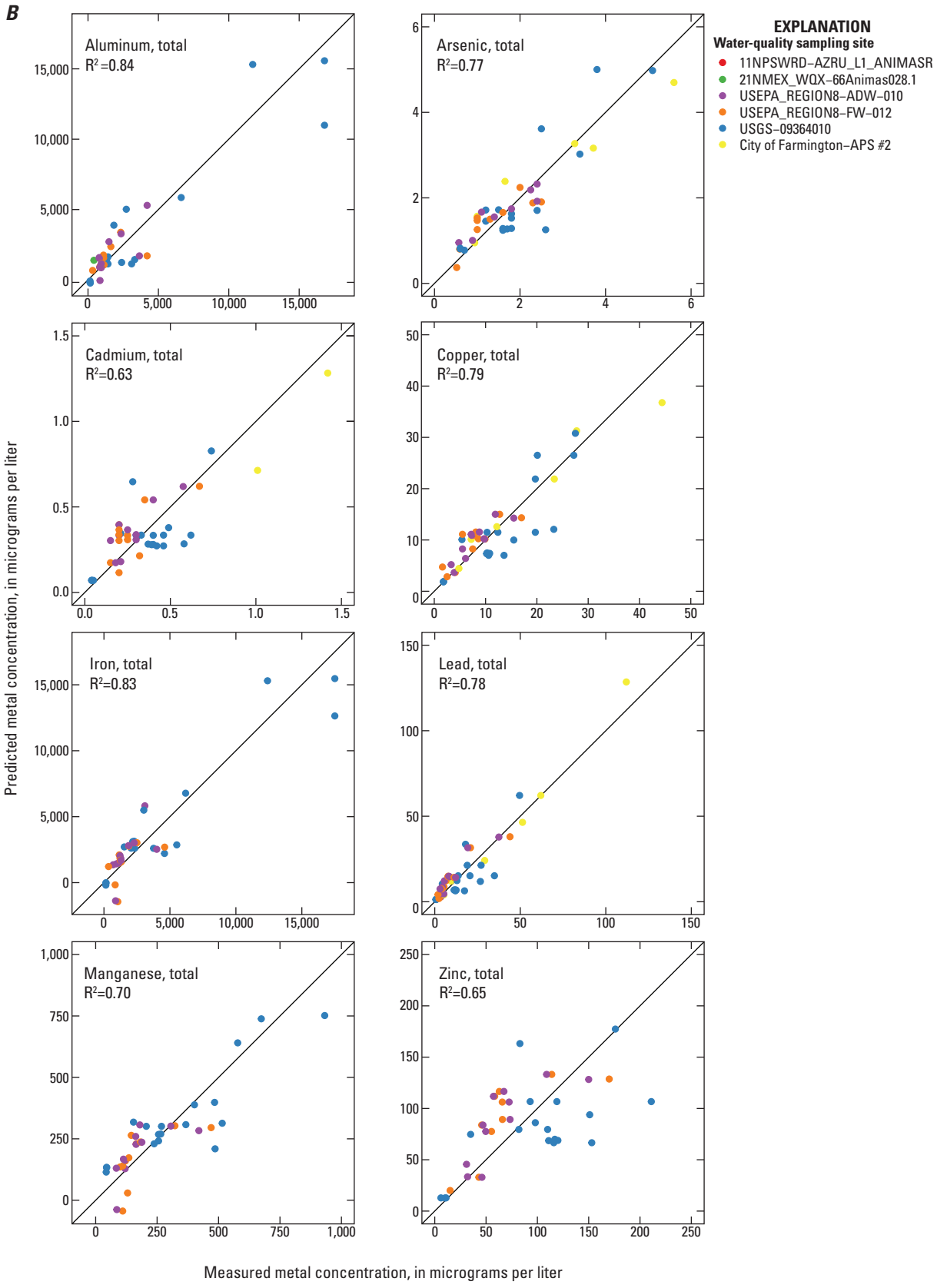

Figure 11. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for Animas River below Aztec, New Mexico (U.S. Geological Survey station 09364010), 2005–17. Blank graphs indicate insufficient data for a regression model.—Continued

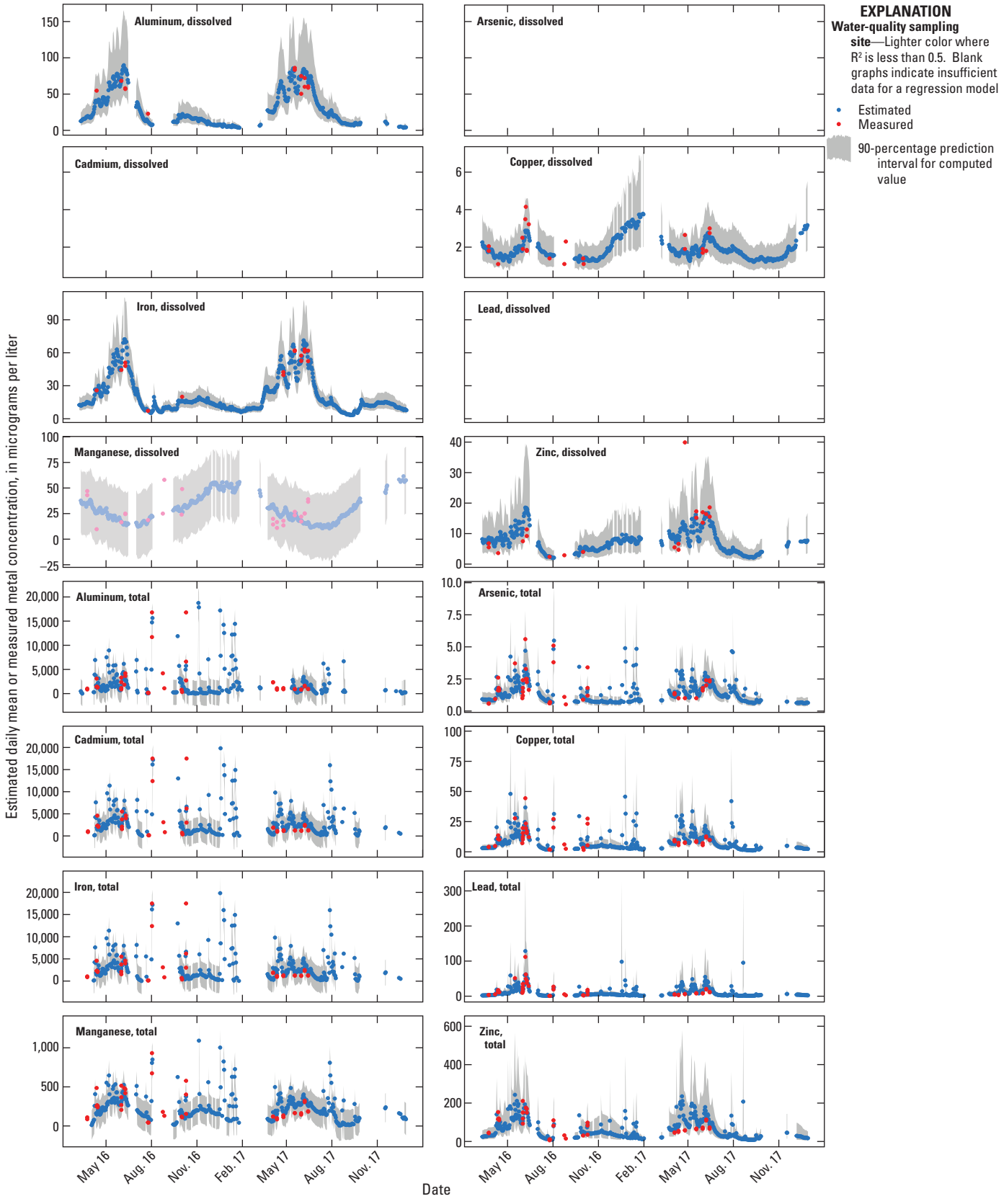

Figure 12. Time series plots of estimated daily and measured metal concentrations for Animas River below Aztec, New Mexico (U.S. Geological Survey station 09364010), for 2016–17. Gaps in estimated concentrations indicate periods of missing surrogate data. Blank graphs indicate insufficient data for a regression model.

San Juan River at Farmington, New Mexico (Station 09365000)

The dataset compiled for this site included 264 water-quality samples collected in the vicinity of the USGS station at San Juan River at Farmington (site 6 in fig. 1) with dissolved or total metal analyses for 1962–2017. Because of the lack of field data in the early part of the record and before installation of the continuous water-quality sensors, the calibration dataset included 29 samples collected from 2015–17 with associated surrogate parameters (streamflow, SC, pH, water temperature, and turbidity). Most dissolved metals had a substantial proportion of censored concentrations ranging from 27 percent for aluminum to 96 percent for cadmium. Total metals had censored concentrations in 20 percent or less of samples with the exception of cadmium, which had 72 percent censored (table 8). Summary statistics for the final dataset and resulting regression equations and statistics are presented in table 8 and graphs of predicted versus measured concentrations for all eight metals are shown in figure 13.

Adequate models (R^2 greater than 0.50) were found for none of the dissolved metals and six of the total metals: aluminum, copper, iron, lead, manganese, and zinc. Six of the dissolved metals (aluminum, cadmium, copper, iron, lead, and zinc) and total cadmium had insufficient data to compute a regression model. There was little consistency in the form of the model among the total metals, as three of them had seasonal terms as the main explanatory variable, one had turbidity, and one had streamflow. The calibration dataset originally included a sample collected on November 5, 2016, following a runoff event that caused a sharp increase in streamflow and turbidity. Because total metal concentrations in this sample were an order of magnitude greater than the other samples in the calibration dataset, the sample strongly leveraged the regression model and made it perform poorly for lower concentration samples. For this reason, this sample was not included in the final calibration dataset.

To further evaluate model performance, daily mean concentrations were estimated from the corresponding regression equations in table 8 using the daily streamflow and water-quality data from the USGS station at San Juan River at Farmington for the period April 2016 through December 2017 (fig. 14). Although the water-quality sensor was operated year round at this station, there were several extended periods of missing record for which concentrations could not be estimated. The wide range of median prediction intervals (110 to 150 percent of the median predicted concentration) (table 8) indicates that concentrations for most total metals could not be reliably estimated using surrogates at this site. One issue may be the small size and representativeness of the calibration dataset, which had only 20–25 samples that were not evenly distributed over the seasons.

The models for aluminum and iron predicted numerous spikes in concentration during short-term runoff events when turbidity was elevated. Unfortunately, few discrete samples were available to verify the model performance during these events. Although the lack of suitable models for this site is partially caused by limitations in the calibration dataset, the lack may also be because the site is located 0.5 mi below the confluence with the Animas River that is responsible for, on average, 50 percent of the flow at the San Juan River at Farmington site. Considering differences in flow regulation between the Navajo Reservoir and unregulated snowmelt, and the source of metals (mining versus sedimentary rocks) between the two rivers, it is not surprising that the metal chemistry at this site might be complicated by the mixing and timing of different solute and water sources. Although beyond the scope of the current study, an approach that has been used to mix inputs from different drainages implements lagging streamflows in the regression model to account for different travel times of source water (Welch and others, 2014). The methods of water-quality sampling (equal-width-increment method versus grab sampling) and the location of the water-quality sensor also may be a consideration if the river is not well mixed at the sampling point or if water-quality samples do not correlate with the location of the water-quality sensor.

In summary, the use of surrogates to predict metal concentrations at this site is not recommended and may be more challenging and require substantially more sampling than is normally needed to establish surrogate relationships for water quality.

Table 8. Summary statistics, regression model equations, and evaluation statistics for dissolved and total metal concentrations for San Juan River at Farmington, New Mexico (U.S. Geological Survey station 09365000), 2015–17.

[Models with R^2 of less than 0.5 are shown in gray. Model terms listed in order of increasing *p*-value. Min., minimum concentration; µg/L, microgram per liter; Med., median concentration; Max., maximum concentration; N, number of samples in calibration dataset; R^2, coefficient of determination adjusted for serial correlation; Std. error, standard error of estimate; Bias corr., bias correction factor (Duan, 1983); Pred. int., median prediction interval as a percent of the median estimated concentration; —, not computed; C, metal concentration; *sin*_2, sin(4π*T*), where *T* is the decimal portion of the year starting on January 1; *cos*_2, cos(4π*T*); *log*, log base 10; *Temp*, water temperature in degrees Celsius; *sin*, sin(2π*T*); *cos*, cos(2π*T*); *Turb*, turbidity in nephelometric turbidity units; *Q*, streamflow in cubic feet per second; *pH*, pH in standard units]

Metal, fraction	Percent censored	Min. (µg/L)	Med. (µg/L)	Max. (µg/L)	N	Model equation and coefficients	R^2	Std. error	Bias corr.	Pred. int.
Aluminum, dissolved	27	7.1	22	52	19	—	—	—	—	—
Arsenic, dissolved	7	0.37	0.62	1.4	28	$C = 0.1945sin\_2 + 0.1328cos\_2 + 0.7167$	0.29	—	—	—
Cadmium, dissolved	96	0.044	0.044	0.044	1	—	—	—	—	—
Copper, dissolved	45	0.54	1.75	2.7	16	—	—	—	—	—
Iron, dissolved	38	11	23.6	59	18	—	—	—	—	—
Lead, dissolved	86	0.124	0.37	0.68	4	—	—	—	—	—
Manganese, dissolved	0	2.9	9.1	16	29	$logC = -0.02995Temp - 0.1338sin - 0.1637cos + 1.207$	0.20	—	—	—
Zinc, dissolved	53	0.85	3.5	6.7	14	—	—	—	—	—
Aluminum, total	0	330	1,900	10,000	25	$C = 14.96Turb + 1034$	0.66	1,160	—	120
Arsenic, total	0	0.56	1.3	3.8	25	$C = 0.004291Turb + 0.0001078Q + 0.8304$	0.40	—	—	—
Cadmium, total	72	0.07	0.16	0.34	7	—	—	—	—	—
Copper, total	4	1.8	5.4	23	24	$logC = 0.1492sin + 0.1159cos + 0.00006060Q + 0.001139Turb + 0.4759$	0.66	0.166	1.06	110
Iron, total	0	1,200	2,720	20,000	25	$C = 27.53Turb + 0.4909Q - 386.1sin - 2594cos + 311.7$	0.78	1,710	—	120
Lead, total	20	2.5	7.5	32	20	$logC = 0.00009351Q + 0.001070Turb + 0.4551$	0.63	0.22	1.14	150
Manganese, total	0	42	130	700	25	$logC = 0.4483sin + 0.8454cos + 0.0001538Q + 0.1021Temp + 0.7340$	0.62	0.21	1.10	150
Zinc, total	0	4.3	42	140	25	$logC = 2.835sin + 0.3305cos + 0.00008232Q - 0.6210pH + 6.391$	0.77	0.206	1.10	150

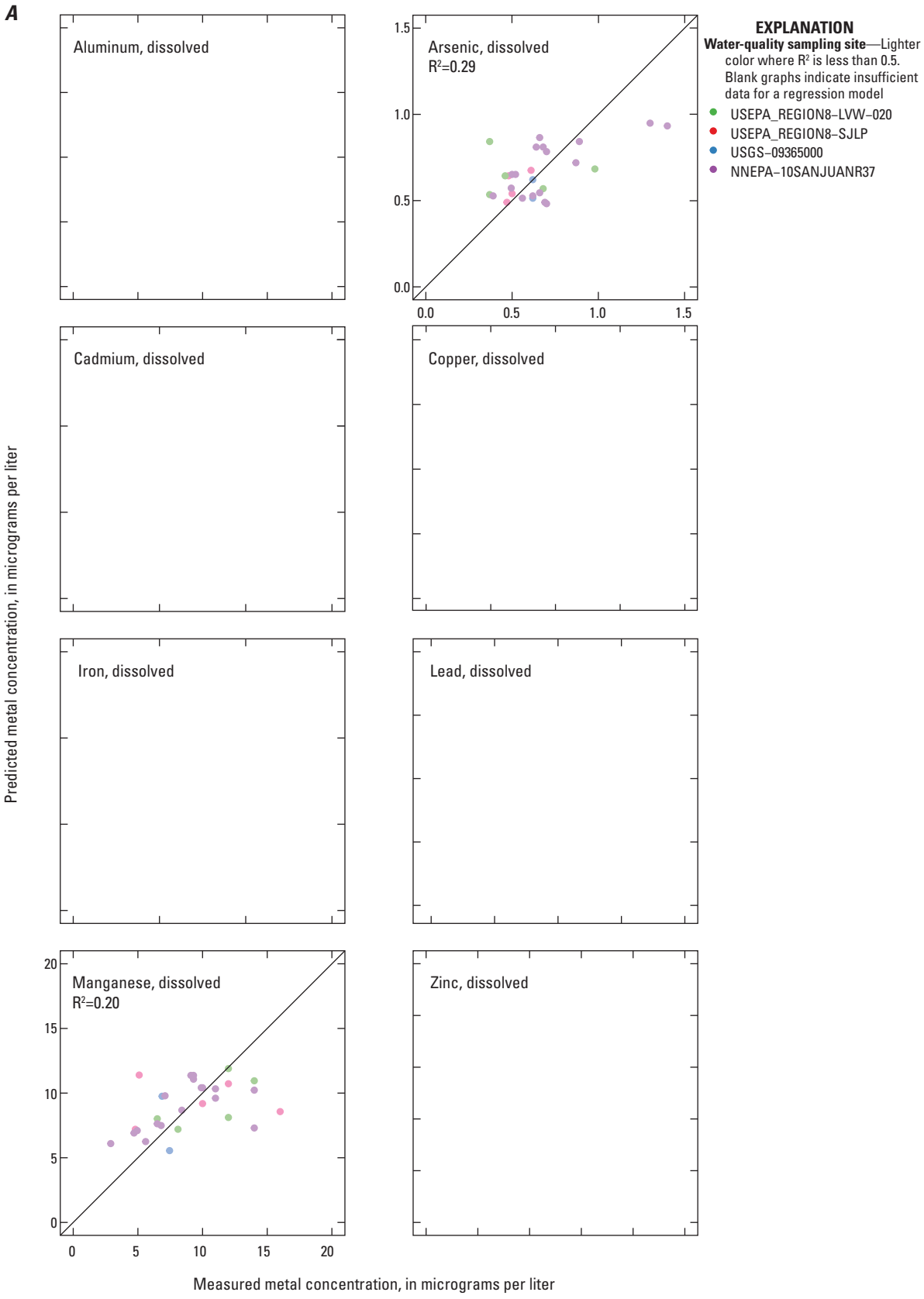

Figure 13. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for San Juan River at Farmington, New Mexico (U.S. Geological Survey Station 09365000), 2015–17. Blank graphs indicate insufficient data for a regression model.

