Regression Models for Estimating Salinity and Selenium Concentrations at Selected Sites in the Upper Colorado River Basin, Colorado, 2009–2012

Open-File Report 2014–1015

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

Prepared in cooperation with the Bureau of Reclamation, the Colorado River Water Conservation District, and the Bureau of Land Management

Regression Models for Estimating Salinity and Selenium Concentrations at Selected Sites in the Upper Colorado River Basin, Colorado, 2009–2012

By Joshua I. Linard and Keelin R. Schaffrath

Open-File Report 2014–1015

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

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia 2014

For product and ordering information:
World Wide Web: *http://www.usgs.gov/pubprod*
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth,
its natural and living resources, natural hazards, and the environment:
World Wide Web: *http://www.usgs.gov*
Telephone: 1-888-ASK-USGS

Suggested citation:
Linard, J.I., and Schaffrath, K.R., 2014, Regression models for estimating salinity and selenium concentrations at selected sites in the Upper Colorado River Basin, Colorado, 2009–2012: U.S. Geological Survey Open-File Report 2014–1015, 28 p., *http://dx.doi.org/10.3133/ofr20141015*.

ISSN 2331-1258 (online)

Contents

Abstract 1
Introduction 1
  Purpose and Scope 3
  Study-Area Description 3
  Study-Site Selection and Data Compilation 5
Model Development 8
Regression Models To Estimate Salinity and Selenium Concentration 9
  Salinity Models 11
  Selenium Models 16
Use of Continuous Data in the Application of Salt and Selenium Regression Models 23
Summary 26
References Cited 27

Figures

Figure 1. The study area and study sites within the Upper Colorado River Basin, Colorado, also showing irrigated land and the Mancos Shale 4
Figure 2. Comparison of measured and estimated salinity concentrations at six sites using either specific conductance or streamflow as an explanatory variable. *A*, SCanyon. *B*, PiceanceRyan. *C*, PiceanceWhite. *D*, Yellow. *E*, DoloresAtBedrock. *F*, DoloresNrBedrock 11
Figure 3. Comparison of measured and estimated selenium concentrations at six sites using specific conductance, streamflow, and (or) seasonality explanatory variables. *A*, Stateline. *B*, GunnWhitewater. *C*, GunnDelta. *D*, NForkGunn. *E*, UncDelta. and *F*, UncColona 17
Figure 4. Example of the use of continuous data with the regression models to estimate *A*, salinity concentration at PiceanceWhite and *B*, selenium concentration at GunnWhitewater using continuously measured specific conductance 24

Tables

Table 1. Summary statistics of salinity and selenium data used to develop regression models for selected study sites within the Upper Colorado River Basin, Colorado 6
Table 2. Model coefficients and evaluation statistics for salinity and selenium regression models developed for selected study sites within the Upper Colorado River Basin, Colorado 10

Conversion Factors

Inch/Pound to SI

Multiply	By	To obtain
	Area	
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Flow Rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

Horizontal coordinate information is referenced to North American Datum of 1983 (NAD 83).
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 either in milligrams per liter (mg/L) or micrograms per liter (µg/L).
Water year, as used in this report, refers to the period that begins on October 1 of the previous year and ends the following September 30 and is designated by the year in which it ends.

Regression Models for Estimating Salinity and Selenium Concentrations at Selected Sites in the Upper Colorado River Basin, Colorado, 2009–2012

By Joshua I. Linard and Keelin R. Schaffrath

Abstract

Elevated concentrations of salinity and selenium in the tributaries and main-stem reaches of the Colorado River are a water-quality concern and have been the focus of remediation efforts for many years. Land-management practices with the objective of limiting the amount of salt and selenium that reaches the stream have focused on improving the methods by which irrigation water is conveyed and distributed. Federal land managers implement improvements in accordance with the Colorado River Basin Salinity Control Act of 1974, which directs Federal land managers to enhance and protect the quality of water available in the Colorado River. In an effort to assist in evaluating and mitigating the detrimental effects of salinity and selenium, the U.S. Geological Survey, in cooperation with the Bureau of Reclamation, the Colorado River Water Conservation District, and the Bureau of Land Management, analyzed salinity and selenium data collected at sites to develop regression models. The study area and sites are on the Colorado River or in one of three small basins in Western Colorado: the White River Basin, the Lower Gunnison River Basin, and the Dolores River Basin. By using data collected from water years 2009 through 2011, regression models able to estimate concentrations were developed for salinity at six sites and selenium at six sites. At a minimum, data from discrete measurement of salinity or selenium concentration, streamflow, and specific conductance at each of the sites were needed for model development. Comparison of the *Adjusted* R^2 and standard error statistics of the two salinity models developed at each site indicated the models using specific conductance as the explanatory variable performed better than those using streamflow. The addition of multiple explanatory variables improved the ability to estimate selenium concentration at several sites compared with use of solely streamflow or specific conductance. The error associated with the log-transformed salinity and selenium estimates is consistent in log space; however, when the estimates are transformed into non-log values, the error increases as the estimates decrease. Continuous streamflow and specific conductance data collected at study sites provide the means to examine temporal variability in constituent concentration and load. The regression models can estimate continuous concentrations or loads on the basis of continuous specific conductance or streamflow data. Similar estimates are available for other sites at the USGS National Real-Time Water Quality Web page (*http://nrtwq.usgs.gov*) and provide water-resource managers with a means of improving their general understanding of how constituent concentration or load can change annually, seasonally, or in real time.

Introduction

Elevated concentrations of salinity and selenium in the tributaries and main-stem reaches of the Colorado River are a water-quality concern and have been the focus of remediation efforts for many years. "Salinity" refers to mineral salts or solids dissolved in water, and selenium is a trace element that bioaccumulates in the food chain. Salt and selenium limit municipal uses of water, reduce agricultural productivity, and, in the case of selenium, can lead to mortality, abnormalities, and reproductive failure in waterfowl and fish (Butler and others, 1996; Presser and Luoma, 2006; Tuttle and Grauch, 2009; Leib and others, 2012). In 1985, the National Irrigation Water Quality Program, a multiagency program within the Department of the Interior, began investigating the effects of irrigation drainage on water quality and on fish and wildlife in the Western United States (Bureau of Reclamation, 2001). The investigations led to the discovery that irrigation drainage contributes a significant part of the nonpoint-source salinity and selenium to the Upper Colorado River Basin and the discovery that high concentrations of selenium are present in water, biota, and sediment (Wright and Butler, 1993; Butler, 1996; Butler and others, 1991, 1996; Butler and Leib, 2002).

Selenium is paradoxical in that it is a nutritional requirement in small amounts but toxic in slightly greater amounts (Lemly, 1993). Various tributaries to the Gunnison and Colorado Rivers are on the State of Colorado's 303(d) list as impaired for selenium (Colorado Department of Public Health and Environment, 2012). Selenium cycles through the aquatic environment and can quickly reach toxic levels as a result of bioconcentration in aquatic organisms (Lemly, 2002). Toxic levels of selenium can cause reproductive failure, deformities, and other adverse effects in birds and fish, including some threatened and endangered fish species (Ohlendorf and others, 1986; Presser and Ohlendorf, 1987; Hamilton, 1998; Lemly, 2002). The Colorado River and portions of the river's tributaries are designated critical habitat for four fish species listed under the Endangered Species Act: the Colorado pikeminnow (*Ptychocheilus lucius*), razorback sucker (*Xyrauchen texanus*), bonytail (*Gila elegans*), and humpback chub (*Gila cypha*).

The geology, land cover, land use, and precipitation characteristics in the basin control streamflow and water chemistry (Kenney and others, 2009). Natural, nonpoint sources of salinity and selenium generally originate from the weathering and dissolution of geological formations that have high salt and selenium content (Presser and Ohlendorf, 1987; Prairie and others, 2005). Salts and trace elements are mobilized through dissolution, surface runoff, and percolation into the groundwater system, which discharges to the river system as base flow (Warner and others, 1985). The application of irrigation water to agricultural lands increases the percolation rate and, consequently, the dissolution and transport of salinity and selenium to streams (Prairie and others, 2005; Kenney and others, 2009).

Land-management practices aimed at limiting the amount of salt and selenium that reaches the stream have focused on improving the methods by which irrigation water is conveyed and distributed. Federal land managers implement these practices in accordance with the Colorado River Basin Salinity Control Act of 1974, which directs Federal land managers to enhance and protect the quality of water available in the Colorado River. To provide the Bureau of Reclamation with information that assists in evaluating and planning salinity-control needs, Liebermann and others (1989) developed multiple linear regression models to estimate annual and monthly salinity concentrations. Streamflow and specific conductance (depending on data availability) were the explanatory variables used in the multiple linear regression models. Streamflow had a negative relation to the concentration of dissolved solids, termed "salinity" in this report, and specific conductance had a positive relation to salinity, which was consistent with patterns described by Hem (1985). To assist the Bureau of Reclamation with understanding salinity, the salinity concentration estimates were converted to loads, which are defined as the weight of material flowing past the sampling site during a specific time interval. Since the publication of the study by Liebermann and others (1989), the U.S. Geological Survey has applied the models on a biennial basis to estimate salinity concentration and load at 20 of the sites analyzed by Liebermann and others (1989). The most recent updates to the models and their estimates were made publically available in a report produced by the Bureau of Reclamation (2011b, *http://www.usbr.gov/uc/progact/salinity/pdfs/PR23.pdf*). A more current study by Mayo and Leib (2012) produced multiple linear regression models capable of estimating selenium concentrations at two sites in the Gunnison and Colorado Rivers. These models used daily streamflow and time data, capturing a seasonal cycling of selenium, as their explanatory variables.

Increased interest in estimates of water quality at temporal resolutions greater than annual or monthly has led to the use of continuously monitored data in the application of multiple linear regression models. Specifically, the U.S. Geological Survey National Real-Time Water Quality Web page (*http://nrtwq.usgs.gov*) computes water quality at numerous sites across the United States from real-time, continuously measured data. Currently, there is an effort to use continuous data measured in the Upper Colorado River Basin in the application of the regression models developed at 20 of the sites analyzed by Liebermann and others (1989).

To enhance the understanding of salinity and selenium in the Upper Colorado River Basin, the network of sites measuring those types of data was expanded. Discrete and continuous data were collected at these additional sites. In an effort to assist in evaluating and mitigating the detrimental effects of salinity and selenium in the Upper Colorado River Basin, the U.S. Geological Survey, in cooperation with the Bureau of Reclamation, the Colorado River Water Conservation District, and the Bureau of Land Management, developed regression models from the discrete salinity and selenium data collected at the new sites.

Purpose and Scope

This report documents regression models developed to assist land managers with evaluating and planning salinity and selenium control needs in the Upper Colorado River Basin. A description of the discrete data collected at 11 sites in the Upper Colorado River Basin from water years 2009 to 2011 is provided. Using those data, regression models for estimating concentrations were developed for salinity at six sites and selenium at six sites. Data from October through April of water year 2012[1] were used to verify the models. The models provide a general understanding of the temporal dynamics of constituent concentration or load on a seasonal or annual basis when applied to continuous data. Moreover, concentration and loads estimated at study sites in real time from continuous data, similar to that presented by the USGS (*http://nrtwq.usgs.gov*), provide water-resource managers with easy access to data for the system they manage.

Study-Area Description

The study area and sites are on the Colorado River or in one of three small basins in Western Colorado (fig. 1). From north to south, the basins are the White River Basin, the Lower Gunnison River Basin, and the Dolores River Basin. The majority of precipitation in the lower elevations, which are classified as a semiarid climate, falls as rain during the late summer and early fall (July through September). The headwaters of all of the basins originate in higher elevations that are considered subalpine zones. The majority of precipitation in the higher elevations falls as snow and accumulates in a seasonal snowpack. The spring snowmelt of the seasonal snowpack dominates the annual streamflow cycle. Streamflow begins to increase in March or April, peaks between May and June, and decreases in July and August. In the lower elevations, smaller increases in streamflow occur in July and August because of summer thunderstorms. Exceptions to the snowmelt hydrograph result from irrigation season return flows and controlled releases from reservoirs in the study area (Butler and others, 1991). Based on the National Land Cover Database in 2001 (Homer and others, 2004), the land cover in Western Colorado is generally classified as shrubs or forest. All of the basins included in the study area have areas of irrigated agriculture. In the entire study area, there were more than 3,650 square kilometers (1,409 square miles) of irrigated parcels in 2005 (Colorado Water Conservation Board, 2010). The Mancos Shale underlies the irrigated parcels in many areas, and studies indicate this formation is the most important contributor of salinity and selenium to the streams in the Gunnison River Basin (Liebermann and others, 1989; Seiler and others, 2003).

---

[1] A water year begins on October 1 of the previous year and ends the following September 30 and is designated by the year in which it ends.

