Groundwater

Forecast of Natural Aquifer Discharge Using a Data-Driven, Statistical Approach

by Kevin G. Boggs[1], Rob Van Kirk[2], Gary S. Johnson[3], and Jerry P. Fairley[3]

Abstract

In the Western United States, demand for water is often out of balance with limited water supplies. This has led to extensive water rights conflict and litigation. A tool that can reliably forecast natural aquifer discharge months ahead of peak water demand could help water practitioners and managers by providing advanced knowledge of potential water-right mitigation requirements. The timing and magnitude of natural aquifer discharge from the Eastern Snake Plain Aquifer (ESPA) in southern Idaho is accurately forecast 4 months ahead of the peak water demand, which occurs annually in July. An ARIMA time-series model with exogenous predictors (ARIMAX model) was used to develop the forecast. The ARIMAX model fit to a set of training data was assessed using Akaike's information criterion to select the optimal model that forecasts aquifer discharge, given the previous year's discharge and values of the predictor variables. Model performance was assessed by application of the model to a validation subset of data. The Nash-Sutcliffe efficiency for model predictions made on the validation set was 0.57. The predictor variables used in our forecast represent the major recharge and discharge components of the ESPA water budget, including variables that reflect overall water supply and important aspects of water administration and management. Coefficients of variation on the regression coefficients for streamflow and irrigation diversions were all much less than 0.5, indicating that these variables are strong predictors. The model with the highest AIC weight included streamflow, two irrigation diversion variables, and storage.

Introduction and Purpose

Water-supply forecasts are important water-management tools, especially in semi-arid to arid regions, where water supply and demand are often out of phase. There are numerous types of forecast models. For example, forecasts have been generated using statistical models, mechanistic models, and calibrated numerical model simulations. Translating forecasts into predictions of natural aquifer discharge can be complicated and is not a commonly applied water-management practice. The desire to forecast groundwater elevations (Kaczmarek 1961; Zaltsberg 1982; Houston 1983; Coppola et al. 2005) and groundwater-surface water interactions (Zaltsberg 1987) is not a recent development. However, groundwater-flow regime forecasts have not typically provided accurate predictions of natural aquifer discharge months ahead of peak groundwater demand.

Numerical models such as MODLFOW (Harbaugh et al. 2000) are often used to forecast groundwater levels and groundwater discharge. The creation and calibration of numerical models can be expensive, and there is inherent uncertainty in model development that drives the need for generating alternative models to assess predictive uncertainty (Poeter 2007). We demonstrate that a simple time-series regression model can be used to quantify certain aspects of groundwater flow and to forecast aquifer discharge, even with restrictive assumptions. Compared to the cost of developing numerical models, statistical tools can often be implemented quickly and efficiently, providing less costly methods for quantifying groundwater flow characteristics while still providing useful results. This paper presents a procedure that can be used in April, to predict annual aquifer discharge at the beginning of irrigation season and months in advance of peak demand, which is typically in July when sufficient water is not always available to meet competing demands. The predictors used for the analysis include variables that could impact the groundwater budget, such as precipitation and irrigation-associated recharge, along with indicators of water supply, such as reservoir storage and snowpack. Our research objectives are (1) to determine whether natural discharge can be forecast using only data known at the time of the forecast, (2) to assess whether the forecast can be improved by using predictors that are not known at the

[1]Corresponding author: University of Idaho, Department of Geological Sciences, Boise, ID 83702; 208-340-1753; Kevin.Boggs@ch2m.com

[2]Henry's Fork Foundation, P.O. Box 550, Ashton, ID 83420 and Humboldt State University, Department of Mathematics, Arcata, CA 95521.

[3]University of Idaho, Department of Geological Sciences, Moscow, ID 83844.

The authors have no conflict of interest to disclose. The authors have no financial disclosure to submit.

Received March 2013, accepted September 2013.

doi: 10.1111/gwat.12133

time of the forecast but can potentially be forecast themselves, and (3) to assess whether analytically-derived, theoretical relationships between aquifer recharge and discharge are observed in recharge-discharge data.