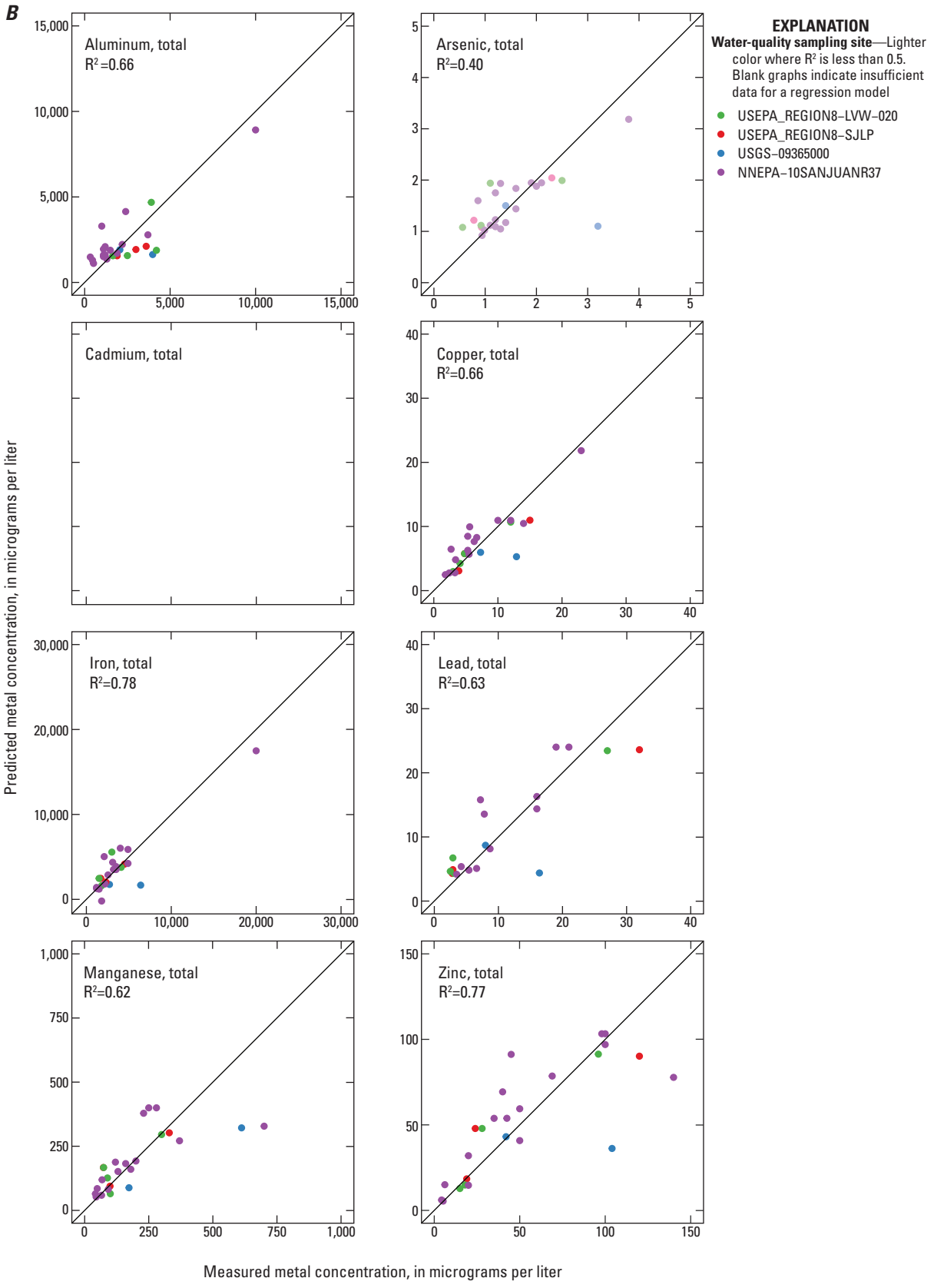

Figure 13. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for San Juan River at Farmington, New Mexico (U.S. Geological Survey Station 09365000), 2015–17. Blank graphs indicate insufficient data for a regression model.—Continued

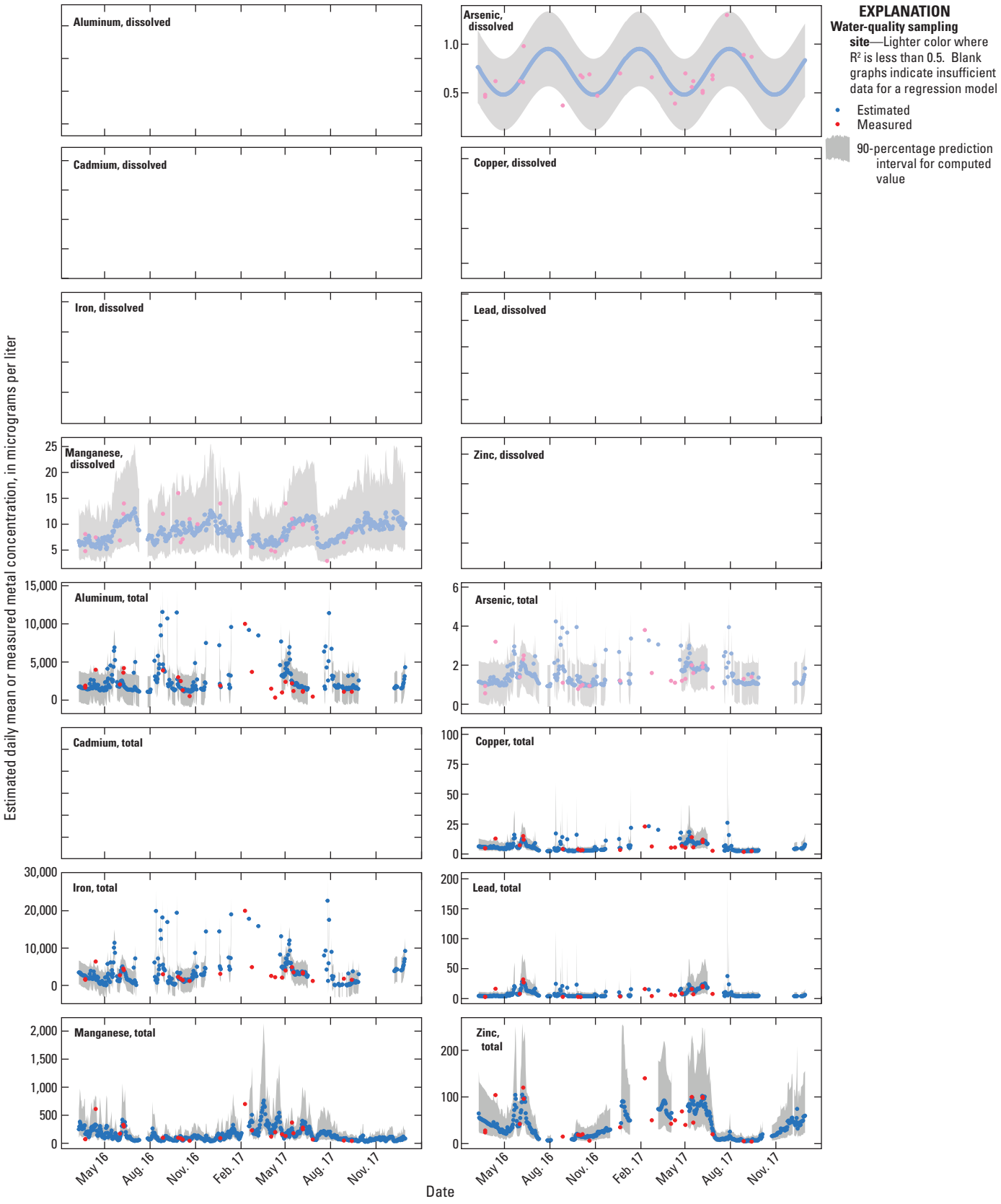

Figure 14. Time series plots of estimated daily mean and measured metal concentrations for San Juan River at Farmington, New Mexico (U.S. Geological Survey station 09365000), for 2016–17. Gaps in estimated concentrations indicate periods of missing surrogate data. Blank graphs indicate insufficient data for a regression model.

San Juan River at Shiprock, New Mexico (Station 09368000)

The dataset compiled for this site included 303 water-quality samples collected in the vicinity of the USGS station at San Juan River at Shiprock (site 7 in fig. 1) with dissolved or total metal analyses for 1958–2017. Because of the lack of field data in the early part of the record and before installation of the continuous water-quality sensors, the calibration dataset included 32 samples collected from 2015–17 with associated surrogate parameters (streamflow, SC, pH, water temperature, and turbidity). All the dissolved metals except manganese had a substantial proportion of censored concentrations, which ranged from 27 percent for arsenic to 93 percent for cadmium. Total metals had censored concentrations in less than 20 percent of samples with the exception of 61 percent for cadmium (table 9). Summary statistics for the final dataset and resulting regression equations and statistics are presented in table 9 and graphs of predicted versus measured concentrations for the eight metals are shown in figure 15.

Adequate models (R^2 greater than 0.50) were found for all of the total metals except cadmium but none were found for any of the dissolved metals. Dissolved cadmium, copper, lead, and zinc and total cadmium had insufficient data to compute a regression model. All the total metal models had turbidity as the main explanatory variable. Three of the total models included pH as a secondary variable and one included streamflow.

To further evaluate model performance, daily mean concentrations were estimated from the corresponding regression equations in table 9 using the daily streamflow and water-quality data from the USGS station at San Juan River at Shiprock for the period April 2016 through December 2017 (fig. 16). Although the water-quality sensor was operated year round at this station, there were many periods of missing record, especially for turbidity when concentrations were not computed.

The models for total metals performed better than for dissolved metals possibly because there was a wider range of concentrations; however, the median prediction intervals were mostly much higher than 100 percent of the median concentration, which indicates a high degree of uncertainty for most estimated concentrations (table 9). Because turbidity was the main explanatory variable in all models, some relatively large spikes in total metal concentrations were predicted during runoff events when turbidity was elevated. However, water-quality sampling during these types of events was limited so this result is difficult to validate.

In summary, use of surrogate models to predict dissolved and total metal concentrations at this site is not recommended until the models can be validated with additional samples collected over a wider range of metal concentrations and flow conditions.

Table 9. Summary statistics, regression model equations, and evaluation statistics for dissolved and total metal concentrations for San Juan River at Shiprock, New Mexico (U.S. Geological Survey station 09368000), 2015–17.

[Models with R^2 of less than 0.5 are shown in gray. Model terms listed in order of increasing *p*-value. Min., minimum concentration; µg/L, microgram per liter; Med., median concentration; Max., maximum concentration; N, number of samples in calibration dataset; R^2, coefficient of determination adjusted for serial correlation; Std. error, standard error of estimate; Bias corr., bias correction factor (Duan, 1983); Pred. int., median prediction interval as a percent of the median estimated concentration; *log*, log base 10; *C*, metal concentration; *Q*, streamflow in cubic feet per second; —, not computed; *sin*, sin(2π*T*), where *T* is the decimal portion of the year starting on January 1; *cos*, cos(2π*T*); *Temp*, water temperature in degrees Celsius; *Turb*, turbidity in nephelometric turbidity units; *pH*, pH in standard units]

Metal, fraction	Percent censored	Min. (µg/L)	Med. (µg/L)	Max. (µg/L)	N	Model equation and coefficients	R^2	Std. error	Bias corr.	Pred. int.
Aluminum, dissolved	28	5.1	22	89.5	21	$logC = 0.4248logQ - 0.09438$	0.25	—	—	—
Arsenic, dissolved	27	0.43	0.665	0.98	24	$C = 0.05236sin - 0.1973cos - 0.3348logQ + 1.706$	0.49	—	—	—
Cadmium, dissolved	93	0.08	0.084	0.088	2	—	—	—	—	—
Copper, dissolved	48	0.83	1.8	8.7	17	—	—	—	—	—
Iron, dissolved	28	4.5	31.4	160	23	$logC = 0.2573sin - 0.1572cos + 1.293$	0.29	—	—	—
Lead, dissolved	89	0.2	0.432	0.59	3	—	—	—	—	—
Manganese, dissolved	3	2.3	5.12	21	32	$logC = 0.0001167Q + 0.05834Temp + 0.05505sin + 0.5612cos - 0.1958$	0.19	—	—	—
Zinc, dissolved	57	3.4	14	128	13	—	—	—	—	—
Aluminum, total	0	315	3,300	91,000	28	$C = 25.86Turb + 2105$	0.88	6,880	—	120
Arsenic, total	0	0.76	2	18	28	$C = 0.004665Turb + 1.176logQ - 2.389$	0.93	0.95	—	90
Cadmium, total	61	0.09	0.17	0.4	11	—	—	—	—	—
Copper, total	11	2.2	10.4	200	25	$C = 0.05584Turb + 5.606$	0.86	17	—	120
Iron, total	0	450	3,560	160,000	28	$C = 45.55Turb + 1503$	0.87	12,600	—	140
Lead, total	14	3.5	7.78	94	24	$C = 0.02458Turb - 31.64pH + 264.6$	0.89	6.6	—	210
Manganese, total	0	22.5	180	2,200	28	$C = 0.5929Turb - 429.3pH + 3670$	0.86	175	—	170
Zinc, total	4	3.6	44	520	27	$C = 0.1362Turb - 126.9pH + 1080$	0.86	42	—	150

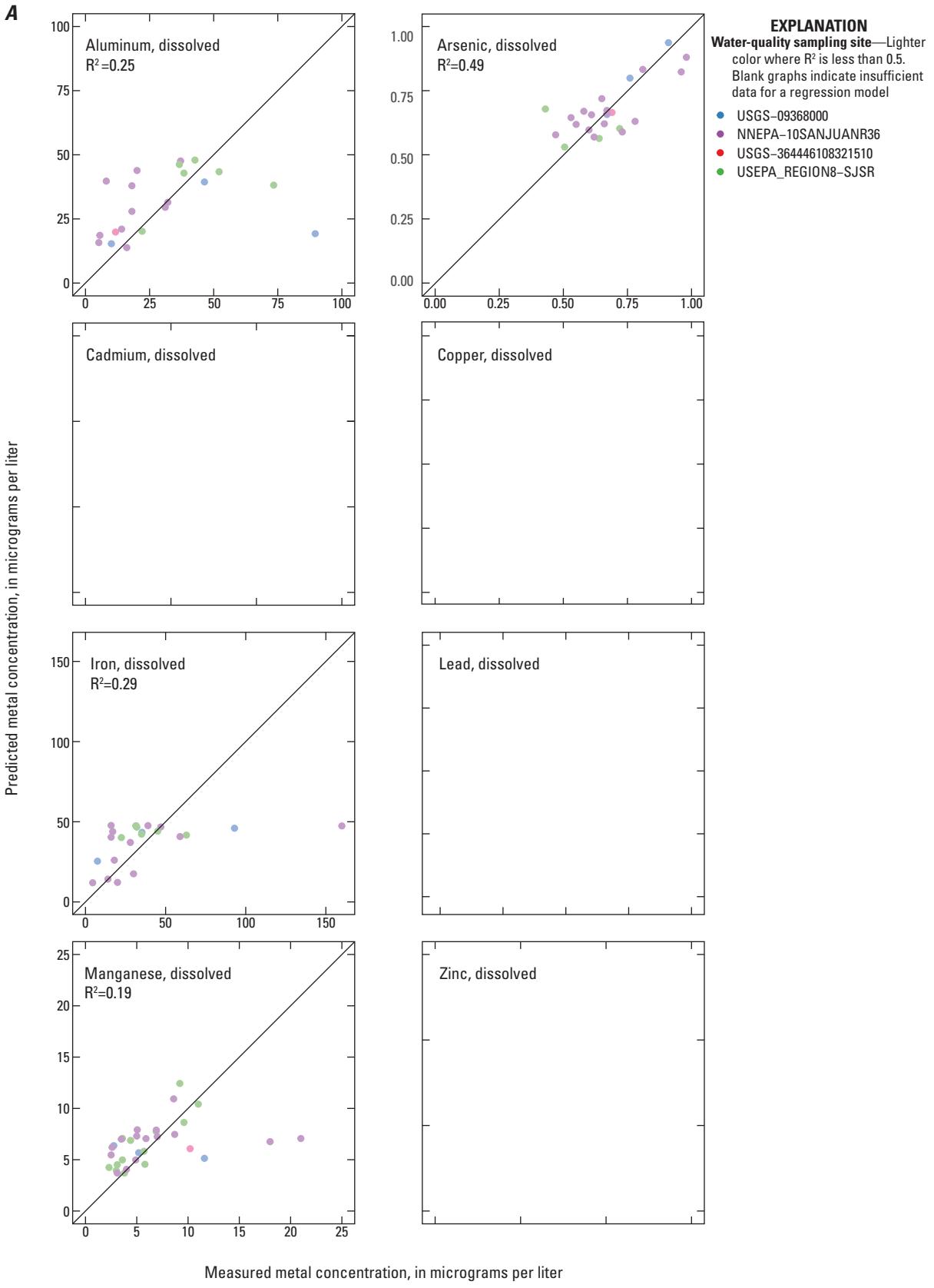

Figure 15. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for San Juan River at Shiprock, New Mexico (U.S. Geological Survey station 09368000), 2015–17. Blank graphs indicate insufficient data for a regression model.