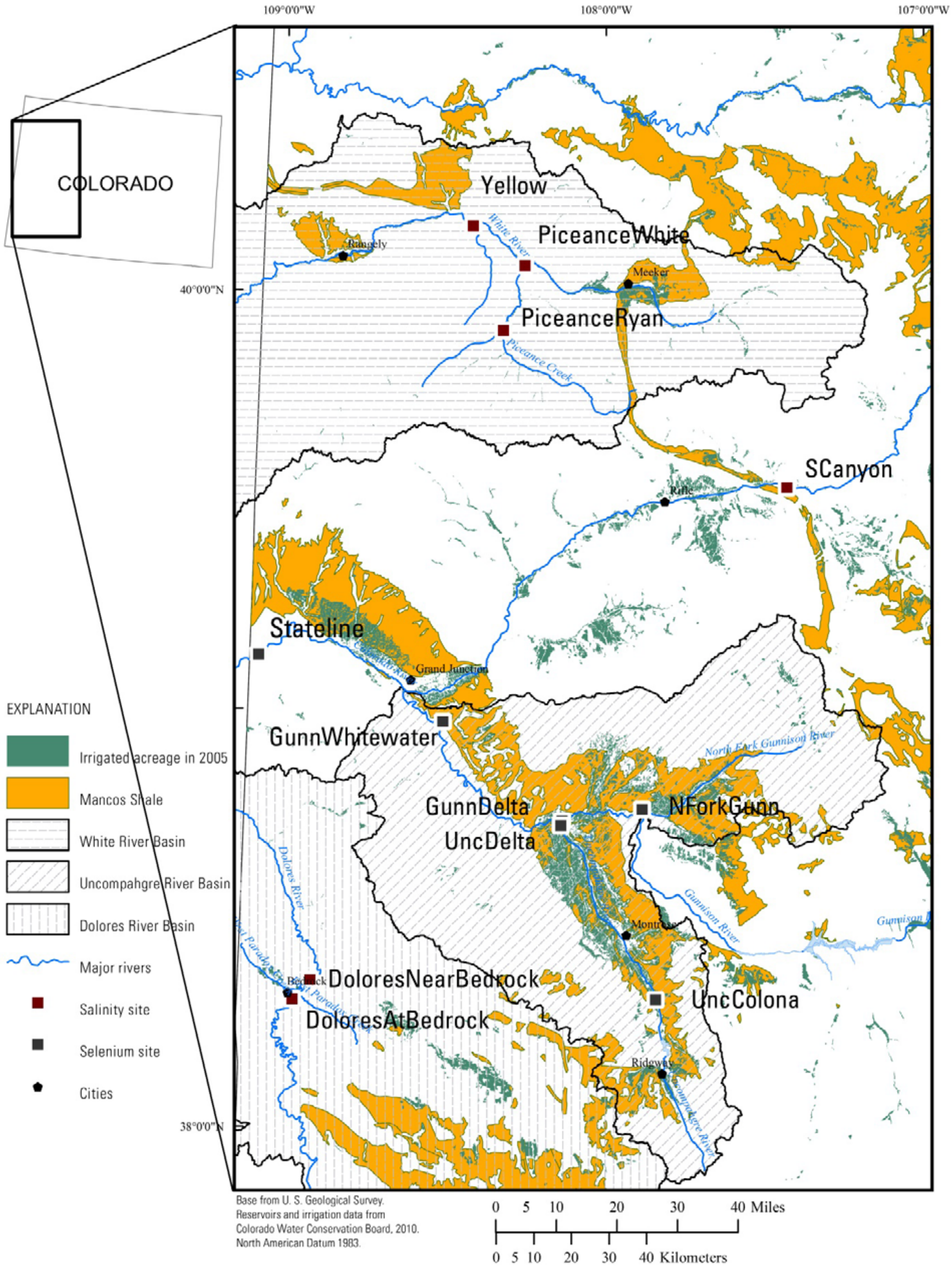

Figure 1. The study area and study sites within the Upper Colorado River Basin, Colorado, also showing irrigated land and the Mancos Shale. (Complete site names given in table 1.)

Study-Site Selection and Data Compilation

Sites chosen for this study were selected on the basis of data availability and cooperator interest. At a minimum, data from discrete measurement of salinity or selenium concentration, streamflow, and specific conductance at each of the sites were needed for model development. In this report, the sites are referred to by the short names listed in table 1. For each site, salinity and selenium concentrations, along with streamflow and specific conductance measured at the time of sampling, were retrieved from the USGS National Water Information System (NWIS) (U.S. Geological Survey, 1998). Discrete sample data generally were available for water years 2009 through 2011. Additional data from October through April of water year 2012 were used to verify the models.

Table 1. Summary statistics of salinity and selenium data used to develop regression models for selected study sites within the Upper Colorado River Basin, Colorado.

[Streamflow, in cubic feet per second; specific conductance, in microsiemens per centimeter at 25 degrees Celsius; selenium concentration, in micrograms per liter; salinity concentration, in milligrams per liter]

U.S Geological Survey streamflow-gaging station number	Site short name (in bold) and full name (in parentheses)	Number of samples	Date		Streamflow			Specific conductance			Constituent concentration		
			Beginning	Ending	Minimum	Median	Maximum	Minimum	Median	Maximum	Minimum	Median	Maximum
Salinity													
09085150	**SCanyon** (Colorado R Abv South Canyon Cr Nr Glenwood Spgs)	18	11/07/08	08/31/11	1,210	2,045	11,500	300	980	1,320	166	547	751
09306200	**PiceanceRyan** (Piceance Creek Bl Ryan Gulch, Nr Rio Blanco, CO)	12	11/13/08	08/23/11	3.0	22	122	892	1,425	2,020	595	967	1,420
09306222	**PiceanceWhite** (Piceance Creek at White River, CO)	12	11/12/08	08/24/11	4.4	30	135	999	1,830	3,320	681	1,242	2,160
09306255	**Yellow** (Yellow Creek Near White River, CO)	12	11/12/08	08/24/11	0.49	1.2	3.5	2,210	3,800	4,260	1,570	2,588	2,890
09169500	**DoloresAtBedrock** (Dolores River at Bedrock, CO)	16	11/19/08	08/18/11	33	62	958	294	768	1,250	177	447	854
09171100	**DoloresNrBedrock** (Dolores River Near Bedrock, CO)	15	11/18/08	08/18/11	38	64	1,000	320	1,870	3,770	192	1,065	2,450

Table 1. Summary statistics of salinity and selenium data used to develop regression models for selected study sites within the Upper Colorado River Basin, Colorado.—Continued

[Streamflow, in cubic feet per second; specific conductance, in microsiemens per centimeter at 25 degrees Celsius; selenium concentration, in micrograms per liter; salinity concentration, in milligrams per liter]

U.S Geological Survey streamflow-gaging station number	Site short name (in bold) and full name (in parentheses)	Number of samples	Date		Streamflow			Specific conductance			Constituent concentration		
			Beginning	Ending	Minimum	Median	Maximum	Minimum	Median	Maximum	Minimum	Median	Maximum
						Selenium							
09163500	**Stateline** (Colorado River Near Colorado-Utah State Line)	18	10/7/08	9/30/11	2,560	3,750	19,300	373	1,060	1,280	1.3	3.6	5.3
09152500	**GunnWhitewater** (Gunnison River Near Grand Junction, CO)	16	10/15/08	09/20/11	1,230	1,860	6,150	390	765	991	1.6	4.2	7.2
09144250	**GunnDelta** (Gunnison River at Delta, CO)	9	07/21/09	09/15/11	827	1,360	13,500	214	538	764	0.48	2.30	3.4
09136100	**NForkGunn** (North Fk Gunnison River Above Mouth Nr Lazear, CO)	16	04/28/09	09/14/11	98	187	4,010	186	1,069	1,460	0.47	2.94	4.6
09149500	**UncDelta** (Uncompahgre River at Delta, CO)	10	07/21/09	09/15/11	166	330	883	735	1,314	1,570	5.4	11.5	20.0
09147500	**UncColona** (Uncompahgre River at Colona CO)	12	10/14/09	08/29/11	56	220	700	373	529	624	0.42	0.83	1.2

The number of discrete samples collected at the six salinity sites differed, and the range of values of each data type also differed between sites. Although the number of samples differed between the salinity sites, a data comparison between the sites gives a qualitative indication of the varied water quality in the Upper Colorado River Basin. The ranges also indicate the limits for which the regression models developed from the data are representative. The SCanyon (09085150) site had the largest streamflows (minimum, 1,210 cubic feet per second (ft^3/s); maximum, 11,500 ft^3/s). The range of the specific conductance (minimum, 300 microsiemens per centimeter at 25 degrees Celsius (µS/cm); maximum, 1,320 µS/cm) was at the low end of the overall range for the salinity sites, and the salinity concentrations (minimum, 166 milligrams per liter (mg/L); maximum, 751 mg/L) were the lowest measured at any of the salinity sites (table 1). In the White River Basin, 12 discrete water-quality samples were collected within water years 2009–11 from each the salinity sites. Based on the discrete data, the White River Basin sites—PiceanceRyan (09306200), PiceanceWhite (09306222), and Yellow (09306255)—had the lowest streamflows (minimums less than or equal to 4.4 ft^3/s; maximums less than or equal to 135 ft^3/s) (table1). The White River Basin sites also had minimum specific conductance greater than or equal to 892 µS/cm, which was greater than that measured at any of the salinity sites, and had the highest minimum concentrations of salinity (greater than or equal to 595 mg/L) of the salinity sites. In the Dolores River Basin—DoloresAtBedrock (09169500) and DoloresNrBedrock (09171100)—streamflow was about an order of magnitude greater than at the White River Basin sites, although the specific conductances were similar (table 1). DoloresNrBedrock is downstream from DoloresAtBedrock, and East and West Paradox Creeks enter the Dolores River between those two sites (fig. 1). East Paradox Creek flows through Paradox Valley, which is the most concentrated source of salt in the Colorado River Basin (Bureau of Reclamation, 2011a). Contrasting those data, discrete water-quality sample data collected in the Dolores River Basin indicated minimum salinity concentrations in the low range of the salinity sites (less than or equal to 192 mg/L).

Similar to the salinity sites, the number of discrete samples collected at the six selenium sites differed, and the range of values of each data type also differed between sites. Streamflows at Stateline (09163500) were the highest among the selenium sites (minimum, 2,560 ft^3/s; maximum, 19,300 ft^3/s), and specific conductances were near the middle of the ranges of the other sites (table 1). The range of selenium concentrations at Stateline (minimum, 1.3 micrograms per liter (µg/L); maximum, 5.3 µg/L) were near the middle of the ranges of concentrations measured at the selenium sites (table 1). At the GunnWhitewater site (09152500), the minimum streamflow (1,230 ft^3/s) and minimum specific conductance (390 µS/cm) were the second highest of the selenium sites (table 1). The selenium concentrations (minimum, 1.6 µg/L; maximum, 7.2 µg/L) at GunnWhitewater were the near highest measured at the selenium sites (table 1). The streamflows increased from NForkGunn (09136100, upstream) to GunnDelta (09144250, downstream), and the lowest minimum specific conductances (less than or equal to 214 µg/L) were measured at those sites (table 1). The selenium concentrations (minimums less than or equal to 0.48 µg/L and maximums less than or equal to 4.6 µg/L) were some of the lowest of the six selenium sites (table 1). The lowest selenium concentrations (minimum, 0.42 µg/L; maximum, 1.2 µg/L) were measured at UncColona (09147500); UncDelta (09149500, downstream of UncColona) had the highest selenium concentrations (minimum, 5.4 µg/L; maximum, 20.0 µg/L) of the six selenium sites (table 1).

Model Development

Model development used water year 2009–11 data associated with the sampling that included the discrete measurements of constituent concentration (response variables), specific conductance (in microsiemens per centimeter at 25 degrees Celsius) and streamflow (in cubic feet per second). Consistent with the ordinary least squared regression procedure described by Liebermann and others (1989), the input values for constituent concentration, streamflow, and specific conductance were natural-log transformed.

Whereas the equations for estimating salinity concentration used either specific conductance or streamflow as the explanatory variable, the procedure to develop the models to estimate selenium was slightly different in that variables representing seasonality were included as potential explanatory variables. Sine and cosine terms were included to address seasonal differences and account for the possibility of two and (or) three annual cycles (in equation 1: k=1 or 2) (Helsel and Hirsch, 2002). Both variables (sine and cosine) are required to account for the amplitude, or magnitude, and the day of the peak (Helsel and Hirsch, 2002), where time, T, is the seasonality term representing the decimal portion of the year starting January 1. A generalized form of a regression equation using streamflow and seasonality to estimate selenium concentration is equation 1:

$$\ln \hat{C} = b_o + b_1(\ln Q) + b_2 \sin(k2\pi T) + b_3 \cos(k2\pi T) \quad (1)$$

where, b_0 is the regression equation intercept, b_n is the coefficient of the nth explanatory variable, T is the seasonality term representing the decimal portion of the year starting January 1, and lnQ is streamflow in units of natural-log-transformed cubic feet per second.

Analysis of all the possible combinations of explanatory variables yielded the final model equations for selenium. The choice of the best model was based on values of the Prediction Error Sum of Squares (*PRESS*), standard error of the model (*se*), *Adjusted* R^2, and the collinearity of the model as determined by the variance inflation factor (*VIF*) statistic (Helsel and Hirsch, 2002). Low values of the first two statistics and the highest *Adjusted* R^2 indicated the best model. Explanatory variables within a model were considered collinear when the *VIF* statistics were greater than 10 (Helsel and Hirsch, 2002). The explanatory variables in the final models were chosen such that instances of collinearity did not occur.