Study Area

We apply our forecast procedure to the Box Canyon Spring, one spring in the Thousand Springs Complex, which discharges water from the Eastern Snake Plain Aquifer (ESPA) in southern Idaho (Figure 1). The ESPA is a critical water resource that it used to irrigate significant areas of land and to provide drinking water to over 300,000 people. Groundwater levels and natural discharge on the ESPA have steadily decreased since the 1950s (Figure 2), and water managers have spent significant resources over the last decade to develop an aquifer management plan (Idaho Water Resource Board 2008). The aquifer management plan includes a goal to increase discharge in the Thousand Springs Complex. To continue to manage competing interests, there is a need to develop accurate, advanced knowledge of aquifer discharge in the Thousand Springs area near Hagerman, Idaho, which is the primary discharge area for the ESPA (Figure 1). The forecasting procedure presented in this paper could help water managers and users by providing them with a simple, accurate, low-cost planning tool.

Regional Hydrogeologic Setting

The eastern Snake River Plain extends nearly 200 miles across southern Idaho (Figure 1). The ESPA is primarily formed of highly fractured, interfingered Quaternary basalt flows, with lenses of sediment between the flows (Smith 2004). More detailed descriptions of the geology of the eastern Snake River Plain are provided by Anderson (1991), Whitehead (1986), and Kuntz et al. (1992).

Approximately 69% of ESPA recharge comes from irrigation seepage (i.e., infiltration of excess irrigation water through the root zone), 23% from tributary basins, 9% from percolation of precipitation on the plain, and 8% from seepage from the Snake River itself, based on the State of Idaho's water budget for the ESPA numerical model for the years 1981 through 2008 (Idaho Department of Water Resources 2013). Average annual discharge exceeds average annual recharge: approximately 68% of annual recharge to the ESPA is discharged back to the Snake River, and about 28% is discharged through pumping to meet crop evapotranspiration (ET) requirements; the balance is accommodated by ET from wetlands, urban pumping, and change in aquifer storage. Average flow from the Thousand Springs Complex exceeds 5000 cubic feet per second (cfs; 140 m^3/s).

Response Variable

The USGS has reported flow from several springs in the Thousands Spring complex daily since the 1950s (http://waterdata.usgs.gov/id/nwis/sw/). We used discharge at Box Canyon Springs near Wendell, Idaho (USGS gauging station 13095500) as the response variable in our model. Average annual discharge from the spring has decreased steadily since the mid-1960s to early 1970s (Figure 2), which is consistent with the decreasing trend in the total discharge from the Thousand Springs Complex.

Predictor Variables

We chose predictor variables to represent the major recharge and discharge components of the ESPA water budget, including variables that reflect overall water

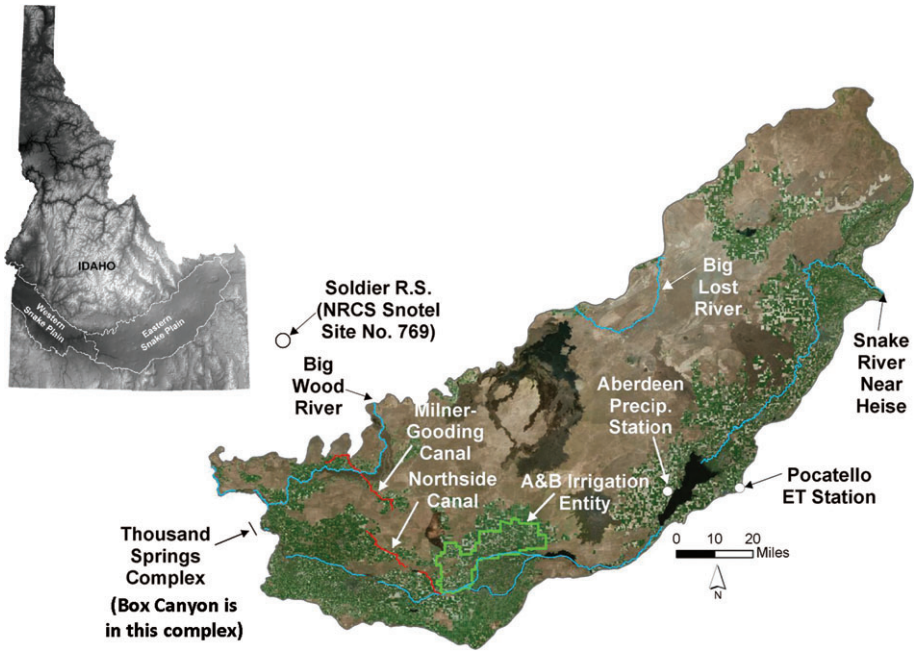

Figure 1. Location of the Eastern Snake Plain Aquifer.