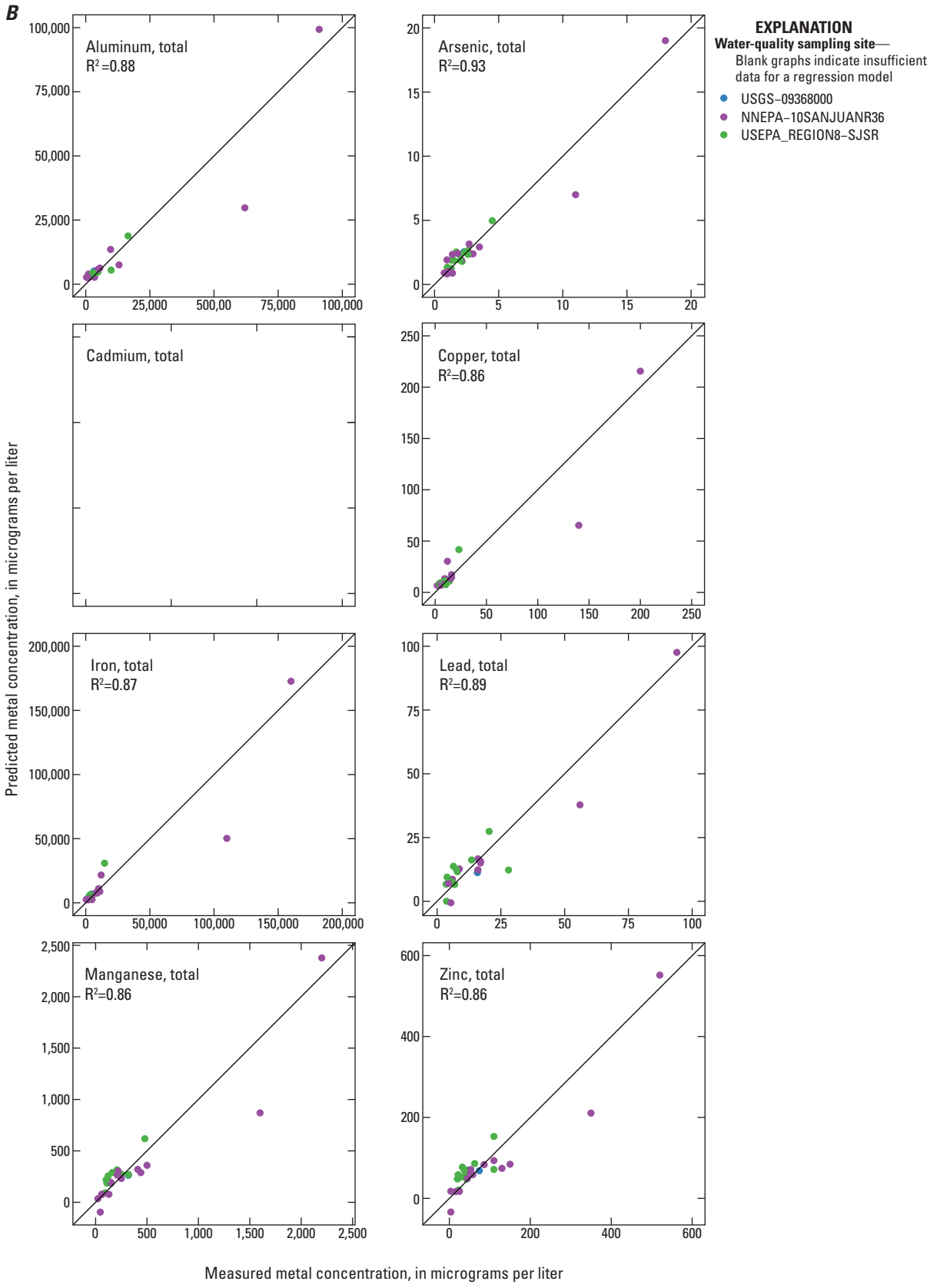

Figure 15. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for San Juan River at Shiprock, New Mexico (U.S. Geological Survey station 09368000), 2015–17. Blank graphs indicate insufficient data for a regression model.—Continued

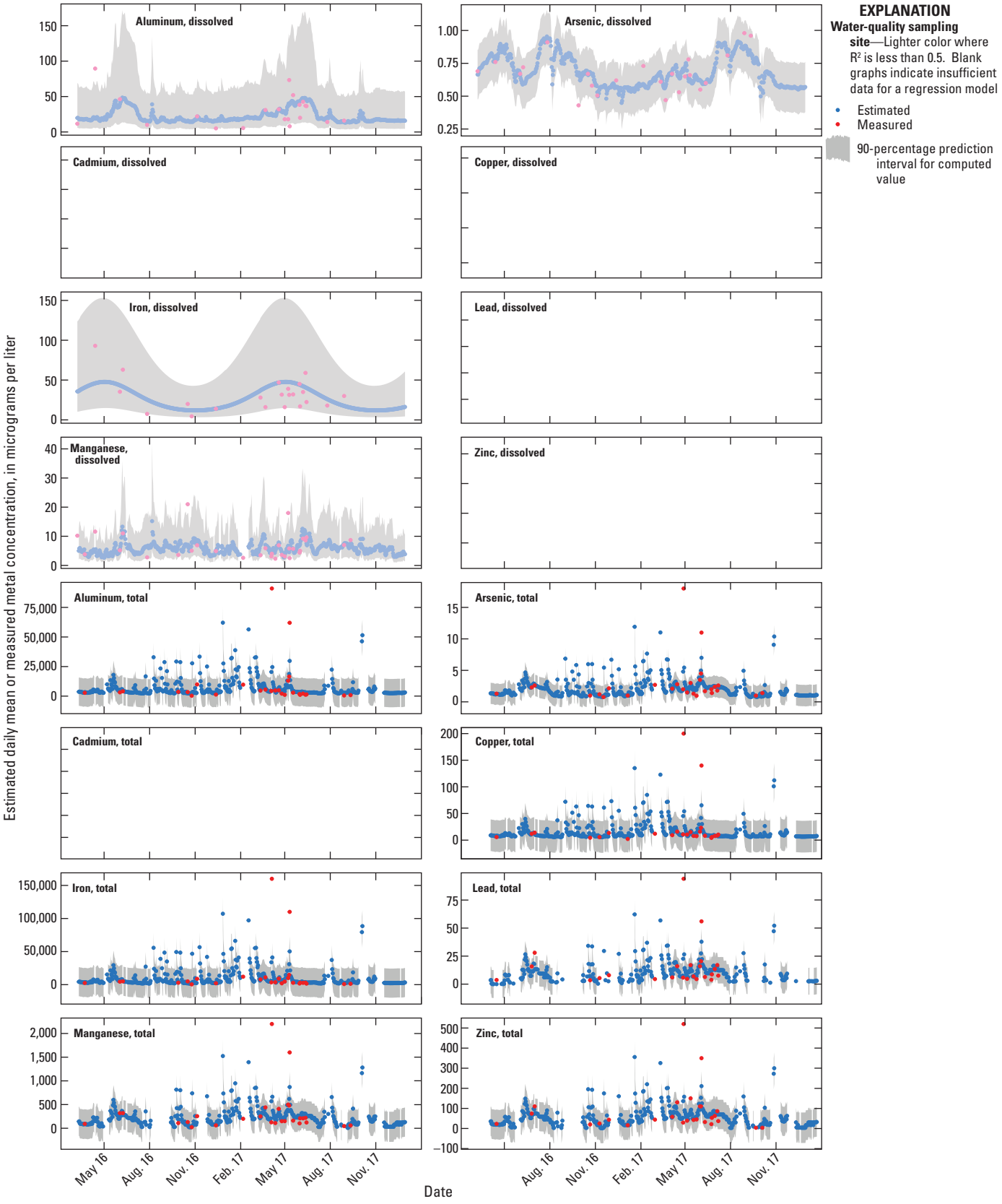

Figure 16. Time series plots of estimated daily mean and measured metal concentrations for San Juan River at Shiprock, New Mexico (U.S. Geological Survey station 09368000), for 2016–17. Gaps in estimated concentrations indicate periods of missing surrogate data. Blank graphs indicate insufficient data for a regression model.

San Juan River at Four Corners, Colorado (Station 09371010)

The dataset compiled for this site included 294 water-quality samples collected in the vicinity of the USGS station at San Juan River at Four Corners (site 8 in fig. 1) with dissolved or total metal analyses for 1968–2017. Because of the lack of field data in the early part of the record and before installation of the continuous water-quality sensors, the calibration dataset included 56 samples collected from 2015–17 with associated surrogate parameters (streamflow, SC, pH, water temperature, and turbidity). All the dissolved metals had a substantial proportion of censored concentrations, which ranged from 21 percent for manganese to 91 percent for cadmium. Total metals had censored concentrations in less than 20 percent of samples with the exception of cadmium, which had 54 percent censored (table 10). Summary statistics for the final dataset and resulting regression equations and statistics are presented in table 10 and graphs of predicted versus measured concentrations for the eight metals are shown in figure 17.

Adequate models (R^2 greater than 0.50) were found for only one dissolved metal (zinc) and for all eight of the total metals. Dissolved cadmium and lead had insufficient data to compute a regression model. All the total metal models had turbidity as the main explanatory variable and streamflow, pH, SC, and seasonal terms as secondary variables.

To further evaluate model performance, daily mean concentrations were estimated from the corresponding regression equations in table 10 and the daily streamflow and water-quality data from the USGS station at San Juan River at Four Corners, Colo., for the period April 2016 through December 2017 (fig. 18). Although the water-quality sensor was operated year round at this station, there were several extended periods of missing record during which concentrations could not be estimated.

Zinc was the only dissolved metal with a model R^2 of more than 0.5; however, the median prediction interval exceeded the median concentration by 340 percent (table 10), indicating that estimated zinc concentrations may be unreliable. The models for total metals generally performed better than the models for dissolved metals possibly because of a wider range of concentrations and fewer censored values. Because turbidity was the main explanatory variable in all models, some relatively large spikes in total metal concentrations were predicted during runoff events when turbidity was elevated. However, water-quality sampling during these types of events was limited, so this result is difficult to validate. For all the total metals except arsenic, the median prediction interval exceeded the median concentration by 100 percent or more, indicating a high degree of uncertainty in the metal predictions.

In summary, use of surrogate models to predict dissolved and total metal concentrations at this site is not recommended until the models can be validated with additional samples collected over a wider range of metal concentrations and flow conditions. In addition, few discrete samples were available to verify the model performance especially during fall and winter months.

Table 10. Summary statistics, regression model equations, and evaluation statistics for dissolved and total metal concentrations for San Juan River at Four Corners, Colorado (U.S. Geological Survey station 09371010), 2015–17.

[Models with R^2 of less than 0.5 are shown in gray. Model terms listed in order of increasing *p*-value. Min., minimum concentration; µg/L, microgram per liter; Med., median concentration; Max., maximum concentration; N, number of samples in calibration dataset; R^2, coefficient of determination adjusted for serial correlation; Std. error, standard error of estimate; Bias corr., bias correction factor (Duan, 1983); Pred. int., median prediction interval as a percent of the median estimated concentration; *C*, metal concentration; *sin*, sin(2π*T*), where *T* is the decimal portion of the year starting on January 1; *cos*, cos(2π*T*); *SC*, specific conductance in microsiemens per centimeter at 25 degrees Celsius; *log*, log base 10; *Q*, streamflow in cubic feet per second; —, not computed; *sin_2*, sin(4π*T*); *cos_2*, cos(4π*T*); *pH*, pH in standard units; *Temp*, water temperature in degrees Celsius; *Turb*, turbidity in nephelometric turbidity units]

Metal, fraction	Percent censored	Min. (µg/L)	Med. (µg/L)	Max. (µg/L)	N	Model equation and coefficients	R^2	Std. error	Bias corr.	Pred. int.
Aluminum, dissolved	35	10	33.8	535	36	$C = -145.1 sin + 100.3 cos + 0.8960 SC + 351.7 logQ - 1.369$	0.47	—	—	—
Arsenic, dissolved	52	0.52	0.8	1.2	30	$logC = 0.1096 sin\_2 + 0.008042 cos\_2 + 0.2754 pH + 0.00002116 Q - 2.429$	0.38	—	—	—
Cadmium, dissolved	91	0.03	0.063	0.095	5	—	—	—	—	—
Copper, dissolved	29	0.64	1.9	14.6	44	$logC = 0.0004090 SC + 0.1397$	0.02	—	—	—
Iron, dissolved	47	4.7	39.9	410	35	$logC = 0.6666 logQ + 0.3736 sin\_2 - 0.1831 cos\_2 - 0.4913$	0.27	—	—	—
Lead, dissolved	75	0.063	0.481	3.4	15	—	—	—	—	—
Manganese, dissolved	21	0.5	4.45	14.4	46	$C = -4.454 pH + 0.0003087 Q + 39.79$	0.29	—	—	—
Zinc, dissolved	57	2.2	7.35	104	26	$logC = -0.5264 sin\_2 - 0.3356 cos\_2 - 0.8116 pH + 7.606$	0.60	0.35	1.29	340
Aluminum, total	0	268	3,560	95,000	56	$logC = 0.8617 logTurb - 0.4360 pH + 5.305$	0.70	0.29	1.23	210
Arsenic, total	19	0.76	2	12	46	$logC = 0.0007205 Turb + 0.1395 sin - 0.02579 cos + 0.00002288 Q + 0.04663$	0.81	0.11	1.03	60
Cadmium, total	54	0.09	0.2	0.815	26	$logC = 0.0005810 Turb - 0.2564 sin\_2 - 0.0490 cos\_2 - 0.9053$	0.61	0.17	1.06	100
Copper, total	4	1.7	8.9	63	55	$logC = 0.6110 logTurb + 0.1163 sin - 0.004477 cos - 0.0006105 SC - 0.1404$	0.71	0.19	1.09	110
Iron, total	0	318	4,180	50,000	56	$logC = 0.9027 logTurb - 0.5059 pH + 5.843$	0.81	0.23	1.13	140
Lead, total	11	0.8	9.16	46	51	$logC = 0.6517 logTurb + 0.3452 logQ + 0.1335 sin - 0.1446 cos - 1.733$	0.71	0.21	1.11	140
Manganese, total	2	3.8	190	1,100	55	$logC = 0.7808 logTurb + 0.5116 logQ + 0.02360 sin\_2 - 0.1280 cos\_2 - 1.092$	0.81	0.21	1.10	130
Zinc, total	11	4.3	48	260	51	$logC = 0.6397 logTurb + 0.1759 sin - 0.2262 cos - 0.2806 pH + 2.413$	0.77	0.20	1.10	120