Model diagnostic plots, used in addition to the evaluation statistics for model selection, included a normal-probability plot of residuals and plots of the standardized residuals. Normality of model residuals was apparent when the normal-probability plot was approximately linear and the plots of the standardized residuals had no pattern and showed uniformity of scatter (homoscedasticity). The p-values of the model coefficients had to be marginally significant, or less than 0.05, to be included in the final model. This final criterion has one exception: only one of the model coefficients on the sine-cosine pair of the Fourier series had to be significant for the pair to be included in the model (Helsel and Hirsch, 2002).

Evaluation of the performance of the final models consisted of comparing estimated salinity and selenium concentrations to water-quality sample data from October 2011 through April 2012. There were three water-quality samples at each salinity site available for verification of the salinity models. A total of 13 samples was available for verification of models developed at selenium sites, consisting of 2 samples each at NForkGunn and GunnDelta, 4 samples each at UncColona and GunnWhitewater, and 1 sample at UncDelta. The estimates of concentration from the regression models were the natural log of the estimated concentration, so they needed retransformation into appropriate units. A bias correction factor is required to address retransformation bias (Bradu and Mundlak, 1970), and this factor was calculated for each model by using the smearing method described by Duan (1983). Comparison of the retransformed estimates to the water-quality sample data indicated model performance.

Regression Models For Estimating Salinity and Selenium Concentrations

The regression models for estimating salinity and selenium concentration were developed through evaluating diagnostic statistics and plots. Through the model development process, it became apparent that models which included both specific conductance and streamflow tended to exhibit collinearity; evidence consisted of either *VIF* statistics exceeding a value of 10 and (or) a switch from the expected positive or negative relation with salinity or selenium. Consequently, the final models use either specific conductance or streamflow as an explanatory variable rather than both (table 2).

Table 2. Model coefficients and evaluation statistics for salinity and selenium regression models developed for selected study sites within the Upper Colorado River Basin, Colorado.

[Streamflow, in units of natural-log transformed cubic feet per second; SC, specific conductance, in units of natural-log transformed microsiemens per centimeter at 25 degrees Celsius; --, variables not used in the regression; R^2, coefficient of determination; Standard error in units of natural-log transformed concentration; bold values indicate statistical significance (p less than 0.05)]

U.S. Geological Survey streamflow-gaging station number	Site short name	Y-axis intercept	Streamflow coefficient	SC coefficient	Seasonal terms, k2πT				Model p-value	Adjusted R^2	Standard error	Bias correction factor
					k=1		*k=2*					
					Sine coefficient	Cosine coefficient	Sine coefficient	Cosine coefficient				
					Salinity regression models							
09085150	SCanyon	11.2929	**-0.6558**	--	--	--	--	--	<0.0001	0.992	0.048	1.001
		-0.671	--	**1.0133**	--	--	--	--	<0.0001	0.998	0.023	1.000
09306200	PiceanceRyan	7.5371	**-0.2226**	--	--	--	--	--	<0.0001	0.938	0.049	1.001
		-0.8999	--	**1.0709**	--	--	--	--	<0.0001	0.982	0.027	1.000
09306222	PiceanceWhite	8.1063	**-0.3043**	--	--	--	--	--	<0.0001	0.922	0.082	1.003
		-0.0539	--	**0.9541**	--	--	--	--	<0.0001	0.990	0.029	1.000
09306255	Yellow	7.8445	**-0.1743**	--	--	--	--	--	<0.0001	0.342	0.130	1.007
		0.3078	--	**0.9170**	--	--	--	--	<0.0001	0.978	0.024	1.000
09169500	DoloresAtBedrock	7.0376	**-0.2433**	--	--	--	--	--	<0.0001	0.339	0.324	1.053
		-0.7738	--	**1.0322**	--	--	--	--	<0.0002	0.969	0.070	1.002
09171100	DoloresNrBedrock	9.5603	**-0.6146**	--	--	--	--	--	0.0001	0.671	0.422	1.087
		-0.5358	--	**0.9931**	--	--	--	--	0.0001	0.982	0.098	1.004
					Selenium regression models							
09163500	Stateline	5.8133	**-0.5529**	--	**-0.2222**	-0.0590	--	--	<0.0001	0.917	0.128	1.007
		-6.0433	--	**1.0530**	**-0.0920**	-0.1049	--	--	<0.0001	0.956	0.093	1.003
09152500	GunnWhitewater	5.8155	**-0.5740**	--	**-0.2384**	0.0897	--	--	<0.0001	0.879	0.152	1.009
		-7.8680	--	**1.4053**	0.0418	**0.1408**	--	--	<0.0001	0.909	0.132	1.006
09144250	GunnDelta	4.8568	**-0.5813**	--	**-0.4201**	-0.0979	--	--	<0.0001	0.974	0.102	1.003
		-8.9887	--	**1.545**	--	--	--	--	<0.0001	0.973	0.104	1.008
09136100	NForkGunn	4.4273	**-0.6349**	--	--	--	--	--	<0.0001	0.955	0.195	1.017
		-6.387	--	**1.0764**	--	--	--	--	<0.0001	0.986	0.108	1.005
09149500	UncDelta	5.5657	**-0.5440**	--	**-0.2338**	**0.1313**	**0.0861**	**0.2009**	<0.0001	0.993	0.037	1.000
		-10.9850	--	**1.8861**	**0.2214**	-0.0909	--	--	0.0002	0.930	0.120	1.004
09147500	UncColona	0.8729	**-0.2227**	--	--	--	--	--	0.0251	0.350	0.276	1.031
		-9.5214	--	**1.4805**	--	--	--	--	0.0009	0.648	0.203	1.018

Salinity Models

Comparison of *Adjusted* R^2 and standard error statistics of the two salinity models developed at each site showed that the models (p-value < 0.05) using specific conductance as the explanatory variable performed better than those using streamflow (table 2). A model performing perfectly would fit a 1:1 line exactly. The SCanyon models performed similarly, although the model using specific conductance as the explanatory variable more closely fit the 1:1 line than the model using streamflow (fig. 2*A*). In the White River Basin this relation was most evident when comparing measured and estimated salinity concentrations at Yellow; the measured concentration deviates from the estimated concentration more for the model using streamflow as the explanatory variable than for the model using specific conductance (fig. 2*D*). The utility of specific conductance as an explanatory variable for estimating salinity concentration was most evident in the Dolores River Basin. Models using streamflow as an explanatory variable did not perform as well (greater deviation from the 1:1 line) at lower streamflow, which corresponds to high concentration, than they did at higher streamflow, which corresponds to low concentration (fig. 2*E*, *F*).

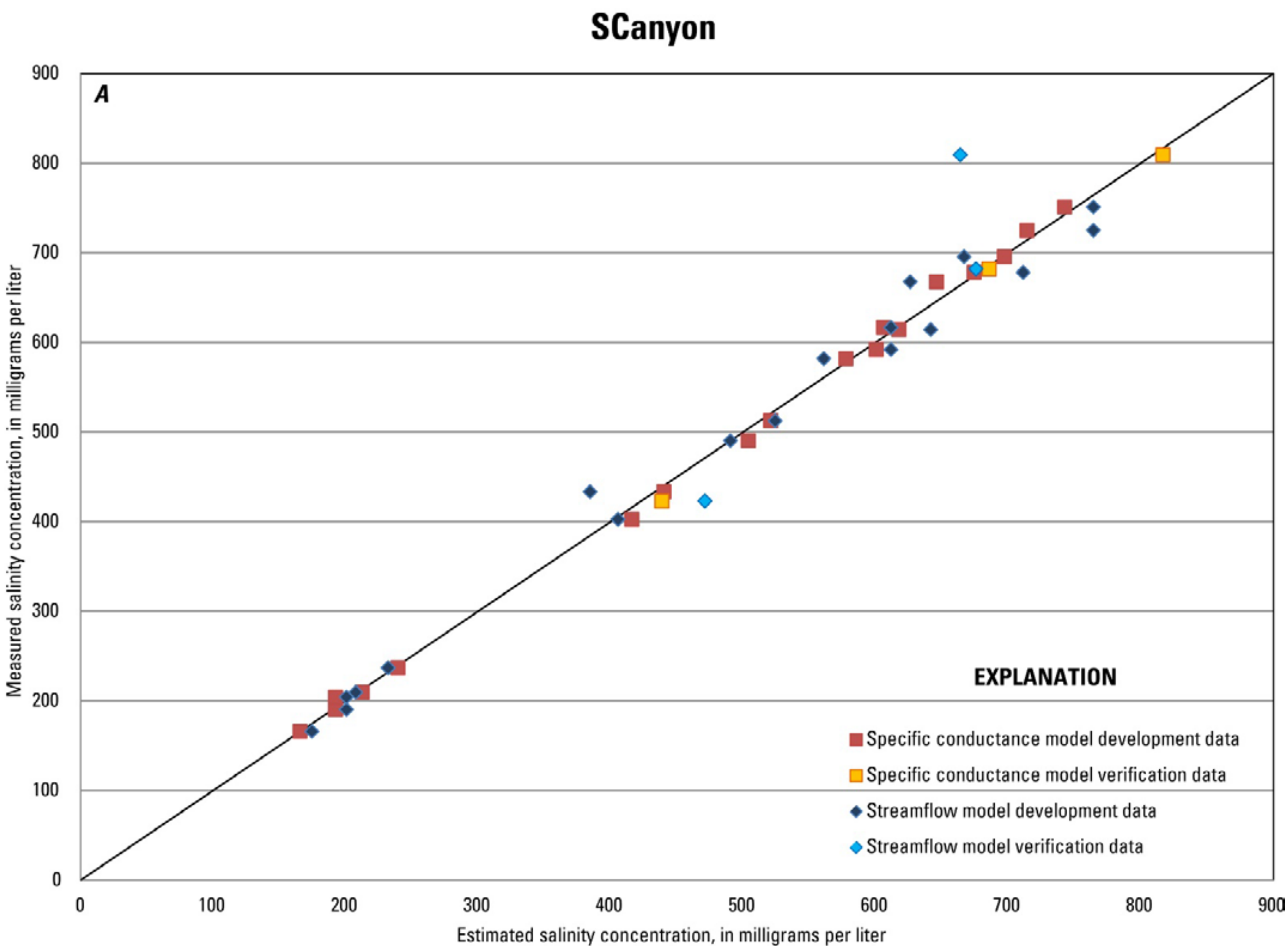

Figure 2. Comparison of measured and estimated salinity concentrations at six sites using either specific conductance or streamflow as an explanatory variable. *A*, SCanyon. *B*, PiceanceRyan. *C*, PiceanceWhite. *D*, Yellow. *E*, DoloresAtBedrock. *F*, DoloresNrBedrock.