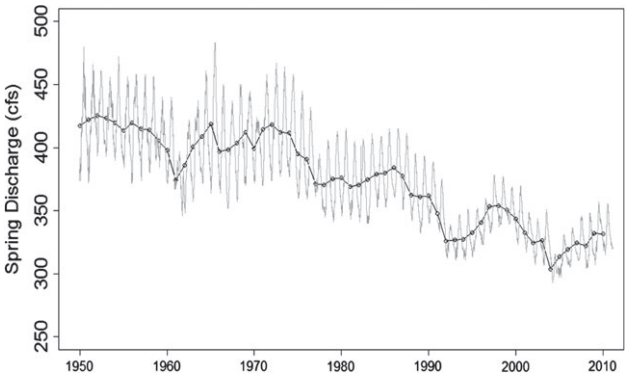

Figure 2. Average annual (solid black line) and daily (solid gray line) Box Canyon Spring (USGS 13095500) discharge (April to April water year) in cubic feet per second (cfs). Data source: http://waterdata.usgs.gov/usa/nwis/uv?13095500

supply and important aspects of water administration and management. Previous analytical work suggests that we need to focus on recharge variability close to the Thousand Springs Complex to predict its discharge (Boggs et al. 2010). There are numerous tributary streams that contribute recharge to the ESPA via underflow through alluvial valley fill. Flow in many of these tributary streams is ungauged, and many of the streams are far from the Thousand Springs Complex. We use the Big Wood and Big Lost rivers to provide proxies for tributary basin underflow to the ESPA and irrigation diversions for the areas because they have been gauged since 1915 and 1912, respectively, and because they are significant tributaries, in terms of recharge. Precipitation at the National Weather Service Aberdeen Experiment Station (100010) was selected because of its period of record (1939 to present) and central location on the ESPA. The A&B Irrigation District uses hundreds of wells to pump groundwater for irrigation purposes and has reported total annual pumping values since the district began pumping in 1960, providing the most readily available pumping data on the ESPA. Therefore, these pumping data were selected as a proxy for pumping on the ESPA as a whole. The canal systems representing the irrigation delivery water budget item were selected because of their proximity to the Box Canyon Spring and because of their long period of record.

Water diverted into canals on the ESPA comes from one of two sources: surface water diversions (natural flow) and reservoirs. Two of the canals, Milner-Gooding and Big Wood, divert primarily natural flow, whereas the Northside canal diverts primarily storage water. Because total irrigation seepage is the difference between diversion and consumptive use, we included as a predictor the alfalfa reference ET at Pocatello (central location on the plain, data obtained from Allen and Robison 2009).

Irrigation application methods could also affect net recharge due to irrigation. Conversion from flood to sprinkler application occurred steadily on the ESPA from the 1950s through the 1990s, and many water users and managers on the ESPA contend that water-level and aquifer-discharge declines on the ESPA can be attributed to the conversion from flood to sprinkler irrigation. Contor (2004) performed an extensive review of sprinkler- and gravity-irrigated land area for an ESPA numerical modeling effort. We included fraction of irrigated land under sprinkler irrigation as a predictor via an empirical fit to Contor's (2004) data similar to that used by Boggs et al. (2010).

Snow water equivalent (SWE) data from the NRCS Soldier Ranger Station (SNOTEL site 769), north of the ESPA, was used to represent overall water supply in the basin because of its period of record and its proximity to the ESPA (http://www3.wcc.nrcs.usda.gov/nwcc/site?sitenum=769&state=id). Much of the water that falls on the mountains surrounding the ESPA is stored in surface-water reservoirs. To represent storage, we used the sum of reservoir storage in Jackson Lake and Palisades Reservoir, which are located upstream of all points of diversion and together account for half of the total storage capacity in the Snake River system above the Thousand Springs Complex.