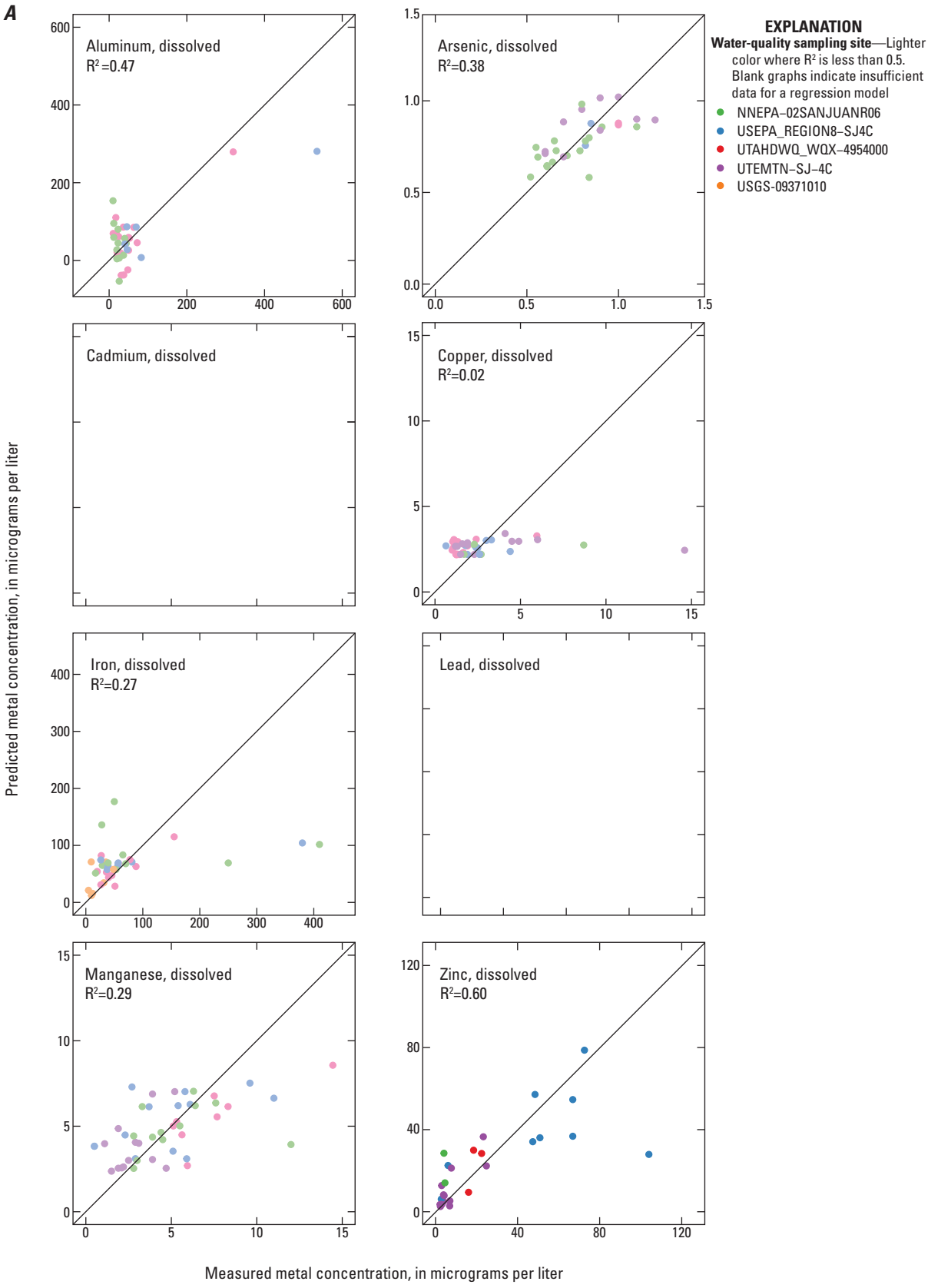

Figure 17. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for San Juan River at Four Corners, Colorado (U.S. Geological Survey station 09371010), 2015–17. Blank graphs indicate insufficient data for a regression model.

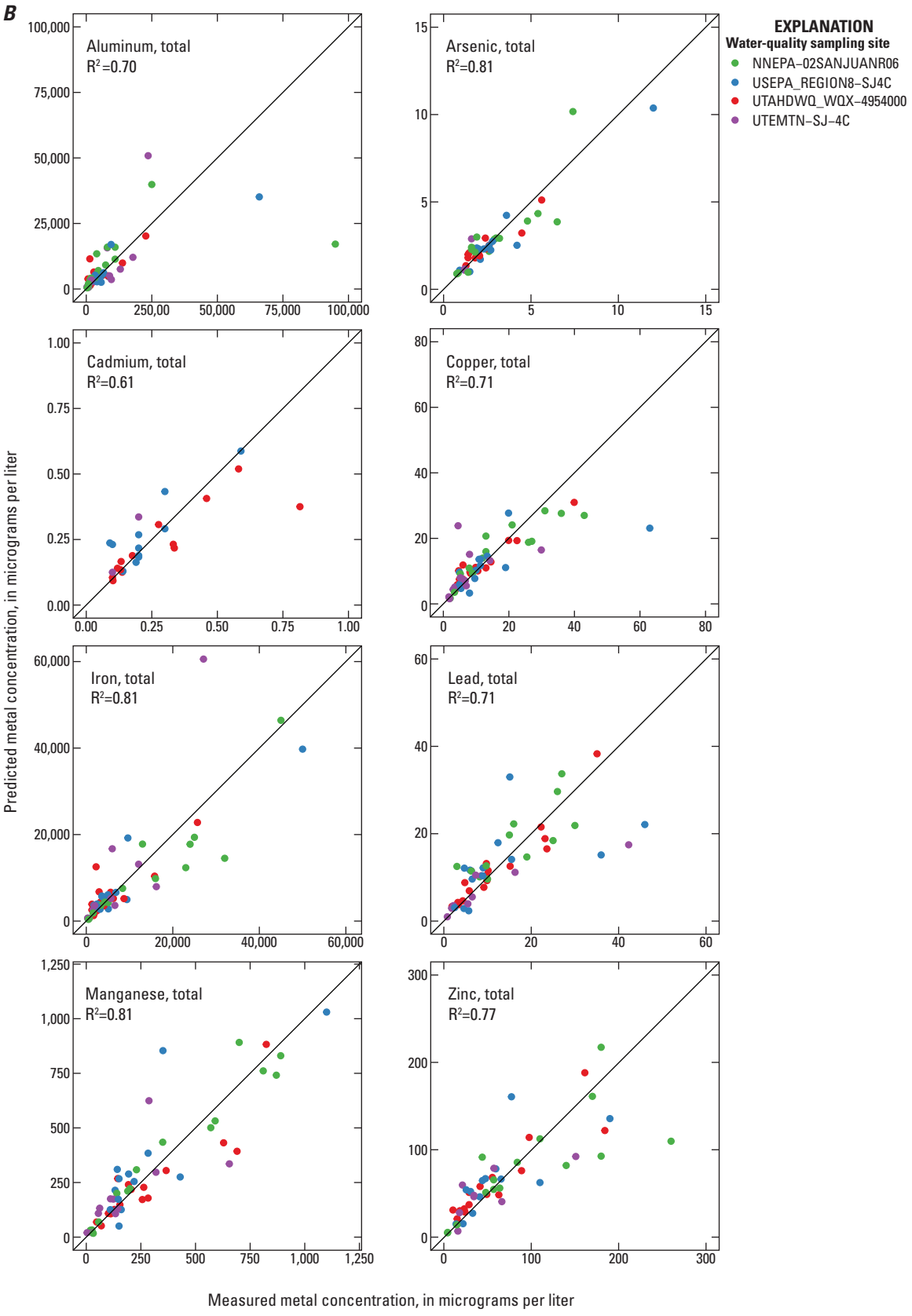

Figure 17. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for San Juan River at Four Corners, Colorado (U.S. Geological Survey station 09371010), 2015–17. Blank graphs indicate insufficient data for a regression model.—Continued

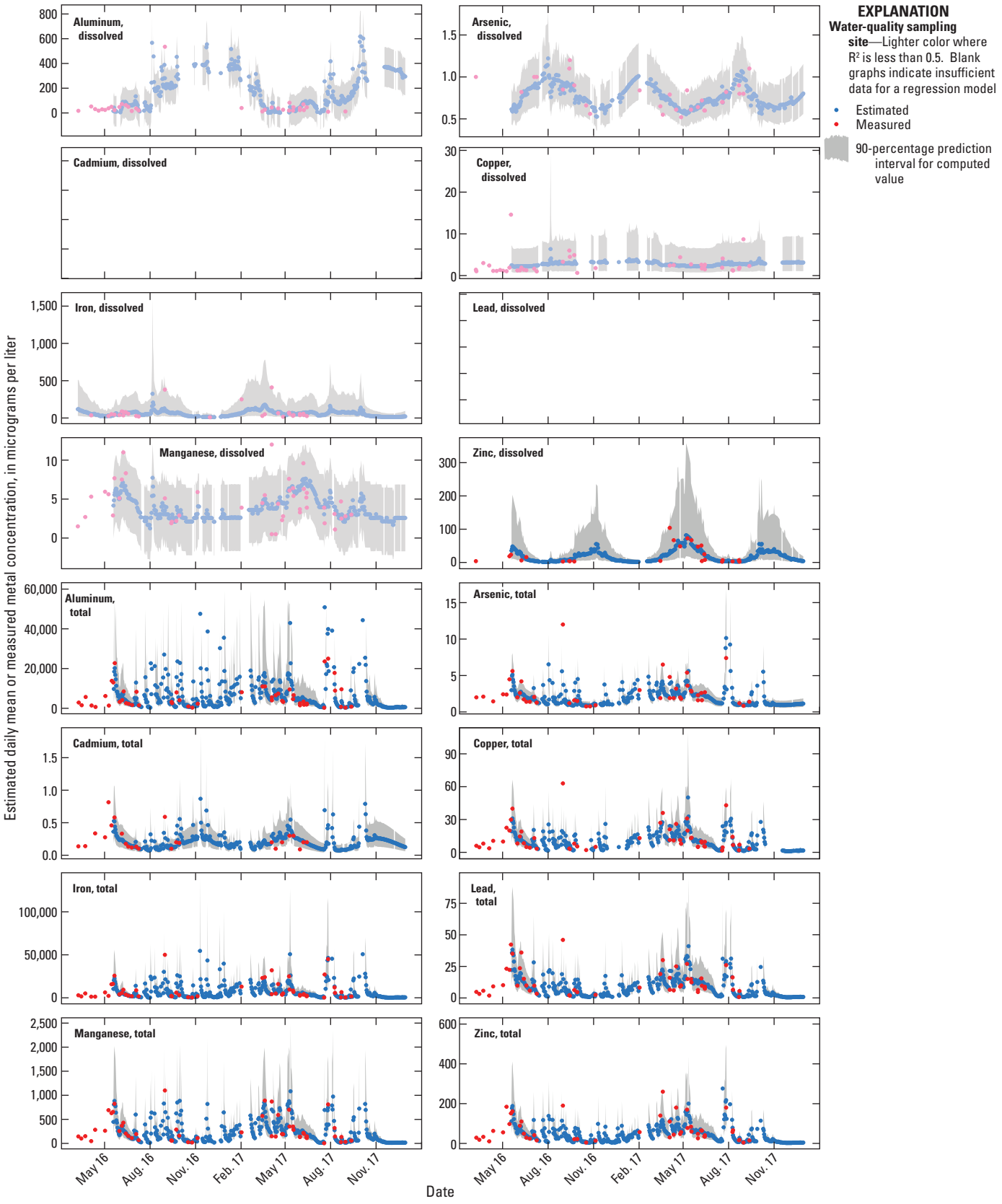

Figure 18. Time series plots of estimated daily mean and measured metal concentrations for San Juan River at Four Corners, Colorado (U.S. Geological Survey station 09371010), for 2016–17. Gaps in estimated concentrations indicate periods of missing surrogate data. Blank graphs indicate insufficient data for a regression model.

San Juan River near Bluff, Utah (Station 09379500)

The dataset compiled for this site included 1,711 water-quality samples collected in the vicinity of the USGS station at San Juan River near Bluff (site 9 in fig. 1) with dissolved or total metal analyses for 1928–2017. Because of the lack of field data in the early part of the record and before installation of the continuous water-quality sensors, the calibration dataset included 46 samples collected from 2015–17 with associated surrogate parameters (streamflow, SC, pH, water temperature, and turbidity). Of the dissolved metals, a substantial proportion of concentrations were censored for cadmium (88 percent), lead (70 percent), and zinc (59 percent). Of the total metal concentrations, only cadmium was censored (42 percent) at greater than 10 percent (table 11). Summary statistics for the final dataset and resulting regression equations and statistics are presented in table 11 and graphs of predicted versus measured concentrations for all eight metals are shown in figure 19.

Adequate models (R^2 greater than 0.50) were found for two of the dissolved metals (aluminum and arsenic) and all eight of the total metals. Dissolved cadmium, lead, and zinc had insufficient data to compute regression models. In the dissolved models, SC or seasonal terms were the main explanatory variable. All the total models had turbidity as the main explanatory variable. Five of the total models had seasonal terms as secondary variables, two had pH, and one had temperature.

To further evaluate model performance, daily mean concentrations were estimated from the corresponding regression equations in table 11 using the daily streamflow and water-quality data from the USGS station at San Juan River near Bluff for the period April 2016 through December 2017 (fig. 20). At this station, the continuous water-quality record was fairly complete with only a few periods of missing turbidity data. Although the dissolved aluminum model R^2 was more than 0.5, the median prediction interval exceeded the median concentration by 110 percent (table 11), indicating a fairly large degree of uncertainty in the model predictions. The prediction interval for dissolved arsenic concentrations was slightly better (40 percent), although metals with little variability in concentrations may not necessitate a predictive model.

The models for total metals generally performed better than for dissolved metals possibly because of a wider range of concentrations and fewer censored values. Because turbidity was the main explanatory variable in all the total models, some relatively large spikes in total metal concentrations were predicted during runoff events when turbidity was elevated. However, water-quality sampling during these types of events was limited, so this result is difficult to validate. For all the total metals except arsenic and manganese, the median prediction interval exceeded the median concentration by 100 percent or more, indicating a fairly large degree of uncertainty in the model predictions (table 11).

In summary, use of surrogate models to predict dissolved and total metal concentrations at this site is not recommended until the models can be validated with additional samples collected over a wider range of turbidity, flow conditions, and seasons.

Evaluation of Surrogate Models Developed for the Animas and San Juan Rivers

The usefulness of surrogate models to predict metal concentrations varies among the USGS stations on the Animas and San Juan Rivers as a result of the processes that control metal concentrations in the rivers and the size and distribution of discrete samples in the calibration datasets. Models did an adequate job of predicting most metals at Cement Creek at Silverton, Colo., the upstream-most site, largely because metal concentrations are elevated at this site and occur primarily in the dissolved phase. Because of the high elevation of this site, runoff is dominated by spring snowmelt, and metals, similar to major dissolved constituents, showed substantial dilution during peak flows. As a result, SC or streamflow provided good surrogates for many of the metals at this site.

Further downstream at the Animas River below Silverton, Colo. (site 2), turbidity became increasingly important for predicting concentrations of some metals because of an increase in the proportion of metals in the particulate phase. Indeed, Schemel and others (2000) showed that the mixing of Cement Creek water with Animas River water increased the pH of the river resulting in formation of colloidal aluminum and iron, which also contained copper and lead. SC remained a good surrogate for manganese and zinc at this site because solubility and sorption of these metals were not as strongly affected by increasing pH as were other metals. Between Silverton and Durango, metal concentrations in the Animas River continued to decline because of both (1) colloid formation and deposition to bed sediments and (2) dilution by surface water and groundwater with neutral pH and low metal content. Although surrogates were relatively poor predictors for most dissolved metals at the Animas River at Durango, Colo. (site 3), turbidity remained a strong predictor of total metals, likely because metal-rich colloids were remobilized from bed sediment and transported mainly during snowmelt runoff. This is consistent with results showing that more than 80 percent of metals in the colloidal component of bed sediments at this site were derived from metal sources upstream from Silverton (Church and others, 1997).

In general, relations between surrogates and metal concentrations became weaker downstream from Durango because inputs of more alkaline water from tributaries and groundwater reduced metal solubility and diluted metal concentrations. For dissolved metals, concentrations continued to decline, with many concentrations near or below the analytical reporting limits. Concentrations close to the reporting limit have greater uncertainty, possibly making it more difficult

Table 11. Summary statistics, regression model equations, and evaluation statistics for dissolved and total metal concentrations for San Juan River near Bluff, Utah (U.S. Geological Survey station 09379500), 2015–17.