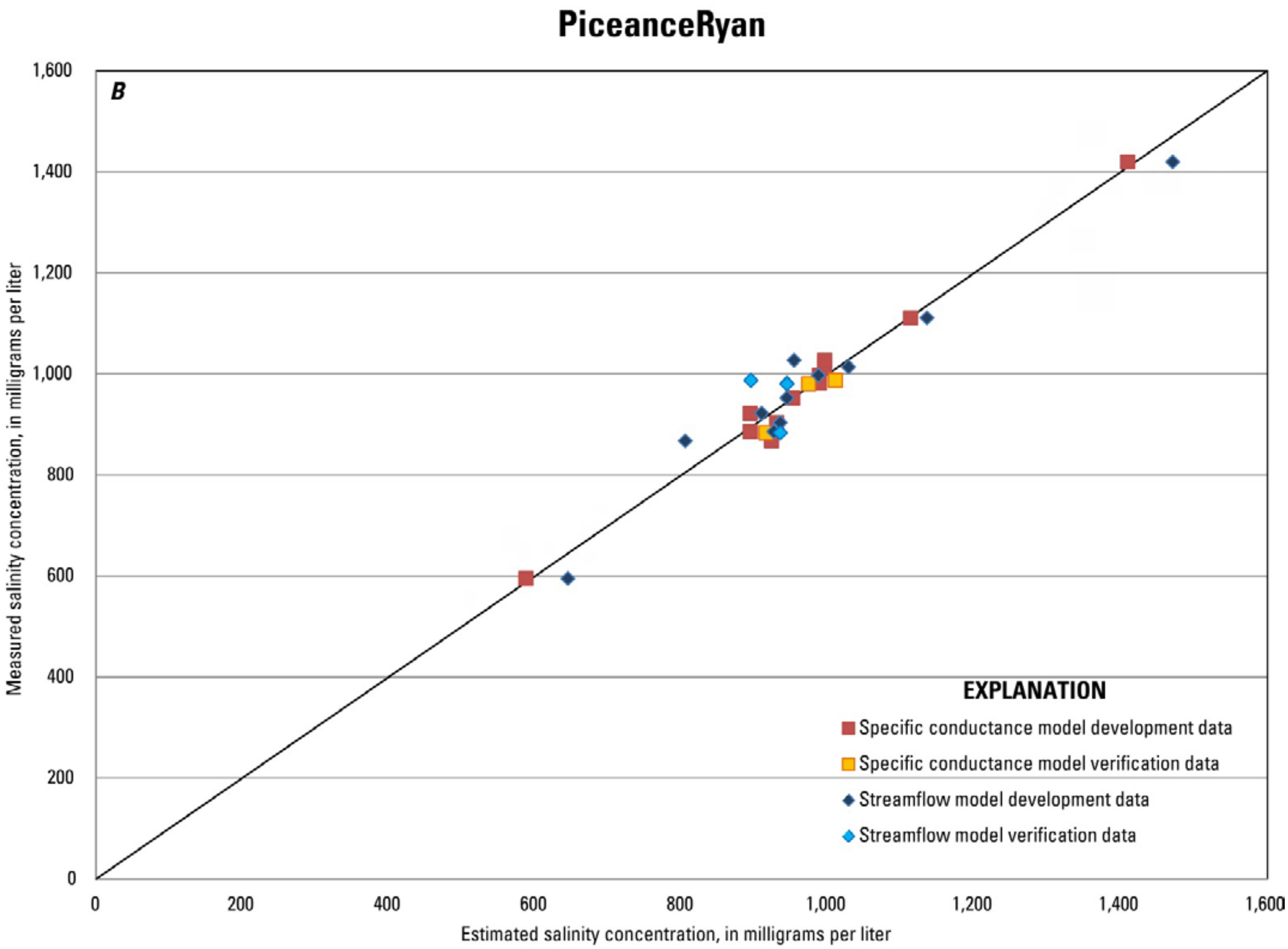

Figure 2. Comparison of measured and estimated salinity concentrations at six sites using either specific conductance or streamflow as an explanatory variable. *A*, SCanyon. *B*, PiceanceRyan. *C*, PiceanceWhite. *D*, Yellow. *E*, DoloresAtBedrock. *F*, DoloresNrBedrock.—Continued

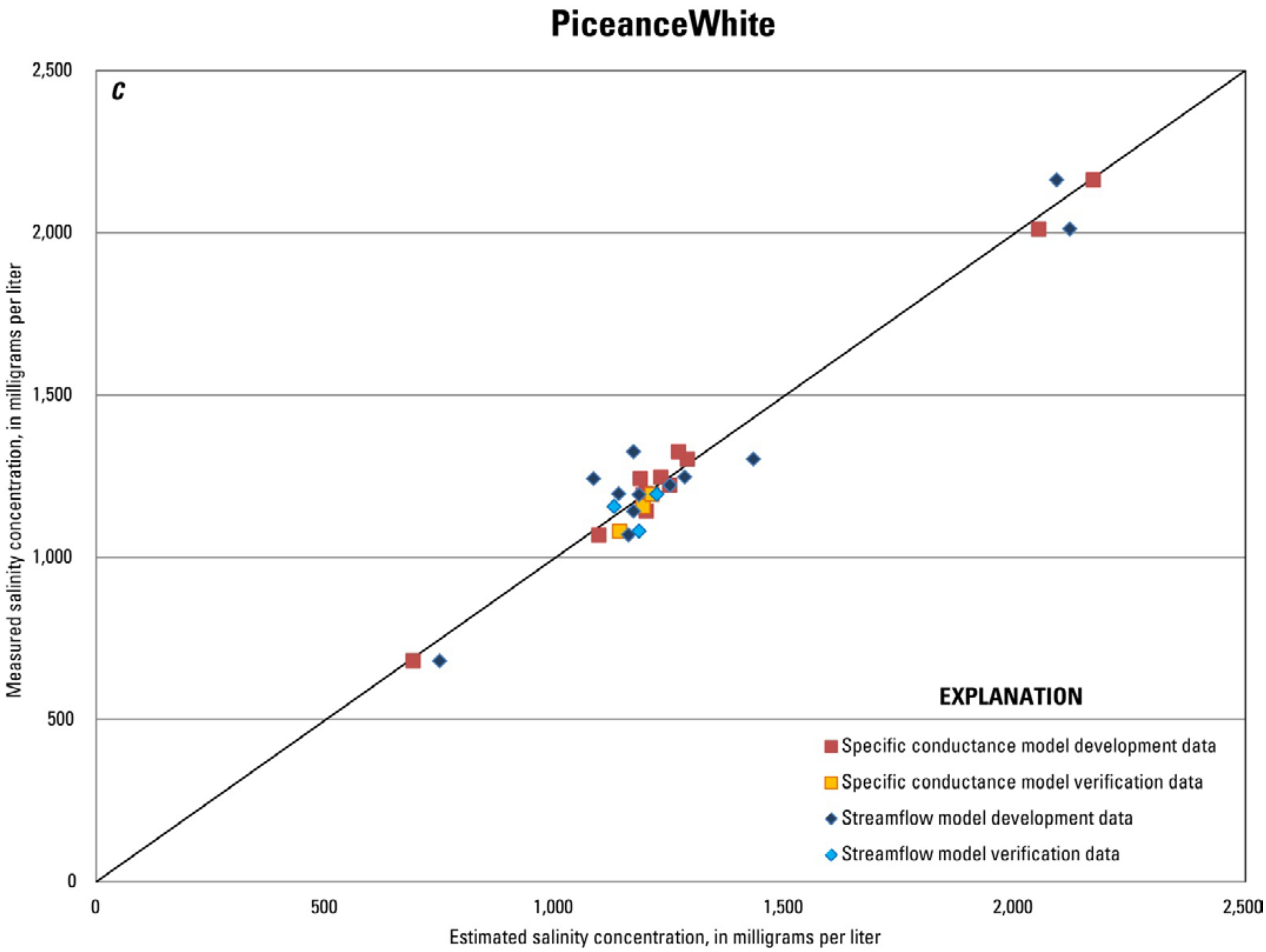

Figure 2. Comparison of measured and estimated salinity concentrations at six sites using either specific conductance or streamflow as an explanatory variable. *A*, SCanyon. *B*, PiceanceRyan. *C*, PiceanceWhite. *D*, Yellow. *E*, DoloresAtBedrock. *F*, DoloresNrBedrock.—Continued

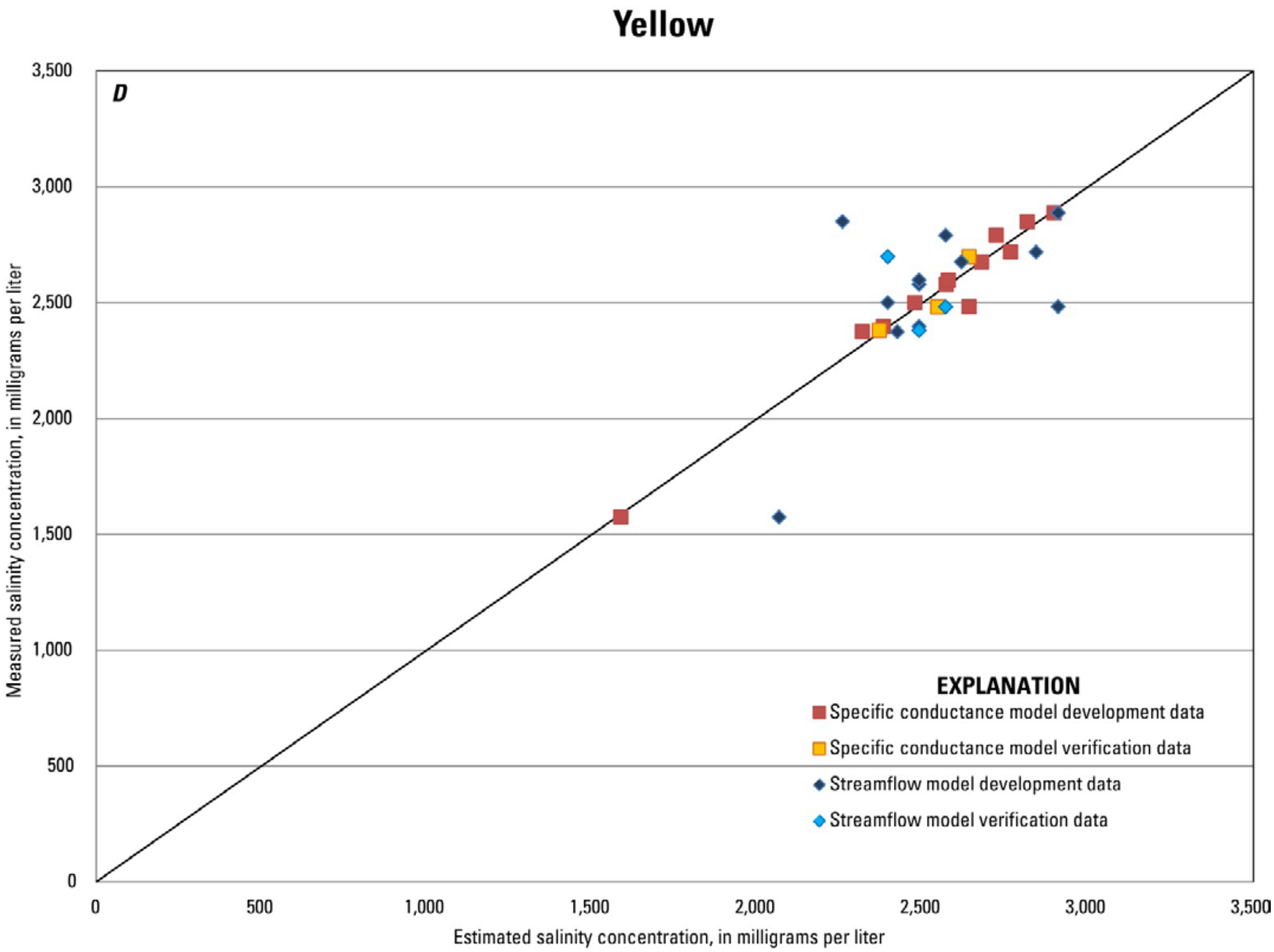

Figure 2. Comparison of measured and estimated salinity concentrations at six sites using either specific conductance or streamflow as an explanatory variable. *A*, SCanyon. *B*, PiceanceRyan. *C*, PiceanceWhite. *D*, Yellow. *E*, DoloresAtBedrock. *F*, DoloresNrBedrock.—Continued

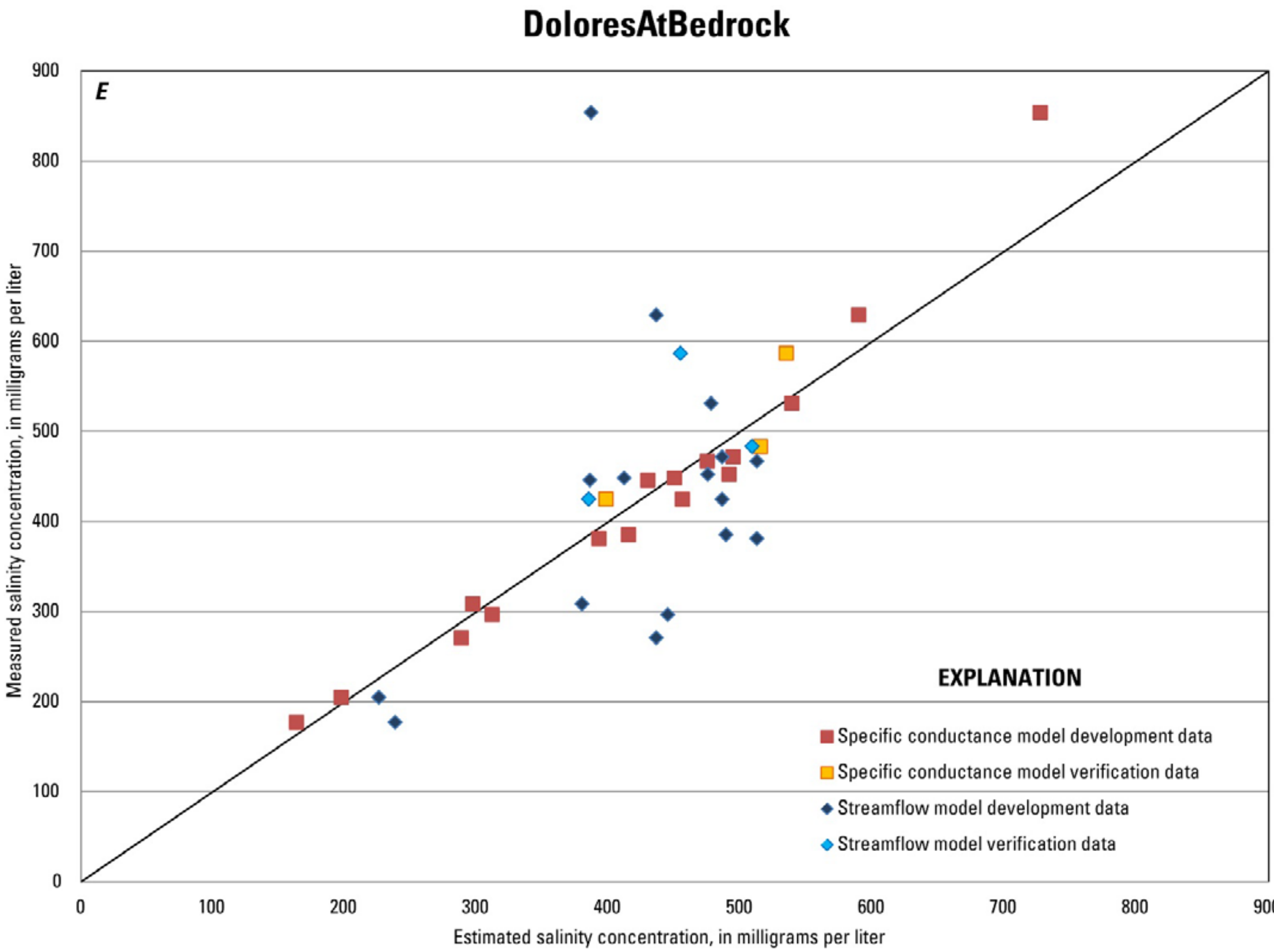

Figure 2. Comparison of measured and estimated salinity concentrations at six sites using either specific conductance or streamflow as an explanatory variable. *A*, SCanyon. *B*, PiceanceRyan. *C*, PiceanceWhite. *D*, Yellow. *E*, DoloresAtBedrock. *F*, DoloresNrBedrock.—Continued