[Models with R^2 of less than 0.5 are shown in gray. Model terms listed in order of increasing *p*-value. Min., minimum concentration; µg/L, microgram per liter; Med., median concentration; Max., maximum concentration; N, number of samples in calibration dataset; R^2, coefficient of determination adjusted for serial correlation; Std. error, standard error of estimate; Bias corr., bias correction factor (Duan, 1983); Pred. int., median prediction interval as a percent of the median estimated concentration; *log*, log base 10; *C*, metal concentration; *sin_2*, sin(4π*T*), where *T* is the decimal portion of the year starting on January 1; *cos_2*, cos(4π*T*); *sin*, sin(2π*T*); *cos*, cos(2π*T*); *Temp*, water temperature in degrees Celsius; —, not computed; *Q*, streamflow in cubic feet per second; *Turb*, turbidity in nephelometric turbidity units; *pH*, pH in standard units; *SC*, specific conductance in microsiemens per centimeter at 25 degrees Celsius]

Metal, fraction	Percent censored	Min. (µg/L)	Med. (µg/L)	Max. (µg/L)	N	Model equation and coefficients	R^2	Std. error	Bias corr.	Pred. int.
Aluminum, dissolved	18	4.6	20	73.9	37	$logC = -0.1153 sin\_2 + 0.1223 cos\_2 - 0.0004672 SC + 1.524$	0.52	0.183	1.09	110
Arsenic, dissolved	16	0.6	0.78	1.8	41	$logC = 0.8552 logSC - 0.02921 sin - 0.3031 cos - 0.01385 Temp - 2.261$	0.54	0.0744	1.01	40
Cadmium, dissolved	88	0.031	0.038	0.052	4	—	—	—	—	—
Copper, dissolved	33	0.96	1.6	7.8	33	$logC = -0.0006108 SC - 0.2621 logQ + 1.438$	0.06	—	—	—
Iron, dissolved	24	6.1	24	99.7	31	$logC = 0.00007608 Q + 1.073$	0.45	—	—	—
Lead, dissolved	70	0.021	0.173	0.67	12	—	—	—	—	—
Manganese, dissolved	17	0.4	2.84	8.51	40	$logC = 0.9126 logQ + 0.001058 SC - 3.222$	0.39	—	—	—
Zinc, dissolved	59	3.3	39.8	77.3	13	—	—	—	—	—
Aluminum, total	0	325	5,100	24,700	46	$logC = 0.7552 logTurb + 0.1119 sin + 0.0533 cos + 1.936$	0.79	0.19	1.09	110
Arsenic, total	0	0.9	2.8	10	46	$logC = 0.4539 logTurb - 0.6353 pH + 4.641$	0.75	0.13	1.04	70
Cadmium, total	42	0.06	0.2	1.18	26	$logC = 0.5956 logTurb - 1.579 pH - 0.08696 sin\_2 - 0.1319 cos\_2 + 10.96$	0.71	0.16	1.07	100
Copper, total	2	0.3	11	49	45	$logC = 0.5432 logTurb + 0.1922 sin - 0.01037 cos - 0.8952 pH + 7.115$	0.78	0.18	1.07	110
Iron, total	0	316	6,200	37,000	46	$logC = 0.7258 logTurb - 0.01677 Temp + 0.1111 sin - 0.09056 cos + 2.297$	0.78	0.21	1.10	130
Lead, total	7	0.4	10.9	48	43	$logC = 0.7114 logTurb + 0.2214 sin - 0.1656 cos - 1.087 pH + 8.173$	0.88	0.16	1.07	100
Manganese, total	0	24.7	264	1,400	46	$logC = 0.6447 logTurb + 0.1617 sin - 0.06650 cos - 0.5322 pH + 5.262$	0.84	0.16	1.06	90
Zinc, total	2	2	53	260	44	$logC = 0.6819 logTurb + 0.2158 sin + 0.05732 cos - 0.9515 logSC + 2.561$	0.83	0.20	1.09	120

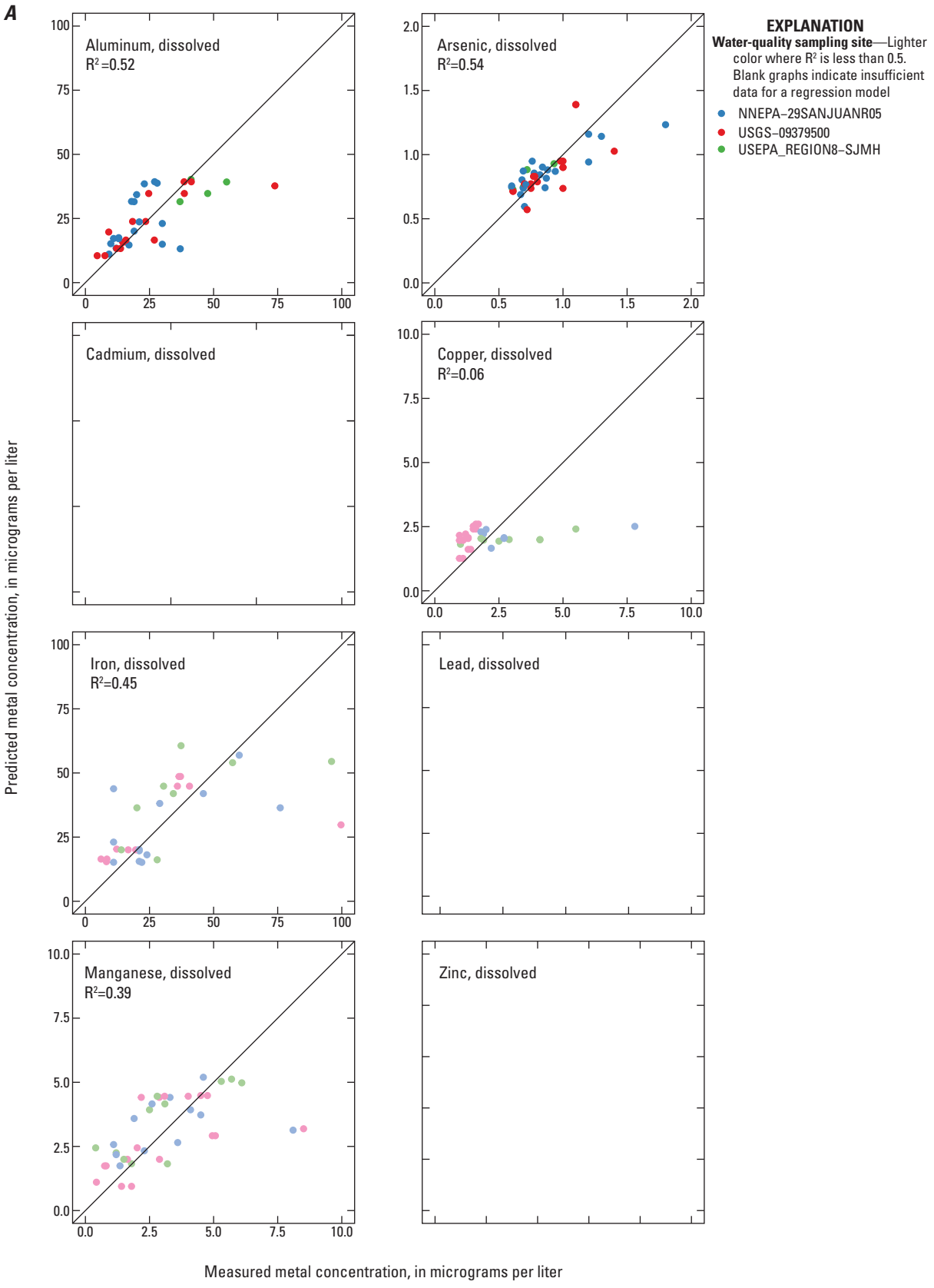

Figure 19. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for San Juan River near Bluff, Utah (U.S. Geological Survey station 09379500), 2015–17. Blank graphs indicate insufficient data for a regression model.

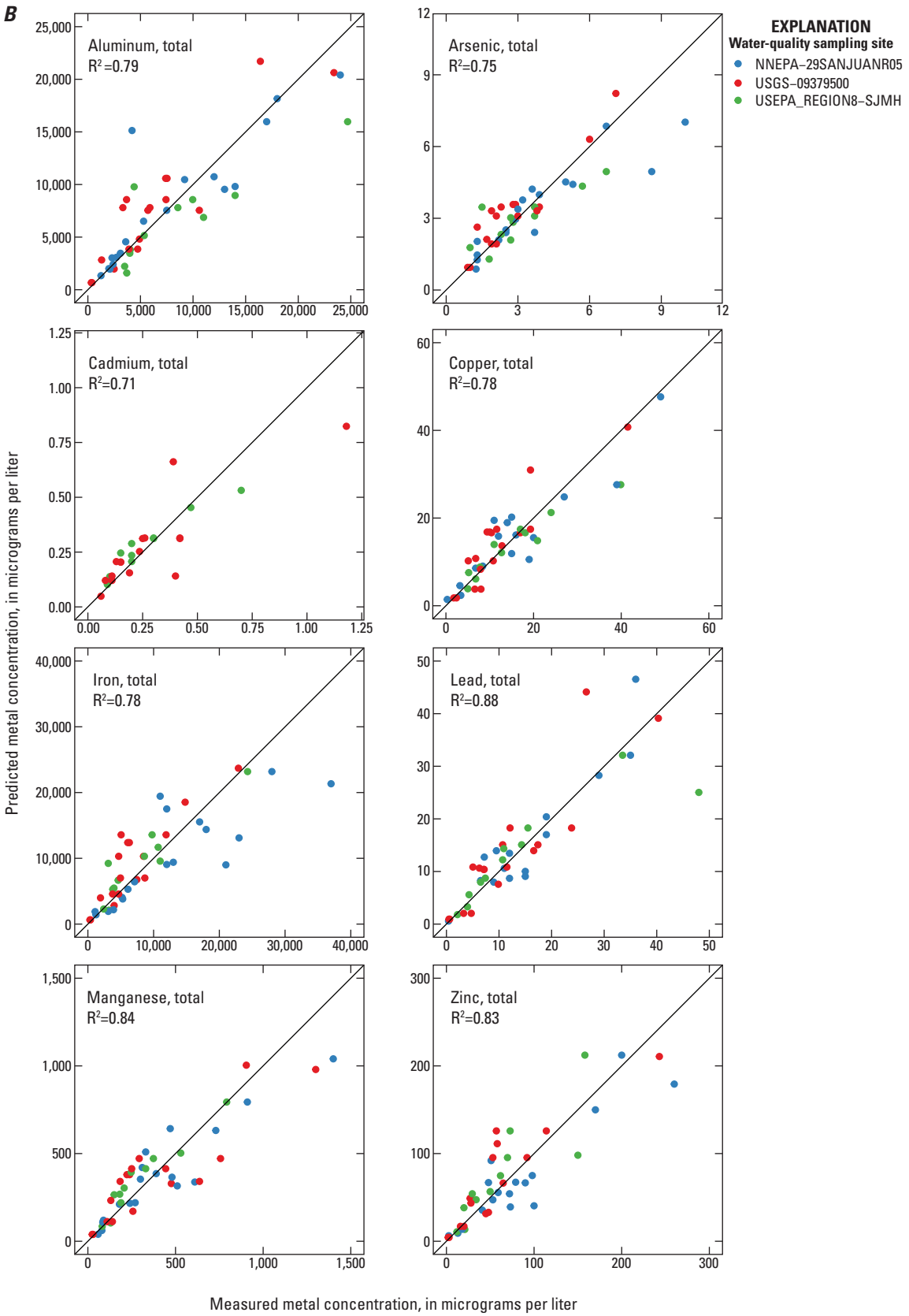

Figure 19. Relation between *A*, predicted and measured dissolved metal concentrations and *B*, predicted and measured total metal concentrations for San Juan River near Bluff, Utah (U.S. Geological Survey station 09379500), 2015–17. Blank graphs indicate insufficient data for a regression model.—Continued

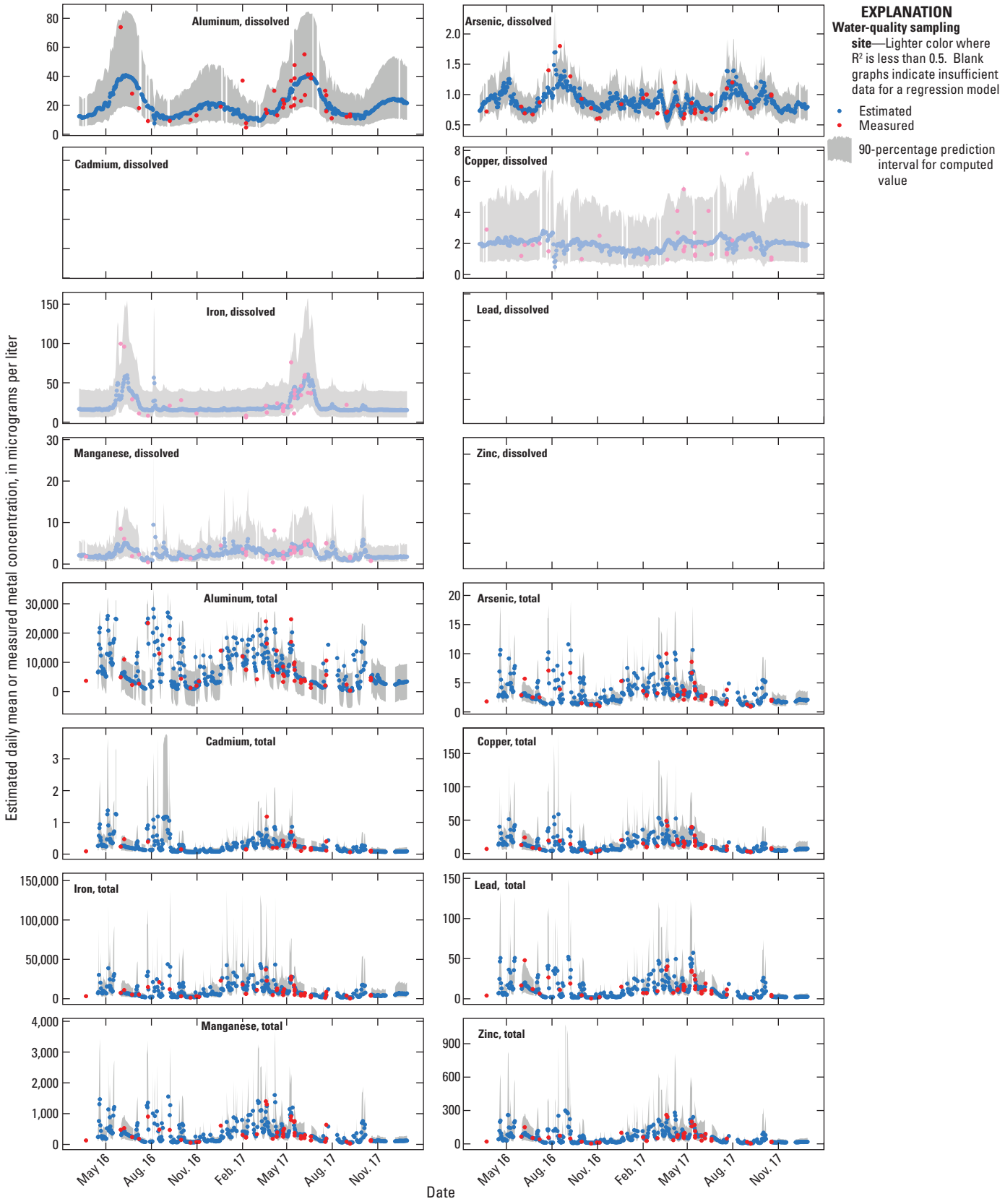

Figure 20. Time series plots of estimated daily mean and measured metal concentrations for San Juan River near Bluff, Utah (U.S. Geological Survey station 09379500), for 2016–17. Gaps in estimated concentrations indicate periods of missing surrogate data. Blank graphs indicate insufficient data for a regression model. Evaluation of Surrogate Models Developed for the Animas and San Juan Rivers.

to achieve good statistical relations. Many of the dissolved metals had so few detected concentrations that model development was not possible at most sites on the San Juan River.

Other factors that might complicate the relation between metal concentrations and surrogates, particularly downstream from Durango, include differences in sources and timing of sediment and water contributions to the river. Downstream from Durango, the bedrock changes to more easily weathered sedimentary rock types that contribute substantial amounts of suspended sediment to the river as indicated by the substantial increase in turbidity (fig. 2). Based on downstream trends in total metal concentrations (tables 3–11), these sedimentary sources appear to contribute aluminum, iron, and arsenic to the river with little to no effect on metals more commonly associated with mining (cadmium, copper, lead, manganese, and zinc) (Wright and others, 2007). Not only is this sediment different in composition from upstream sediment that is affected by metals from mining sources, but inputs primarily occur during short-term runoff events driven by monsoonal moisture as indicated by turbidity spikes occurring during late summer and fall months (fig. 2). The amount and timing of sediment additions during these monsoonal events are highly variable and derived largely from sources outside of the mainstem channel. Another complicating factor in model development for sites on the San Juan River could be timing of flows from different tributaries. Although flows in the Animas River and its tributaries are largely unregulated, flows in the San Juan River upstream from Farmington, N. Mex., are regulated by the Navajo Reservoir and do not always coincide with runoff timing of the Animas River.

In addition to hydrologic and geochemical factors, the success of developing useful surrogate models also depends on having an adequate model-calibration dataset that includes an appropriate number of discrete samples with associated surrogate data and samples collected over a wide range of hydrologic conditions (Rasmussen and others, 2009). The number of samples is often an important metric for determining if a dataset is sufficient, which as suggested by Rasmussen and others (2009), might include as many as monthly samples collected over a 2- to 3-year period. Although several hundred samples with metals data were retrieved for each USGS station, most samples were not accompanied by values for the corresponding surrogate parameters, especially turbidity, which was almost always missing. This reduced the number of available samples for the calibration dataset particularly for total metals because only samples collected after the water-quality sensors began operation in April 2016 could be included in the analysis.

The distribution of the samples over the range of observed streamflow, surrogate values, hydrologic conditions, and seasons for each site also is important for developing adequate statistical models. Duration curves provide a useful tool for evaluating the distribution of samples and are shown in figure 21 comparing results for the Animas River below Silverton, Colo. (site 2), and Animas River near Cedar Hill, N. Mex. (site 4), for streamflow, SC, and turbidity. At the Silverton site, the samples span the range of observed daily streamflow, SC, and turbidity values, which suggests the calibration dataset for this site was suitable for model development. At the Cedar Hill site, the current dataset did not cover the entire range of observed surrogate values. For streamflow and turbidity, samples were missing at the lowest and highest values, whereas for SC, the dataset was not representative of SC at the high end of the range. This suggests that additional samples that target these conditions might yield more representative models. Coordination among sampling agencies might help to improve the distribution of samples over different hydrologic conditions and seasons.

Based on this evaluation, the following recommendations may improve or help to validate the current surrogate models for predicting metal concentrations for sites in this study. Although models were suitable for most dissolved and total metals at the Cement Creek near Silverton and Animas River below Silverton sites, and for total metals at the Animas River at Durango site, the models may be improved by additional sampling during very low flows and along the rising limb of the snowmelt hydrograph, as these are important for metals that exhibit a flushing behavior or are mobilized with metal-rich sediments during the early part of snowmelt runoff. For sites downstream from the Animas River at Durango site, sampling should target the full range of hydrologic conditions including base flow, snowmelt, and summer storm events.

Except for aluminum and iron, concentrations of dissolved metals are so low in the lower reaches of the river that additional sampling will not necessarily improve the regression models and it might be fair to conclude that available surrogates may not provide good proxies for dissolved metals at some sites. However, total metals are higher in concentration and largely driven by the movement of suspended sediment, so the use of turbidity as a surrogate seems plausible. Especially for sites on the San Juan River, storm events are underrepresented in the current dataset and additional samples that target events with high-sediment concentrations would help to validate or refine the current models. Limitations of the optical turbidity sensors, such as limited range, susceptibility to fouling, and sensitivity to suspended particle size, may also be factors in developing reliable models. Consideration should be given to replacing optical sensors with acoustic instruments, which likely would be better suited for the high sediment load of the San Juan River (Landers and others, 2016).

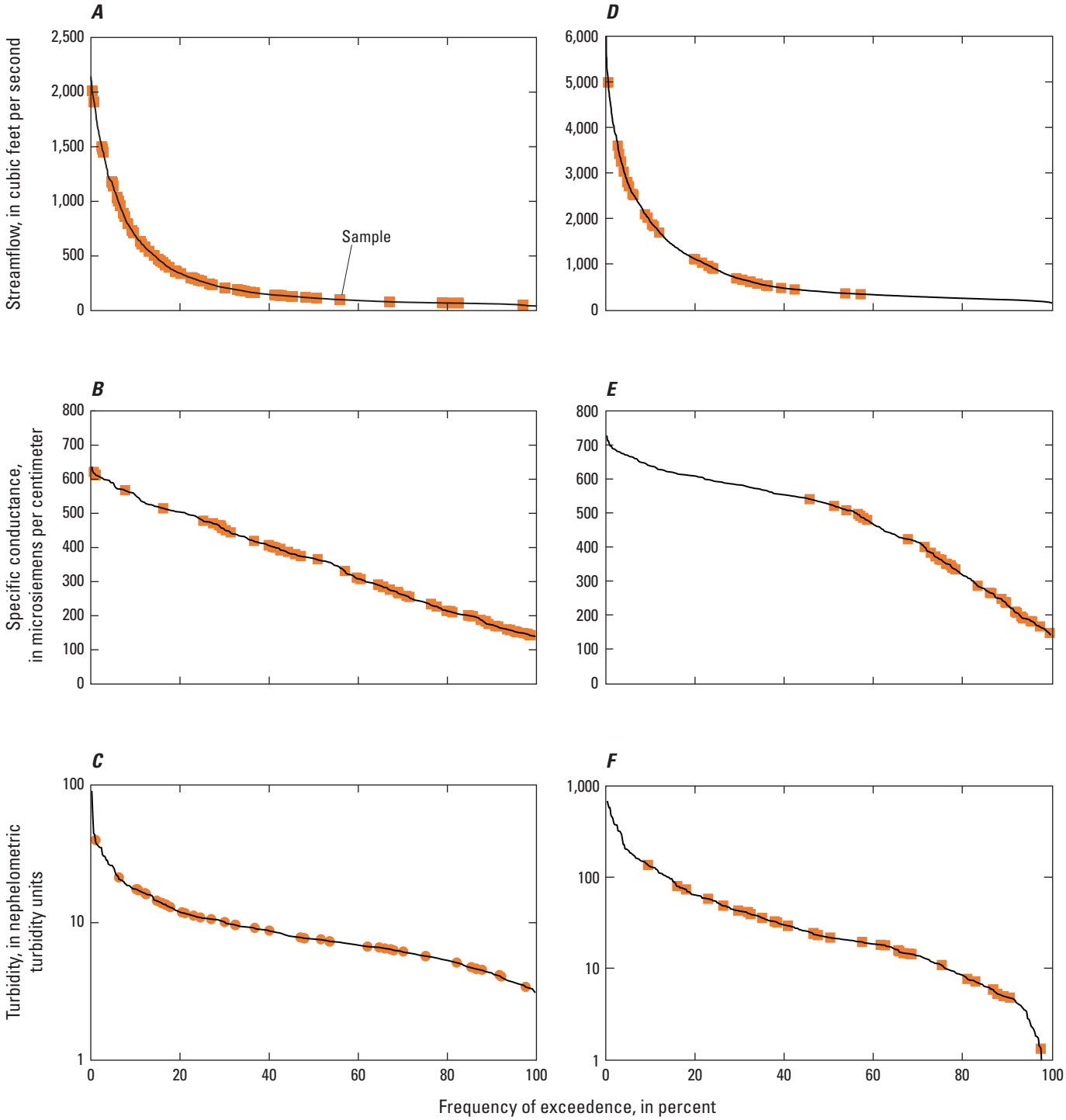

Figure 21. Duration curves for *A*, streamflow; *B*, specific conductance; and *C*, turbidity, and corresponding values associated with discrete water-quality samples collected at the Animas River below Silverton, Colorado, site. The same graphs for the Animas River near Cedar Hill, New Mexico, site are shown in *D*, *E*, and *F*. Duration curves for streamflow are based on daily values for 2008–17 and curves for specific conductance and turbidity are based on daily values for 2016–17.

Summary

Following the Gold King Mine release, there was heightened concern by downstream communities and Tribes in Colorado, New Mexico, and Utah about the risks posed by elevated metal concentrations to public and private drinking-water supplies, water for irrigation and ranching, and recreational fishing and boating. As a result, monitoring of basic water-quality parameters and metal concentrations in the Animas and San Juan Rivers was identified as a priority for characterizing and understanding the water quality of the river. In response to this need, the U.S. Geological Survey, in collaboration with the U.S. Environmental Protection Agency, agreed on a scope of work to install and operate continuous water-quality sensors that would be used in an attempt to develop regression models that correlate one or more parameters collected by the sensors (referred to as "surrogates" and including specific conductance, pH, turbidity, and water temperature) with dissolved and total concentrations of key metals of concern (aluminum, arsenic, cadmium, copper, iron, lead, manganese, and zinc). The purpose of this report is to evaluate the use of site-specific regression models to compute concentrations of selected metals using statistical relations between continuous and discrete water-quality monitoring data collected at nine U.S. Geological Survey streamflow-gaging stations along the Animas and San Juan Rivers.

The usefulness of surrogate models to predict metal concentrations had mixed results for the Animas and San Juan Rivers, which reflects processes that determine metal concentrations in rivers as well as the size and distribution of discrete samples in the calibration datasets. A larger number of regression models were statistically significant for the most upstream sites, where metal concentrations were elevated by drainage from abandoned mines and mineralized bedrock. Models generally did not perform as well at downstream sites, especially for dissolved metals, which occurred at lower concentrations than at the upstream sites. In the lower reaches of the rivers, the input of more alkaline water from tributaries and groundwater reduced metal solubility and diluted metal concentrations.

In general, the models for total metals performed better than the models for dissolved metals. Specific conductance was the most common explanatory variable in dissolved metal models and turbidity was the most common explanatory variable in total metal models.

This report includes some recommendations that might improve, or help to validate, the current surrogate models for predicting metal concentrations for sites in this study. Sampling should target the full range of hydrologic conditions including base flow, snowmelt, and summer storm events. Consideration should be given to replacing optical sensors with acoustic instruments, which likely would be better suited for the high sediment load of the San Juan River.

Acknowledgments

The author thanks Johanna Blake (U.S. Geological Survey) for feedback on development of regression models for the San Juan River sites in New Mexico. Technical reviews provided by Brent Aulenbach (U.S. Geological Survey) and Cory Williams (U.S. Geological Survey) greatly improved the content and clarity of the report.

References Cited

Aulenbach, B.T., 2013, Improving regression-model-based streamwater constituent load estimates derived from serially correlated data: Journal of Hydrology, v. 503, p. 55–66.

Aulenbach, B.T., Joiner, J.K., and Painter, J.A., 2017, Hydrology and water-quality in 13 watersheds in Gwinnett County, Georgia, 2001–15: U.S. Geological Survey Scientific Investigations Report 2017–5012, 82 p.

Church, S.E., Kimball, B.A., Fey, D.L., Ferderer, D.A., Yager, T.J., and Vaughn, R.B., 1997, Source, transport, and partitioning of metals between water, colloids, and bed sediments of the Animas River, Colorado: U.S. Geological Survey Open-File Report 97–151, 135 p., accessed April 1, 2018, at https://pubs.er.usgs.gov/publication/ofr97151.

Church, S.E., von Guerard, P., and Finger, S.E., eds., 2007, Integrated investigations of environmental effects of historical mining in the Animas River watershed, San Juan County, Colorado: U.S. Geological Survey Professional Paper 1651, 1,096 p.

Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, no. 383, p. 605–610.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 522 p.

Landers, M.N., Straub, T.D., Wood, M.S., and Domanski, M.M., 2016, Sediment acoustic index method for computing continuous suspended-sediment concentrations: U.S. Geological Survey Techniques and Methods, book 3, chap. C5, 63 p., accessed April 1, 2018, at https://doi.org/10.3133/tm3C5.

Lorenz, D.L., 2015, smwrBase—An R package for managing hydrologic data, version 1.1.1: U.S. Geological Survey Open-File Report 2015–1202, 7 p., accessed April 1, 2018, at https://doi.org/10.3133/ofr20151202.

Mast, M.A., 2018, Calibration datasets and model archive summaries for regression models developed to estimate metal concentrations at nine sites on the Animas and San Juan Rivers, Colorado, New Mexico, and Utah: U.S. Geological Survey data release, https://doi.org/10.5066/P9THSFE0.

Mazerolle, M.J., 2017, AICcmodavg—Model selection and multimodel inference based on (Q)AIC(c), R package version 2.1-1, accessed April 1, 2018, at https://cran.r-project.org/package=AICcmodavg.

Miller, L.D., Schaffrath, K.R., and Linard, J.I., 2013, Assessment of historical surface-water quality data in southwestern Colorado, 1990–2005: U.S. Geological Survey Scientific Investigations Report 2012–5255, 74 p.

Nania, J., Cozzetto, K., and Duren, S., 2014, Farming, chap. 5 *of* Nania, J., and Cozzetto, K., eds., Considerations for climate change and variability adaptation on the Navajo Nation: Boulder, Colo., University of Colorado Boulder, p. 81–94, accessed April 1, 2018, at http://scholar.law.colorado.edu/books_reports_studies/3/.

Nasrabadi, T., Ruegner. H., Sirdari, Z.Z., Schwientek, M., and Grathwohl, P., 2016, Using total suspended solids (TSS) and turbidity as proxies for evaluation of metal transport in river water: Applied Geochemistry, v. 68, p. 1–9.

Rantz, S.E. and others, 1982, Measurement and computation of streamflow, Volume 1—Measurement of stage and discharge and Volume 2—Computation of discharge: U.S. Geological Survey Water-Supply Paper 2175, 631 p.

Rasmussen, P.P., Gray, J.R., Glysson, G.D., and Ziegler, A.C., 2009, Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods, book 3, chap. C4, 53 p.

Schemel, L.E., Kimball, B.A., and Bencala, K.E., 2000, Colloid formation and metal transport through two mixing zones affected by acid mine drainage near Silverton, Colorado: Applied Geochemistry, v. 15, p. 1003–1018.

Schilling, K.E., Kim, S.W, and Jones, C.S., 2017, Use of water quality surrogates to estimate total phosphorus concentrations in Iowa rivers: Journal of Hydrology—Regional Studies, v. 12, p. 111–121.

Senior, L.A., 2017, Estimated fecal coliform bacteria concentrations using near real-time continuous water-quality and streamflow data from five stream sites in Chester County, Pennsylvania, 2007–16: U.S. Geological Survey Scientific Investigations Report 2017–5075, 46 p.

Stolwijk, A.M., Straatman, H., and Zielhuis, G.A., 1999, Studying seasonality by using sine and cosine functions in regression analysis: Journal of Epidemiology & Community Health, v. 53, p. 235–238.

Sullivan, K., Cyterski, M., Knightes, C., Kraemer, S.R., Washington, J., Prieto, L, and Avant, B., 2017, Analysis of the transport and fate of metals released from the Gold King Mine in the Animas and San Juan Rivers: U.S. Environmental Protection Agency, Office of Research and Development, EPA/600/R-16/296, 328 p, accessed April 1, 2018, at https://www.epa.gov/goldkingmine/fate-transport-analysis.

U.S. Geological Survey, 2018a, USGS 09358000 Animas River at Silverton, CO *in* USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed August 1, 2018, at http://waterdata.usgs.gov/nwis/inventory/?site_no=09358000&agency_cd=USGS.

U.S. Geological Survey, 2018b, USGS 09361500 Animas River at Durango, CO *in* USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed August 1, 2018, at http://waterdata.usgs.gov/nwis/inventory/?site_no=09361500&agency_cd=USGS.

U.S. Geological Survey, 2018c, USGS 09364010 Animas River below Aztec, NM *in* USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed August 1, 2018, at http://waterdata.usgs.gov/nwis/inventory/?site_no=09364010&agency_cd=USGS.

U.S. Geological Survey, 2018d, USGS 09365000 San Juan River at Farmington, NM *in* USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed August 1, 2018, at http://waterdata.usgs.gov/nwis/inventory/?site_no=09365000&agency_cd=USGS.

U.S. Geological Survey, 2018e, USGS 09379500 San Juan River near Bluff, UT *in* USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed August 1, 2018, at http://waterdata.usgs.gov/nwis/inventory/?site_no=09379500&agency_cd=USGS.

Wagner, R.J., Boulger, R.W., Jr., Oblinger, C.J., and Smith, B.A., 2006, Guidelines and standard procedures for continuous water-quality monitors—Station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1–D3, 51 p. plus attachments.

Welch, H.L., Coupe, R.H., and Aulenbach, B.T., 2014, Concentrations and transport of suspended sediment, nutrients, and pesticides in the lower Mississippi-Atchafalaya River subbasin during the 2011 Mississippi River flood, April through July: U.S. Geological Survey Scientific Investigations Report 2014–5100, 44 p.

Wright, W.G., Simon, W., Bove, D.J., Mast, M.A., and Leib, K.J., 2007, Distribution of pH values and dissolved trace-metal concentrations in streams, chap. E10 *of* Church, S.E., von Guerard, P., and Finger, S.E., eds., Integrated investigations of environmental effects of historical mining in the Animas River watershed, San Juan County, Colorado: U.S. Geological Survey Professional Paper 1651, p. 497–541.

Appendix 1. Locations of U.S. Geological Survey Streamflow-Gaging Stations and Associated Water-Quality Sampling Sites used in the Regression Analysis

The appendix provides maps (figs 1.1–1.9) showing the locations of U.S. Geological Survey (USGS) streamflow-gaging stations and associated water-quality sampling sites in Colorado, New Mexico, and Utah used in the regression analysis.