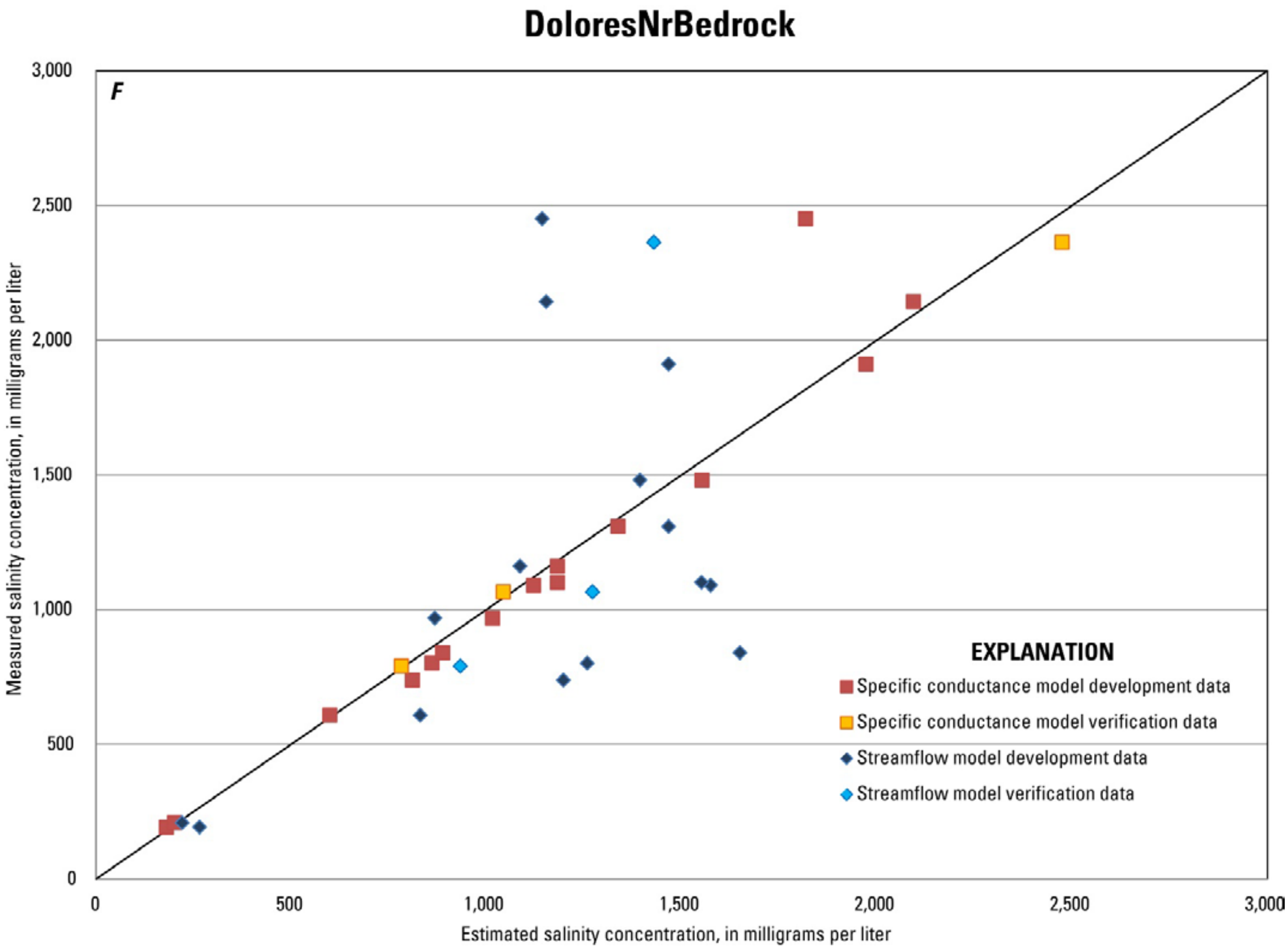

Figure 2. Comparison of measured and estimated salinity concentrations at six sites using either specific conductance or streamflow as an explanatory variable. *A*, SCanyon. *B*, PiceanceRyan. *C*, PiceanceWhite. *D*, Yellow. *E*, DoloresAtBedrock. *F*, DoloresNrBedrock.—Continued

Selenium Models

The addition of multiple explanatory variables improved the ability of the models to estimate selenium concentration at several selenium sites over using solely streamflow or specific conductance. The models developed at Stateline and GunnWhitewater included the seasonal explanatory variable representing two annual cycles ($k2\pi T$, where k equals 1), and the evaluation statistics indicated that the model using specific conductance performed better than the model using streamflow. The models for GunnDelta and UncDelta using streamflow included more seasonal explanatory variables than the models using specific conductance; GunnDelta used two annual cycles compared to none and UncDelta used two and three cycles compared to just two in the specific conductance model. In contrast to all the other models in this report, the evaluation statistics, the *Adjusted* R^2 and standard error, indicated that models using streamflow at GunnDelta and UncDelta (p-value <0.05) performed better than those using specific conductance (table 2). The relation between selenium estimates and the 1:1 line illustrated in figure 3 indicates that models developed at GunnDelta and UncDelta (fig. 3*C*,*E*) perform better at higher streamflow (low concentration) than at lower streamflow. At sites like NForkGunn and UncColona, the *Adjusted* R^2 and standard error indicated that models using a single explanatory variable performed better and that models using specific conductance as the explanatory variable, rather than streamflow, performed best. Regarding the visual comparison of measured concentrations to estimated concentrations, the benefits were less clear (fig. 3*D*). At NForkGunn, the estimates from both salinity models more closely resembled the measured concentrations at higher streamflows, which correspond to lower concentrations, than at lower streamflows, which correspond to higher concentrations (fig. 3*D*). The amount of streamflow at UncColona did not seem to affect the ability of the model to estimate salinity concentration (fig. 3*F*).

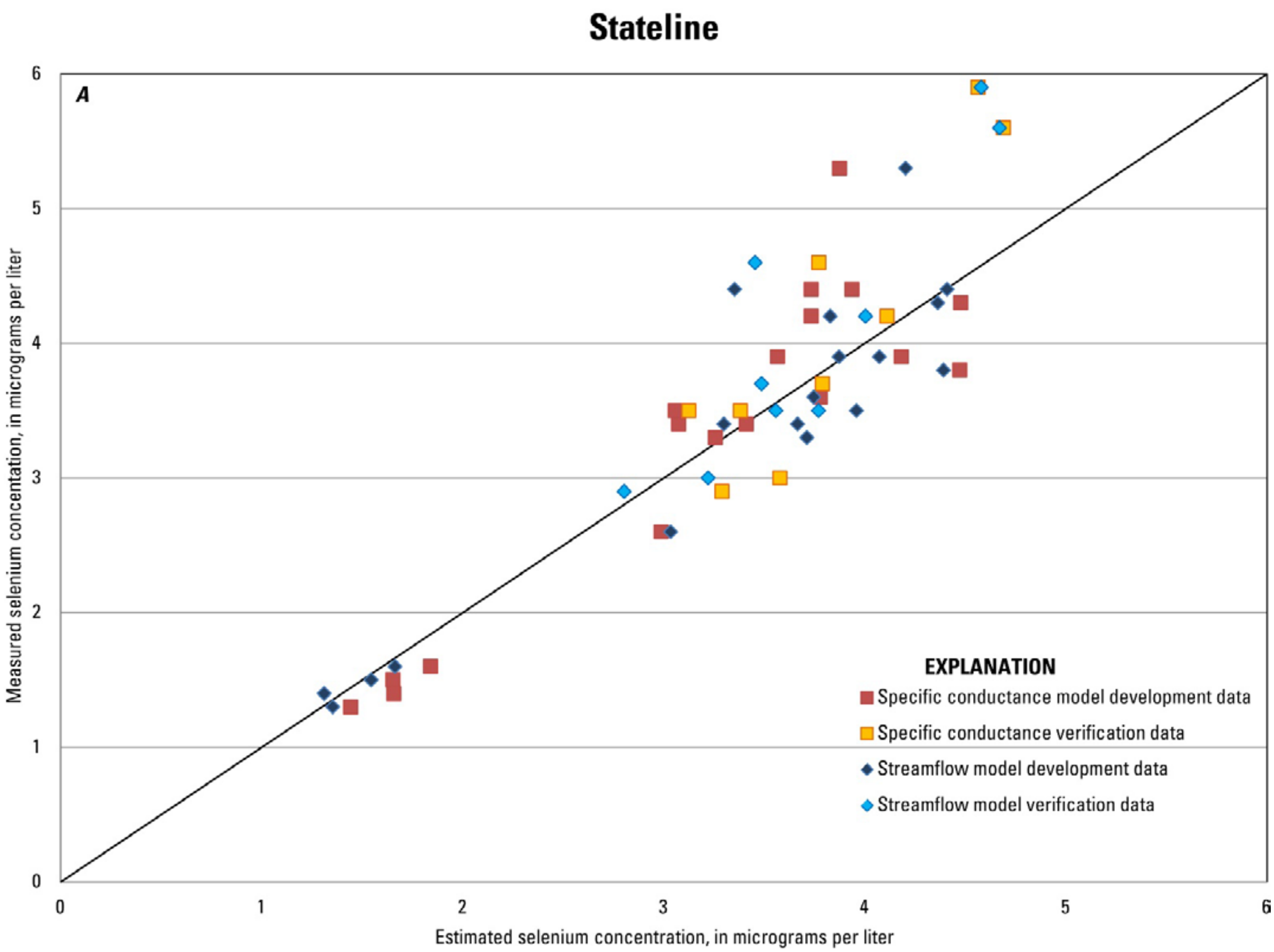

Figure 3. Comparison of measured and estimated selenium concentrations at six sites using specific conductance, streamflow, and (or) seasonality explanatory variables. *A*, Stateline. *B*, GunnWhitewater. *C*, GunnDelta. *D*, NForkGunn. *E*, UncDelta. and *F*, UncColona.