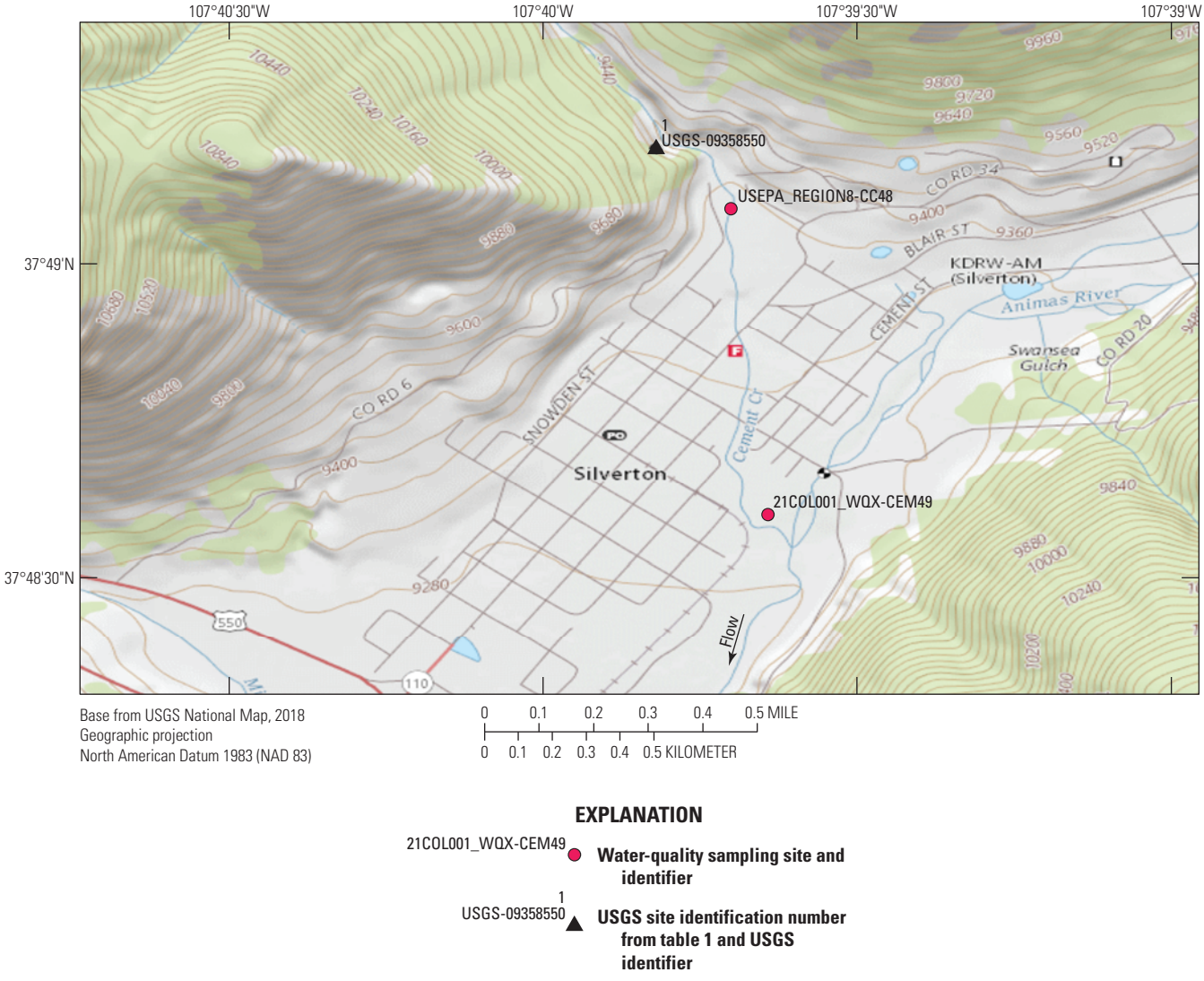

Figure 1.1. Cement Creek at Silverton, Colorado (station 09358550) and associated water-quality sampling sites (see table 1 for a description of the sites). (USGS, U.S. Geological Survey)

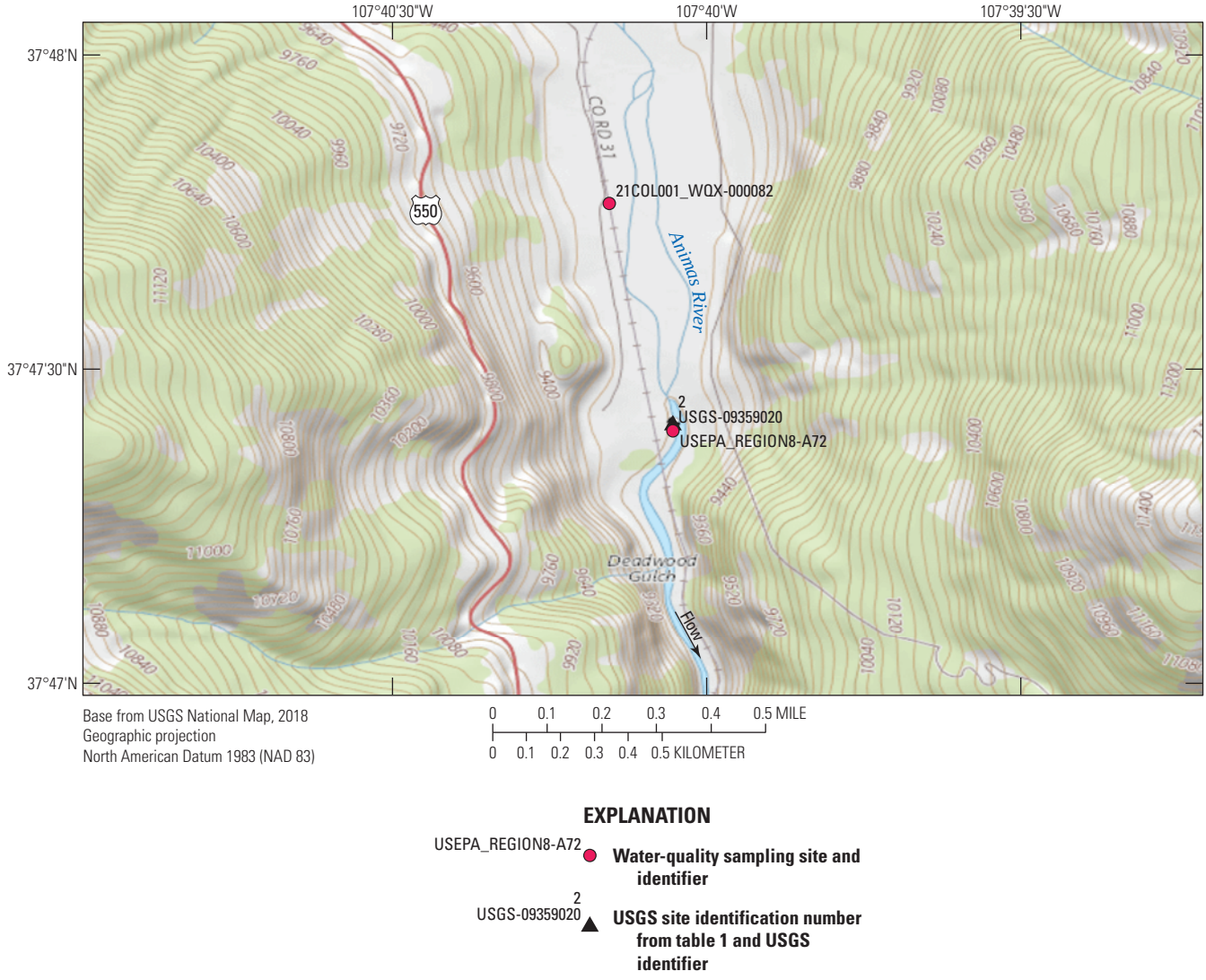

Figure 1.2. Animas River below Silverton, Colorado (station 09359020) and associated water-quality sampling sites (see table 1 for a description of the sites). (USGS, U.S. Geological Survey)

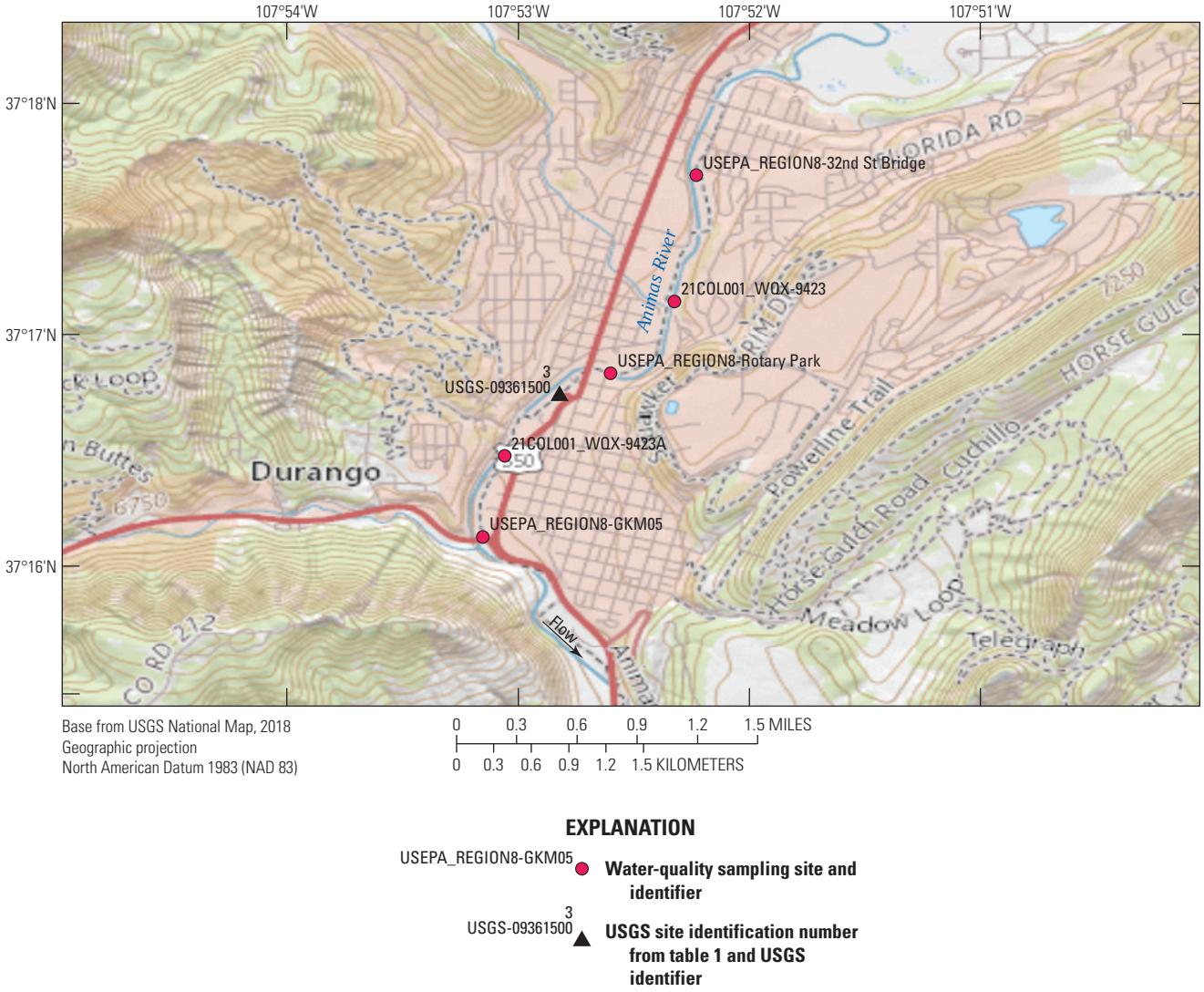

Figure 1.3. Animas River at Durango, Colorado (station 09361500) and associated water-quality sampling sites (see table 1 for a description of the sites). (USGS, U.S. Geological Survey)

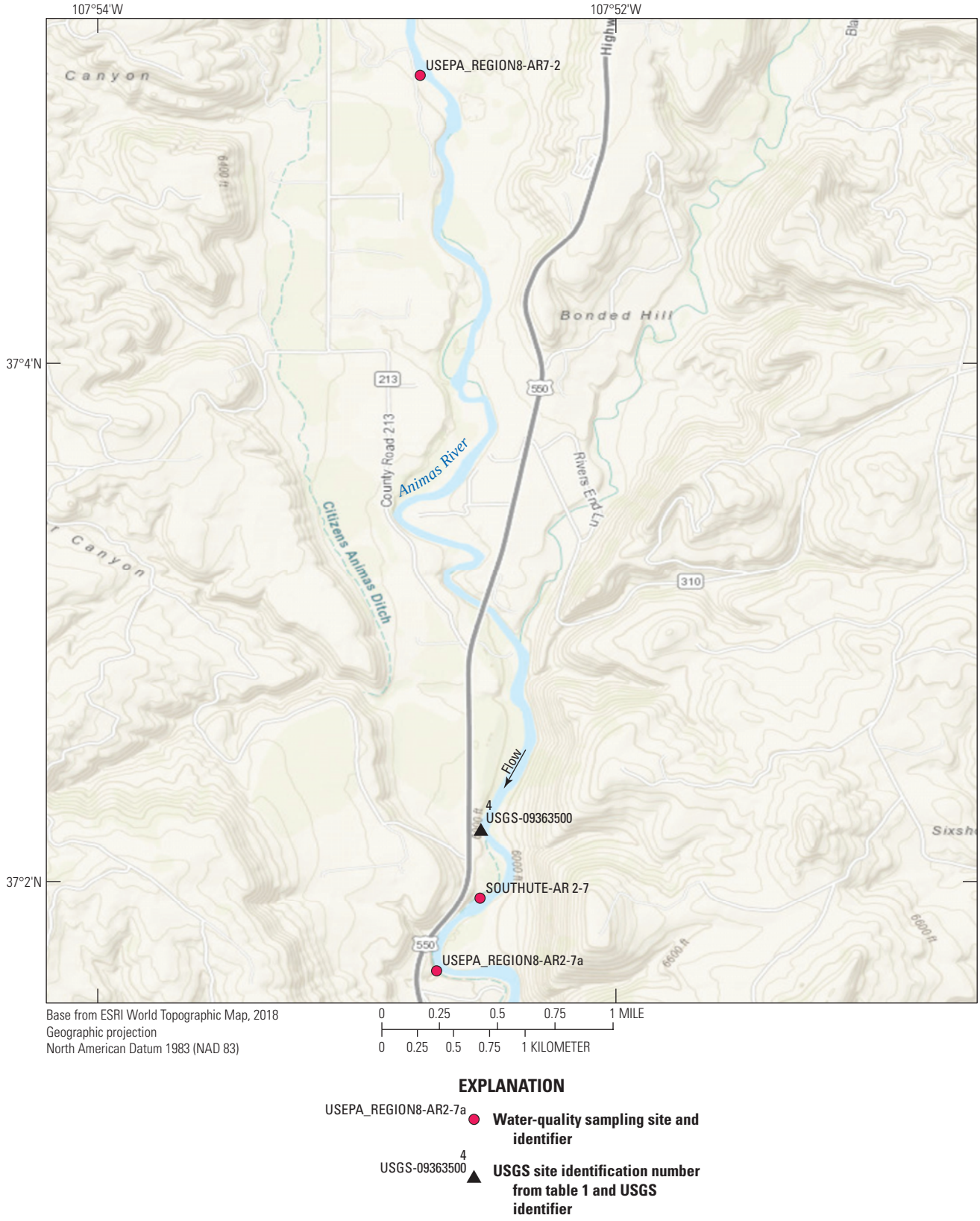

Figure 1.4. Animas River near Cedar Hill, New Mexico (station 09363500) and associated water-quality sampling sites (see table 1 for a description of the sites). (USGS, U.S. Geological Survey)

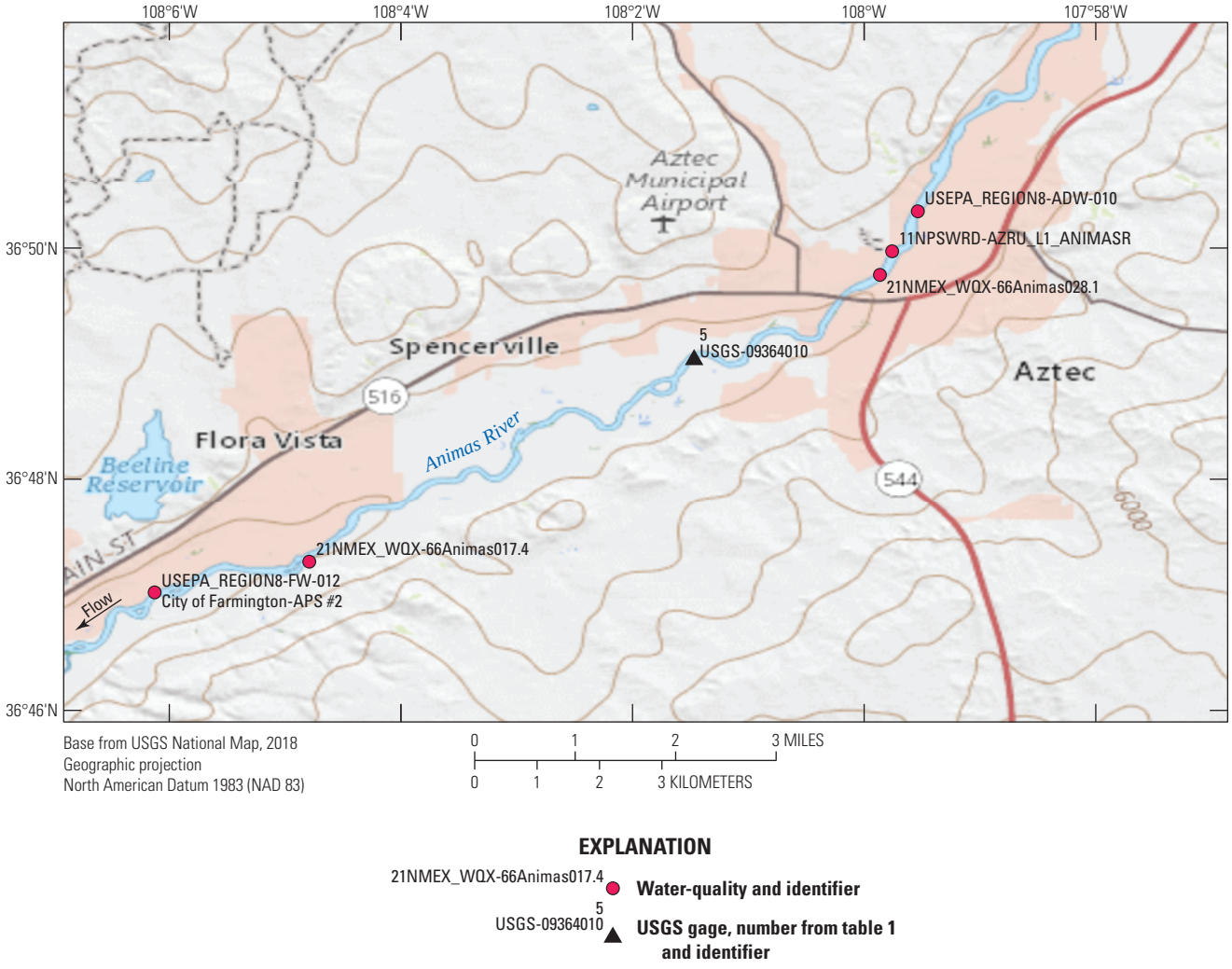

Figure 1.5. Animas River below Aztec, New Mexico (station 09364010) and associated water-quality sampling sites (see table 1 for a description of the sites). (USGS, U.S. Geological Survey)