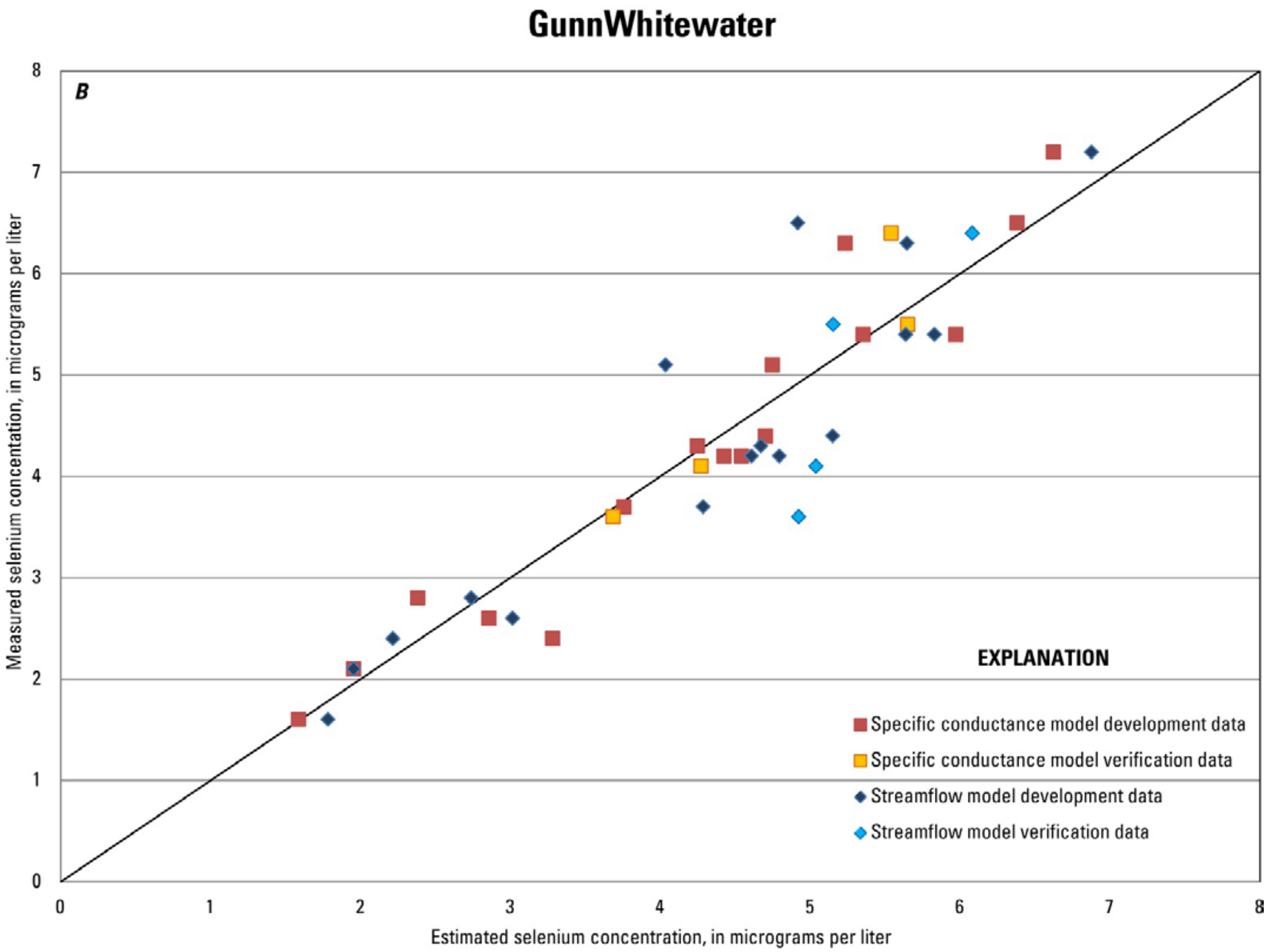

Figure 3. Comparison of measured and estimated selenium concentrations at six sites using specific conductance, streamflow, and (or) seasonality explanatory variables. *A*, Stateline. *B*, GunnWhitewater. *C*, GunnDelta. *D*, NForkGunn. *E*, UncDelta. and *F*, UncColona.—Continued

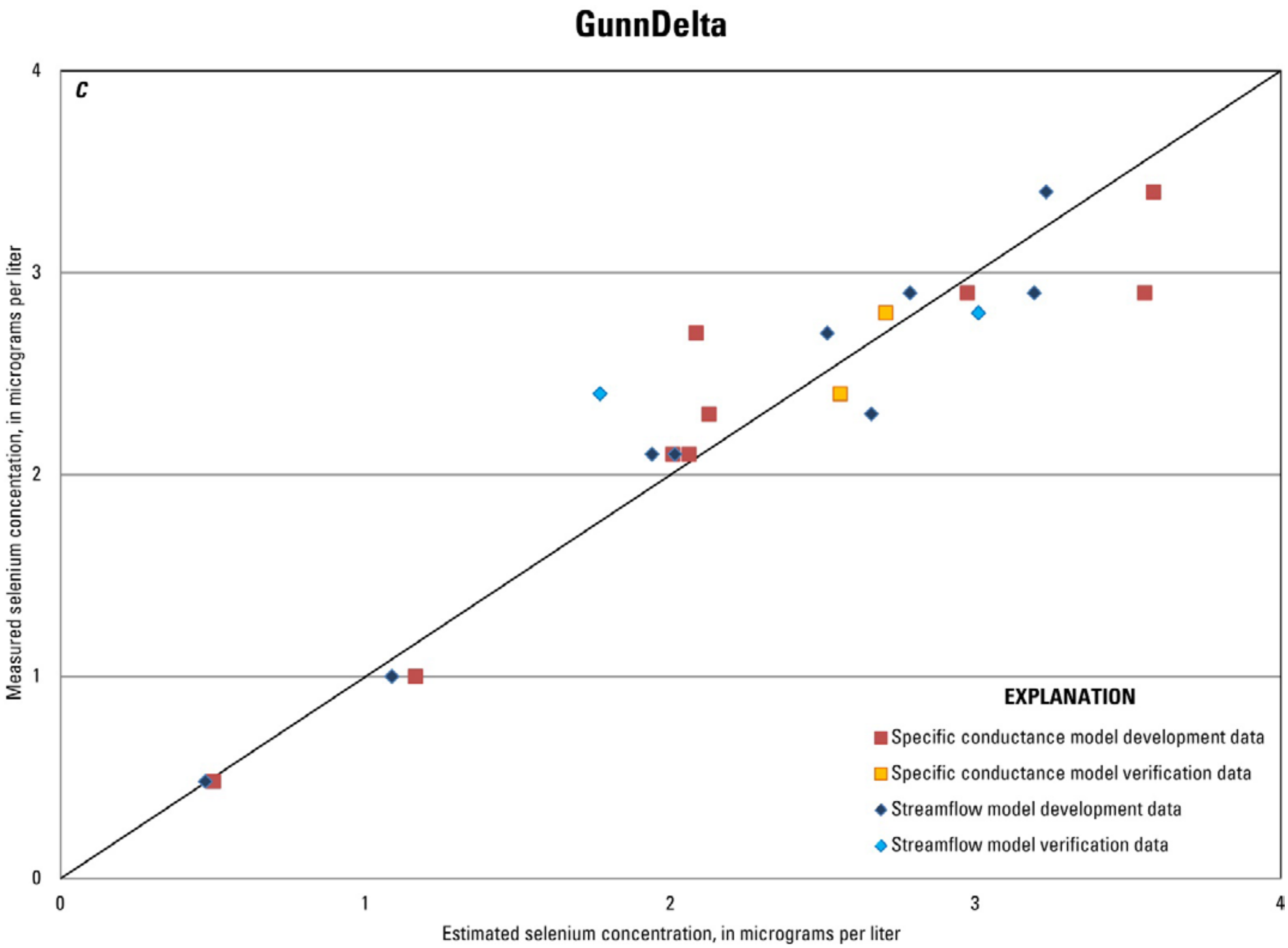

Figure 3. Comparison of measured and estimated selenium concentrations at six sites using specific conductance, streamflow, and (or) seasonality explanatory variables. *A*, Stateline. *B*, GunnWhitewater. *C*, GunnDelta. *D*, NForkGunn. *E*, UncDelta. and *F*, UncColona.—Continued

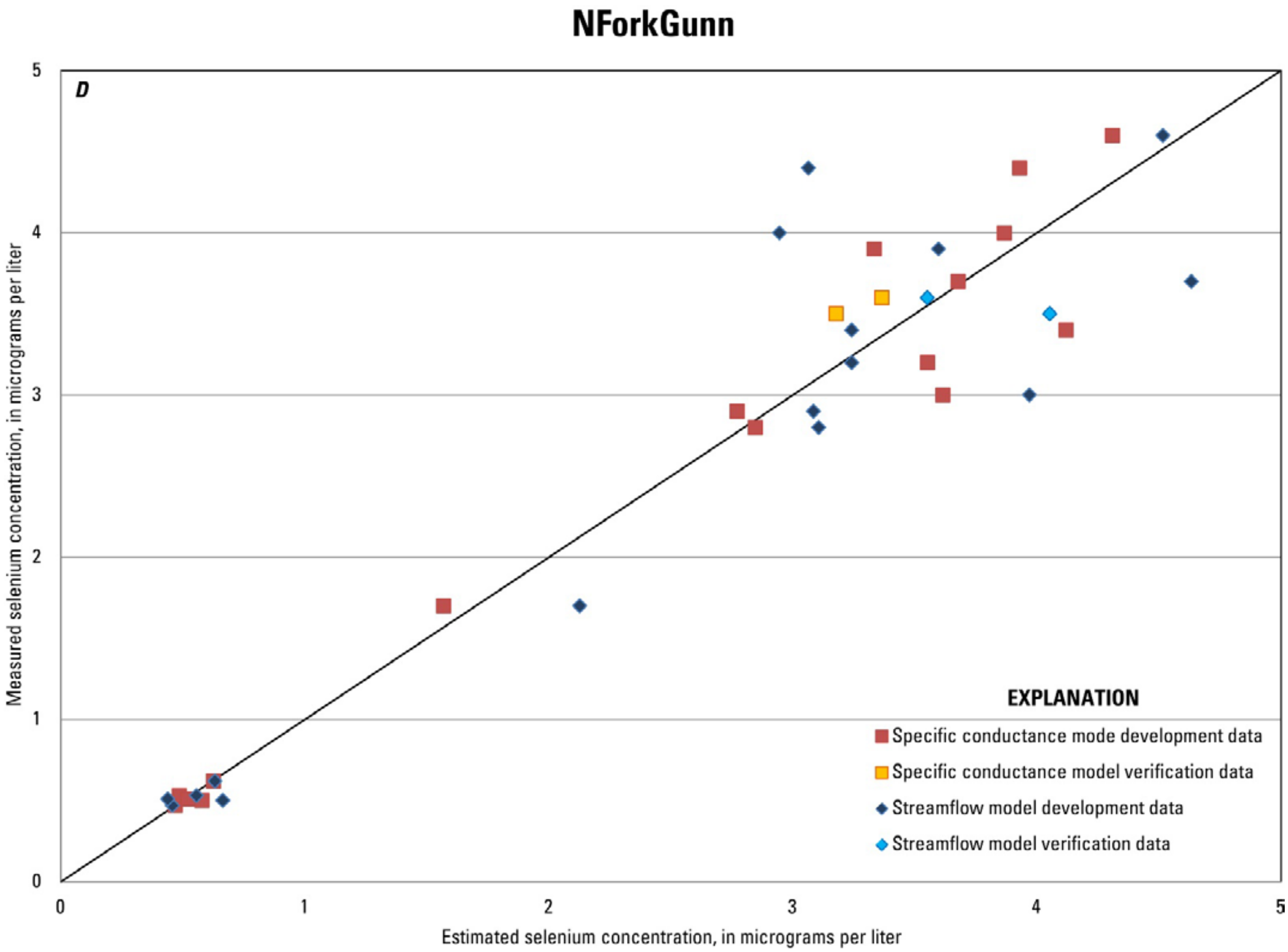

Figure 3. Comparison of measured and estimated selenium concentrations at six sites using specific conductance, streamflow, and (or) seasonality explanatory variables. *A*, Stateline. *B*, GunnWhitewater. *C*, GunnDelta. *D*, NForkGunn. *E*, UncDelta. and *F*, UncColona.—Continued

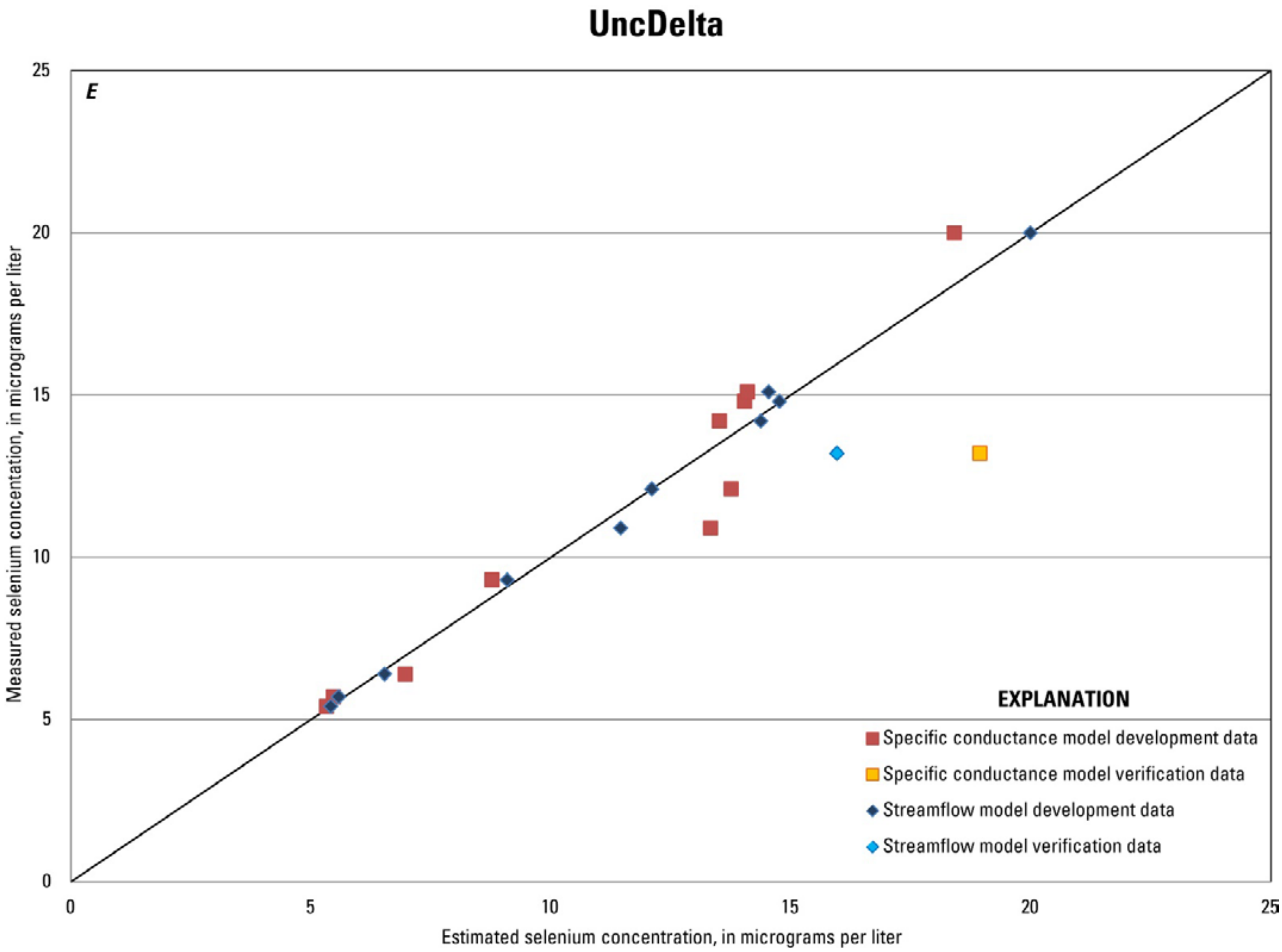

Figure 3. Comparison of measured and estimated selenium concentrations at six sites using specific conductance, streamflow, and (or) seasonality explanatory variables. *A*, Stateline. *B*, GunnWhitewater. *C*, GunnDelta. *D*, NForkGunn. *E*, UncDelta. and *F*, UncColona.—Continued

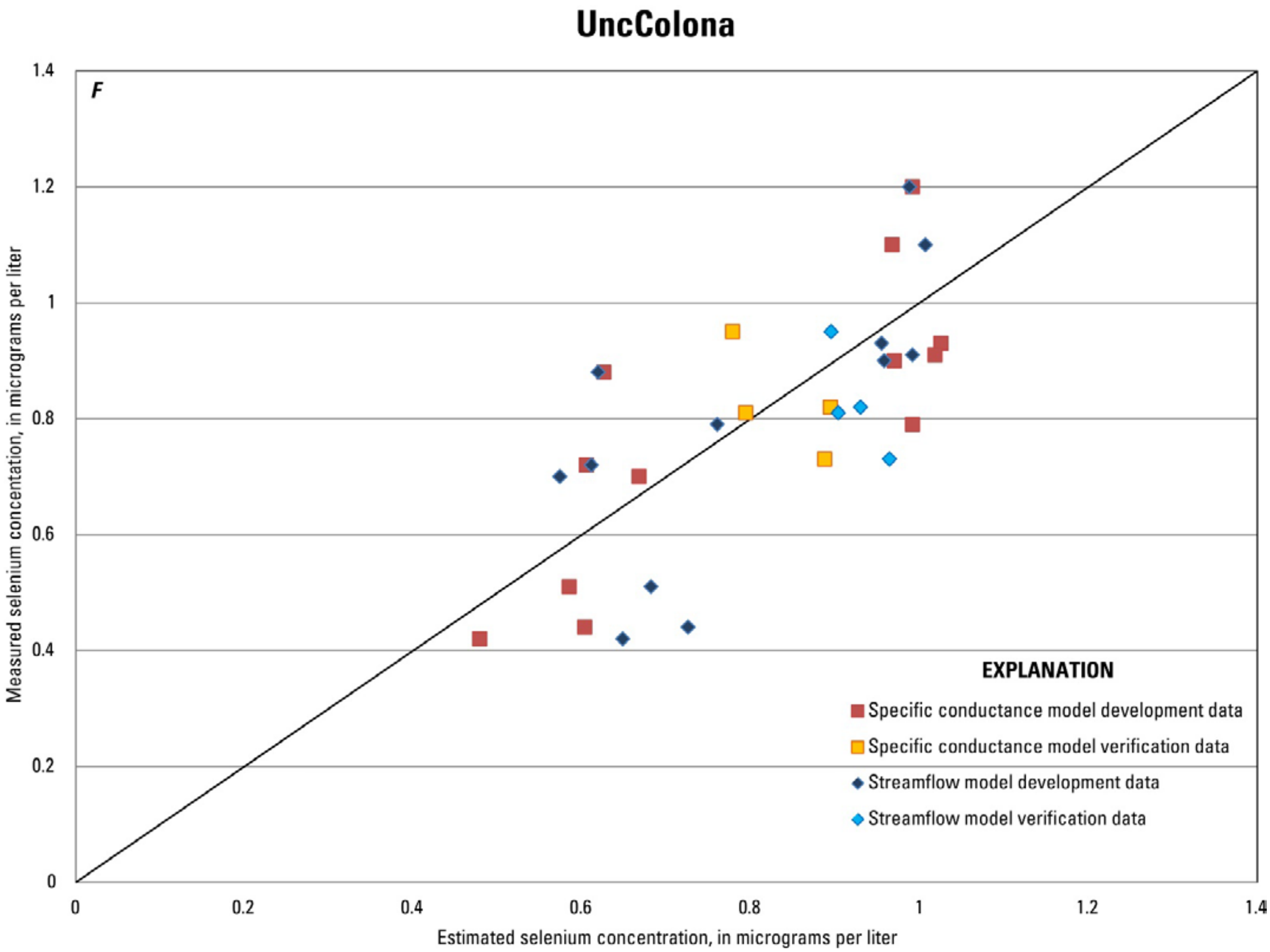

Figure 3. Comparison of measured and estimated selenium concentrations at six sites using specific conductance, streamflow, and (or) seasonality explanatory variables. *A*, Stateline. *B*, GunnWhitewater. *C*, GunnDelta. *D*, NForkGunn. *E*, UncDelta. and *F*, UncColona.—Continued

Use of Continuous Data in the Application of Salt and Selenium Regression Models

Continuous streamflow and specific conductance data collected at study sites provide the means to examine temporal variability in constituent concentration and load. The regression models can estimate continuous concentrations or loads on the basis of continuous specific conductance or streamflow data. In-stream load is computed by multiplying the estimated concentration by the streamflow (cubic feet per second) and a unit conversion constant. An example of concentrations estimated by using continuous specific conductance data is provided for PiceanceWhite and GunnWhitewater (fig. 4; data gaps in the continuous record appear as breaks in the figure). Similar plots are available for other sites at the USGS National Real-Time Water Quality Webpage (*http://nrtwq.usgs.gov*) and provide water-resource managers with a means of improving their general understanding of how constituent concentration or load can change annually, seasonally, or in real time. With respect to the models presented in this study, the amount of relative streamflow should be considered when interpreting estimated salinity or selenium concentration or loads. The error associated with the log-transformed salinity and selenium estimates is consistent in log space; however, when the estimates are transformed into non-log values, the error increases as the estimates decrease (Helsel and Hirsch, 2002).

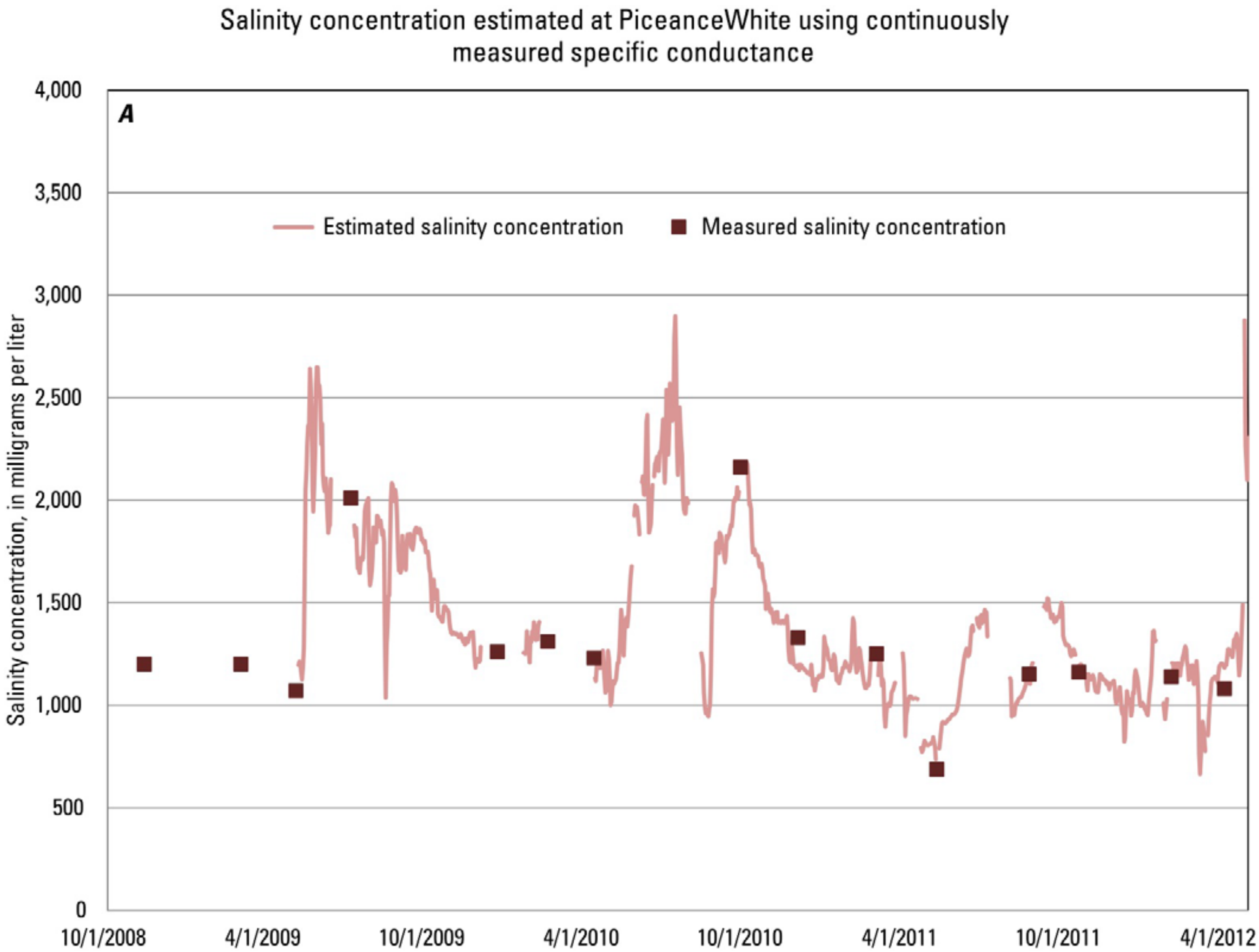

Figure 4. Example of the use of continuous data with the regression models to estimate *A*, salinity concentration at PiceanceWhite and *B*, selenium concentration at GunnWhitewater using continuously measured specific conductance.

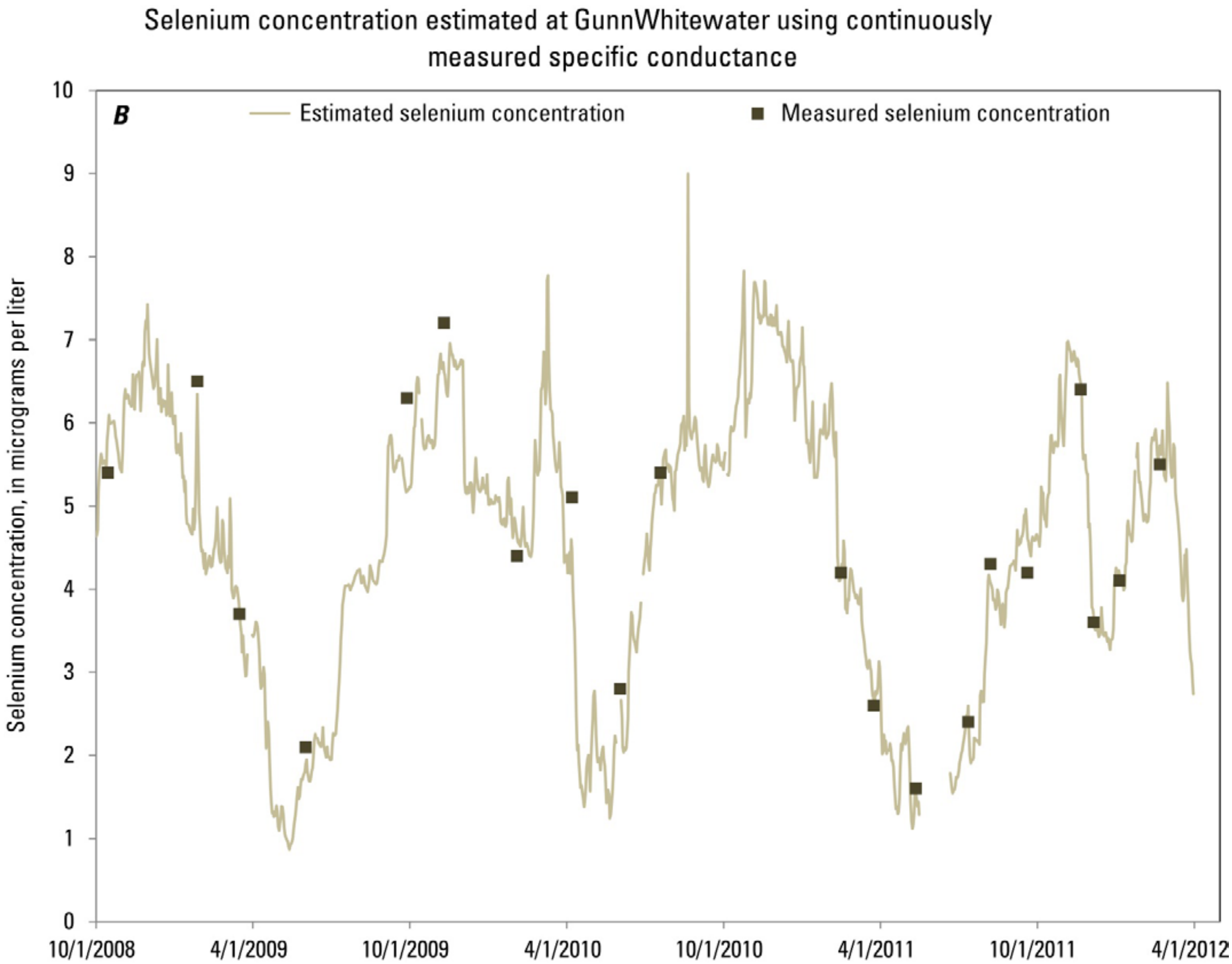

Figure 4. Example of the use of continuous data with the regression models to estimate *A*, salinity concentration at PiceanceWhite and *B*, selenium concentration at GunnWhitewater using continuously measured specific conductance.—Continued