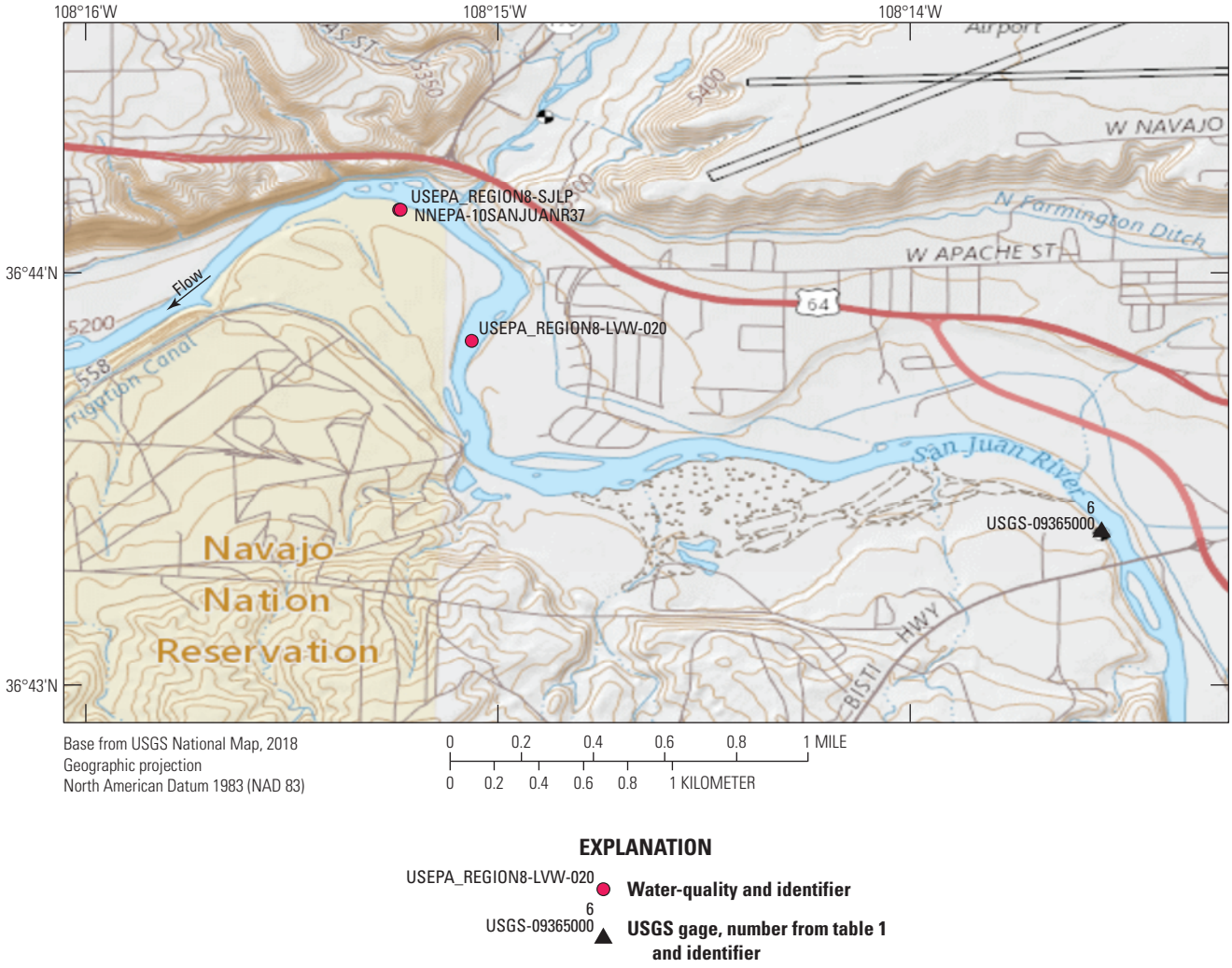

Figure 1.6. San Juan River at Farmington, New Mexico (station 09365000) and associated water-quality sampling sites (see table 1 for a description of the sites). (USGS, U.S. Geological Survey)

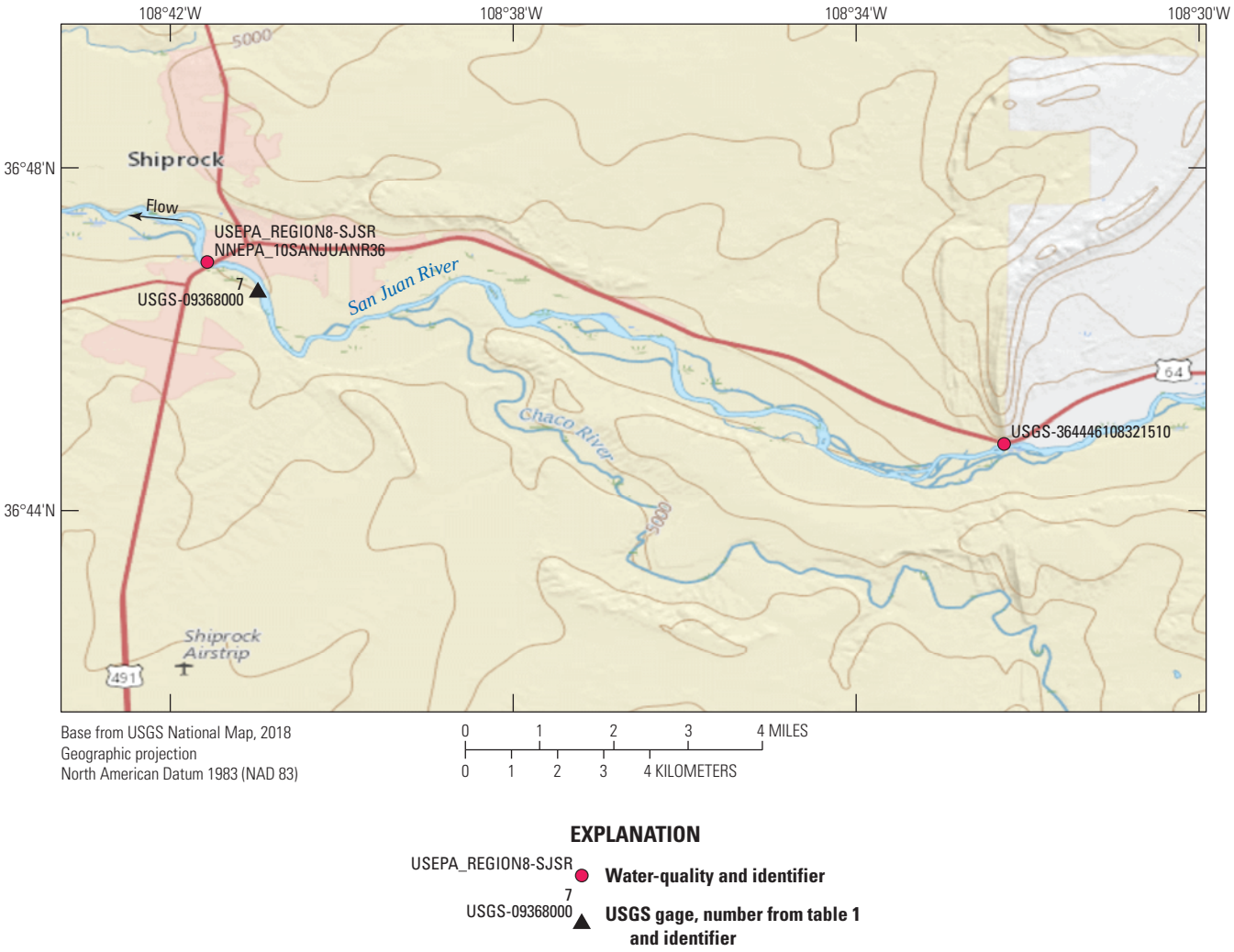

Figure 1.7. San Juan River at Shiprock, New Mexico (station 09368000) and associated water-quality sampling sites (see table 1 for a description of the sites). (USGS, U.S. Geological Survey)

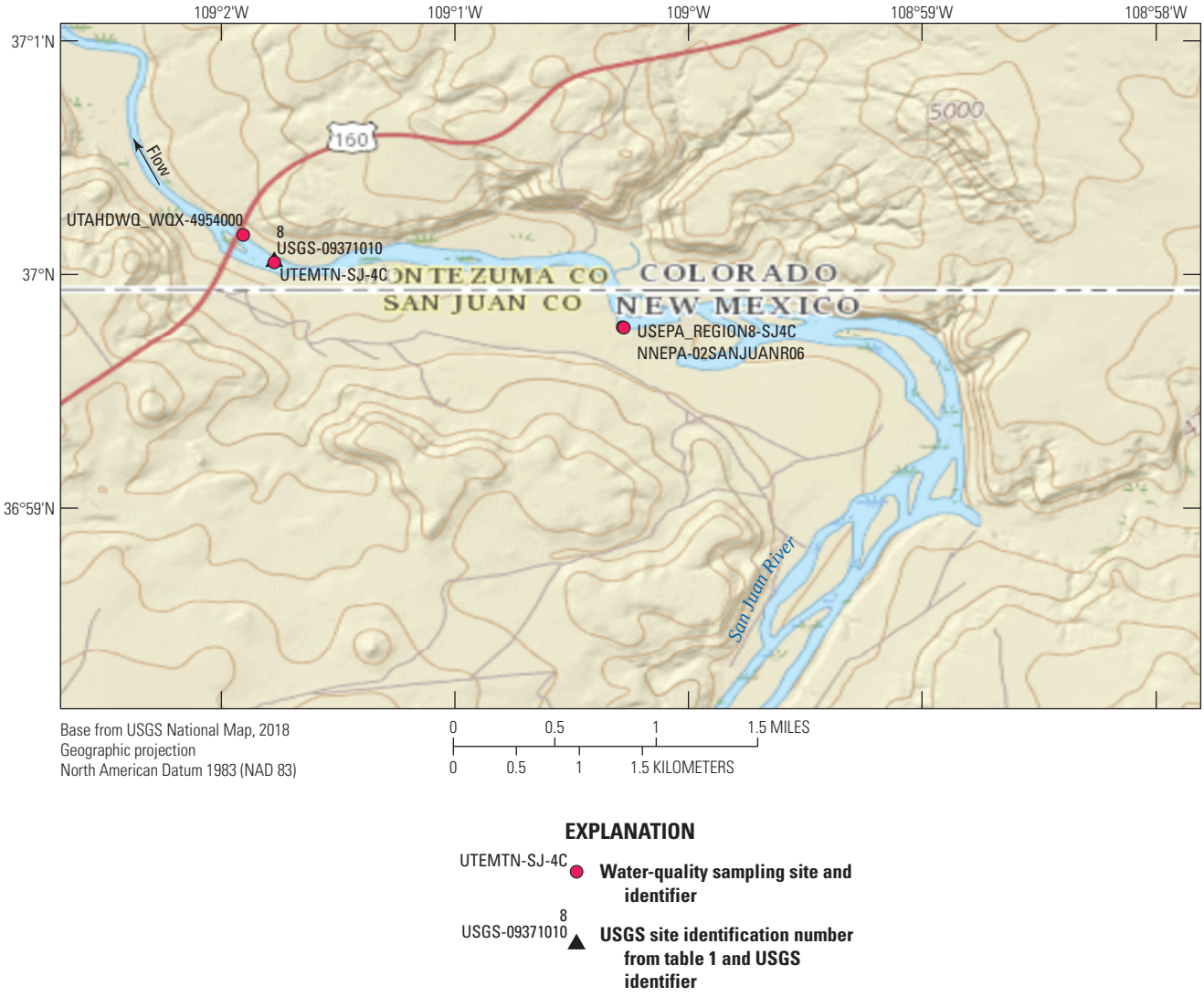

Figure 1.8. San Juan River at Four Corners, Colorado (station 09371010) and associated water-quality sampling sites (see table 1 for a description of the sites). (USGS, U.S. Geological Survey)

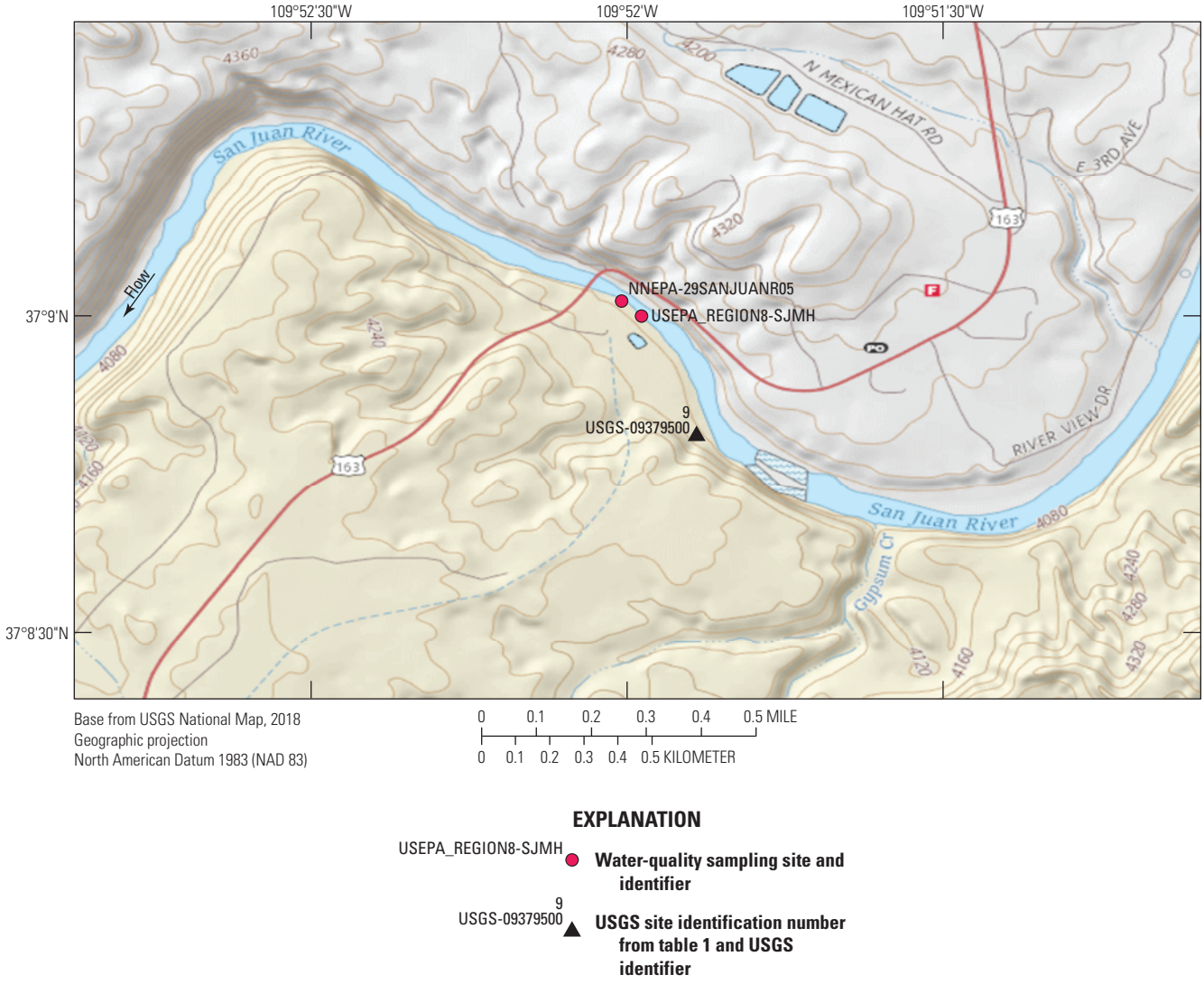

Figure 1.9. San Juan River near Bluff, Utah (station 09379500) and associated water-quality sampling sites (see table 1 for a description of the sites). (USGS, U.S. Geological Survey)

For additional information, contact:
Director, Colorado Water Science Center
U.S. Geological Survey
Box 25046, Mail Stop 415
Denver, CO 80225
(303) 236–6901

Or visit our website at:
https://co.water.usgs.gov/

Publishing support provided by the Denver and West Trenton Publishing Service Centers

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20185116