Summary

Elevated concentrations of salinity and selenium in the tributaries and main-stem reaches of the Colorado River are a water-quality concern and have been the focus of remediation efforts for many years. “Salinity” refers to mineral salts or solids dissolved in water, and selenium is a trace element that bioaccumulates in the food chain. Salt and selenium limit municipal uses of water, reduce agricultural productivity, and, in the case of selenium, can lead to mortality, abnormalities, and reproductive failure in waterfowl and fish. Selenium is paradoxical in that it is a nutritional requirement in small amounts but toxic in slightly greater amounts. Toxic levels of selenium can cause reproductive failure, deformities, and other adverse effects in birds and fish, including some threatened and endangered fish species. Natural, nonpoint sources of salinity and selenium generally originate from the weathering and dissolution of geological formations that have high salt and selenium content. Land-management practices aimed at limiting the amount of salt and selenium that reaches the stream have focused on improving the methods by which irrigation water is conveyed and distributed.

Federal land managers implement improvements in accordance with the Colorado River Basin Salinity Control Act of 1974, which directs Federal land managers to enhance and protect the quality of water available in the Colorado River. In 1989, multiple linear regression models to estimate annual and monthly salinity concentrations were developed for the Bureau of Reclamation to assist in evaluating and planning salinity-control needs. Streamflow and specific conductance (depending on data availability) were the explanatory variables used in the regression models. A 2012 study produced multiple linear regression models capable of estimating selenium concentrations at two sites in the Gunnison and Colorado Rivers. These models used daily streamflow and time data, capturing a seasonal cycling of selenium, as their explanatory variables.

In an effort to assist in evaluating and mitigating the detrimental effects of salinity and selenium, the U.S. Geological Survey, in cooperation with the Bureau of Reclamation, the Colorado River Water Conservation District, and the Bureau of Land Management, analyzed discrete salinity and selenium data collected at new sites to develop regression models for those sites. The study area and sites are on the Colorado River or in one of three small basins in Western Colorado: the White River Basin, the Lower Gunnison River Basin, and the Dolores River Basin. At a minimum, data from discrete measurement of salinity or selenium concentrations, streamflow, and specific conductance at each of the sites were needed for model development. Using data collected from water years 2009 through 2011, regression models, able to estimate concentrations, were developed for salinity at six sites (PiceanceRyan, PiceanceWhite, Yellow, SCanyon, DoloresAtBedrock, and DoloresNrBedrock) and selenium at six sites (NForkGunn, UncColona, GunnWhitewater, GunnDelta, and UncDelta). Evaluation of the performance of the regression models consisted of comparing estimated salinity and selenium concentrations to water-quality sample data from October 2011 through April 2012.

The regression models for estimating salinity and selenium concentration were developed through evaluating diagnostic statistics and plots. Through the model development process, it became evident that models that included both specific conductance and streamflow tended to exhibit collinearity. Consequently, the final models use specific conductance or streamflow as an explanatory variable rather than both. Comparison of the *Adjusted* R^2 and standard error statistics of the two salinity models developed at each site (p-value<0.05) indicated that the models using specific conductance as the explanatory variable performed better than those using streamflow. The addition of multiple explanatory variables improved the ability to estimate selenium concentration at several selenium sites over using solely streamflow or specific conductance. More seasonal explanatory variables were included in the models for GunnDelta and UncDelta using streamflow than the models using specific conductance. In contrast to all the other models in this report, the evaluation statistics indicated that models developed for GunnDelta and UncDelta using streamflow performed better than those using specific conductance. The

error associated with the log-transformed salinity and selenium estimates is consistent in log space; however, when the estimates are transformed into non-log values the error increases as the estimates decrease.

Continuous streamflow and specific conductance data collected at study sites provide the means to examine temporal variability in constituent concentration and load. The regression models can estimate continuous concentrations or loads on the basis of continuous specific conductance or streamflow data. Similar estimates are available for other sites at the USGS National Real-Time Water Quality Webpage (*http://nrtwq.usgs.gov*) and provide water-resource managers with a means of improving their general understanding of how constituent concentration or load can change annually, seasonally, or in real time.

References Cited

Bradu, D., and Mundlak, Y., 1970, Estimation in lognormal linear models: Journal of the American Statistical Association, v. 65, no. 329, p. 198–211.

Bureau of Reclamation, 2001, National Irrigation Water Quality Program brochure: Accessed July 13, 2012, at *http://www.usbr.gov/niwqp/*.

Bureau of Reclamation, 2011a, Dolores Project—Project details: Accessed May 21, 2012, at *http://www.usbr.gov/projects/Project.jsp?proj_Name=Dolores%20Project*.

Bureau of Reclamation, 2011b, Quality of water Colorado River Basin: Accessed March 21, 2012, at *http://www.usbr.gov/uc/progact/salinity/pdfs/PR23final.pdf*.

Butler, D.L., 1996, Trend analysis of selected water-quality data associated with salinity-control projects in the Grand Valley, in the lower Gunnison Basin, and at Meeker Dome, western Colorado: U.S. Geological Survey Water-Resources Investigations Report 95–4274, 38 p.

Butler, D.L., Krueger, R.P., Osmundson, B.C., Thompson, A.L., and McCall, S.K., 1991, Reconnaissance investigation of water quality, bottom sediment, and biota associated with irrigation drainage in the Gunnison and Uncompahgre River basins and at Sweitzer Lake, west-central Colorado, 1988–89: U.S. Geological Survey Water-Resources Investigations Report 91–4103, 99 p.

Butler, D.L., and Leib, K.J., 2002, Characterization of selenium in the Lower Gunnison River Basin, Colorado, 1988–2000: U.S. Geological Survey Water-Resources Investigations Report 02–4151, 26 p.

Butler, D.L., Wright, W.G., Stewart, K.C., Osmundson, B.C., Krueger, R.P., and Crabtree, D.W., 1996, Detailed study of selenium and other constituents in water, bottom sediment, soil, alfalfa, and biota associated with irrigation drainage in the Uncompahgre Project area and in the Grand Valley, west-central Colorado 1991–93: U.S. Geological Survey Water-Resources Investigations Report 96–4138, 136 p.

Colorado Department of Public Health and Environment, 2012, Colorado's section 303d list of impaired waters and monitoring and evaluation list: Accessed May 15, 2012, at *http://www.cdphe.state.co.us/regulations/wqccregs/93_2012(03).pdf*.

Colorado Water Conservation Board, 2010, Colorado Decision Support Systems—Irrigated parcels, Divisions 4, 5, 6, and 7: Accessed May 15, 2012, at *http://cdss.state.co.us/GIS/Pages/GISDataHome.aspx*.

Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, p. 605–610.

Hamilton, S.J., 1998, Selenium effects on endangered fish in the Colorado River Basin, *in* Frankenberger, W.T., and Engderg, R.A., eds., Environmental chemistry of selenium: New York, Marcel Dekker, p. 297–314.

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.

Hem, J.D., 1985, Study and interpretation of the chemical characteristics of natural water (3d ed.): U.S. Geological Survey Water-Supply Paper 2254, 264 p.

Homer, C., Huang, C., Yang, L., Wylie, B.K., and Coan, M., 2004, Development of a 2001 National Land Cover Database for the United States: Photogrammetric Engineering and Remote Sensing, v. 70, no. 7, p. 829–840.

Kenney, T.A., Gerner, S.J., Buto, S.G., and Spangler, L.E., 2009, Spatially referenced statistical assessment of dissolved-solids load sources and transport in streams of the Upper Colorado River Basin: U.S. Geological Survey Scientific Investigations Report 2009–5007, 50 p.

Leib, K.J., Linard, J.I., and Williams, C.A., 2012, Statistical relations of salt and selenium loads to geospatial characteristics of corresponding subbasins of the Colorado and Gunnison Rivers in Colorado: U.S. Geological Survey Scientific Investigations Report 2012–5003, 31 p.

Lemly, A.D., 1993, Guidelines for evaluating selenium data from aquatic monitoring and assessment studies: Environmental Monitoring and Assessment, v. 28, p. 83–100.

Lemly, A.D., 2002, Selenium assessment in aquatic ecosystems—A guide for hazard evaluation and water quality criteria: New York, Springer-Verlag, 161 p.

Liebermann, T.D., Mueller, D.K., Kircher, J.E., and Choquette, A.F., 1989, Characteristics and trends of streamflow and dissolved solids in the Upper Colorado River Basin, Arizona, Colorado, New Mexico, Utah, and Wyoming: U.S. Geological Survey Water-Supply Paper 2358, 64 p.

Mayo, J.W., and Leib, K.J., 2012, Flow-adjusted trends in dissolved selenium load and concentration in the Gunnison and Colorado Rivers near Grand Junction, Colorado, water years 1986–2008: U.S. Geological Survey Scientific Investigations Report 2012–5088, 33 p.

Ohlendorf, H.M., Hoffman, D.J., Saiki, M.K., and Aldrich, T.W., 1986, Embryonic mortality and abnormalities of aquatic birds—Apparent impacts of selenium from irrigation drainwater: Science of the Total Environment, v. 52, p. 49–63.

Prairie, J.R., Rajagopalan, B., Fulp, T.J., and Zagona, E.A., 2005, Statistical nonparametric model for natural salt estimation: Journal of Environmental Engineering, v. 131, no. 1, p. 130–138.

Presser, T.S., and Luoma, S.N., 2006, Forecasting selenium discharges to the San Francisco Bay-Delta estuary—Ecological effects of a proposed San Luis Drain extension: U.S. Geological Survey Professional Paper 1646, 196 p. (Also available at *http://pubs.usgs.gov/pp/p1646*.)

Presser, T.S., and Ohlendorf, H.M., 1987, Biogeochemical cycling of selenium in the San Joaquin Valley, California, USA: Environmental Management, v. 11, no. 6, p. 805–821.

Seiler, R.L., Skorupa, J.P., Naftz, D.L., and Nolan, B.T., 2003, Irrigation-induced contamination of water, sediment, and biota in the western United States—Synthesis of data from the National Irrigation Water Quality Program: U.S. Geological Survey Professional Paper 1655, 123 p.

Tuttle, M.L., and Grauch, R.I., 2009, Salinization of the upper Colorado River—Fingerprinting geologic salt sources: U.S. Geological Survey Scientific Investigations Report 2009–5072, 62 p.

U.S. Geological Survey, 1998, National Water Information System (NWIS): U.S. Geological Survey Fact Sheet 027–98, 2 p., accessed February 21, 2013, at *http://waterdata.usgs.gov*.

Warner, J.W., Heimes, F.J., and Middelburg, R.F, 1985, Ground-water contribution to the salinity of the Upper Colorado River Basin: U.S. Geological Survey Water-Resources Investigations Report 84–4198, 113 p.

Wright, W.G., and Butler, D.L., 1993, Distribution and mobilization of dissolved selenium in ground water of the irrigated Grand and Uncompahgre Valleys, western Colorado, *in* Allen, R.G., and Neale, C.M.U., eds., Management of irrigation and drainage systems—Integrated perspectives: American Society of Civil Engineers, Proceedings of the 1993 National Conference on Irrigation and Drainage Engineering, Park City, Utah, July 21–23, 1993: p. 770–777.

ISSN 2331-1258 (online)
http://dx.doi.org/10.3133/ofr20141015