Recharge and Discharge Water Year Convention

Theory indicates that when upgradient boundaries have little effect on discharge, and recharge is periodic and uniformly distributed across the domain, discharge is also periodic, with period equal to that of recharge, but lagged by 1/8 period (Townley 1995). The ESPA more or less satisfies these conditions, with recharge of annual period (Boggs et al. 2010), so discharge should lag recharge by about 45 d. Over the common period of record, the mean date of minimum in the annual cycles of both irrigation infiltration and streamflow (the two largest sources of recharge to the ESPA) was February 25. The mean date of minimum in annual cycles of discharge at Box Canyon Spring was about 45 d later, April 13 (Figure 3), consistent with theory. Thus, we define the recharge water year as February 25 to February 24 and the discharge water year as April 13 to April 12. Our water year is designated by the calendar year in which it starts. The year starting February 25, 1999, for example, is called the "1999" recharge water year, and the corresponding discharge water year would start on April 13, 1999. Using the minimum value to define recharge and discharge water years not only aligns phase lag in annual cycles with the analytical theory but also keeps seasonal recharge and its theoretical effects in the same year as discharge. The traditional October-to-September water year or November-through-October irrigation year splits the effect of seasonal recharge between two different water years. Administratively, the irrigation season in the upper Snake River system begins on April 1, and in most years, the onset of physical irrigation ranges from April 1 at the lowest elevations to June 1 at the highest elevations. Therefore, we generate our forecast in early April, which coincides with the beginning of irrigation season and the beginning of the discharge water year.

The application of the water year convention to any given variable depended on the physical nature of the variable and its data resolution (Table 1). Because the

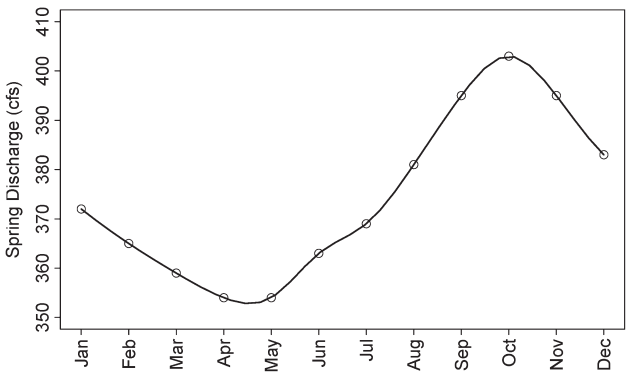

Figure 3. Average monthly Box Canyon Spring discharge (USGS Station 13095500) for the period of record (1950 through 2013).

aquifer discharge data were available on a daily basis, we summed daily values over the period April 13 to April 12 to obtain annual values for the response variable. For the streamflow, diversion, and pumping variables, we summed daily values over the period February 25 to February 24 to obtain annual values, except where data were available only on a monthly or seasonal basis, in which case we summed monthly values for March through February to obtain the annual values. Although direct precipitation over the ESPA is nearly uniformly distributed throughout the year, essentially all precipitation that falls during the growing season is lost to ET. Only precipitation that falls in the winter months, the majority of which is in the form of snow, recharges the aquifer. Using our recharge water year convention, precipitation in November through February does not contribute to spring discharge until the following year. Therefore, for the Aberdeen precipitation variable, the value for water year 2000, for example, is the sum of monthly precipitation over November and December 1999 and January through March 2000. Similarly, mountain snowpack accumulated over the winter of 1999–2000 would not contribute to water supply until the 2000 discharge year, so we used SWE on April 1, 2000, for example, as the value for water year 2000. The annual values for percentage of irrigation applied with sprinklers and for reservoir storage were also taken to be their point-in-time values on April 1 to coincide with the start of the new discharge year and the timing of the forecast. Thus, the fraction of irrigation applied with sprinklers and the values of the water-supply variables (precipitation, SWE, reservoir storage) for the upcoming discharge year are known at the time of the forecast. However, the values of variables representing the largest components of the aquifer budget (streamflow, irrigation diversion, crop ET, and groundwater pumping) for the upcoming discharge year would not be known at the time of the forecast.

Statistical Methods

We used time-series regression models in which the predictors are potentially lagged in time and averaged over several water years to reflect that aquifer recharge is lagged in time and attenuated in magnitude before it is discharged from the aquifer (Townley 1995; Knight et al. 2005; Criss and Winston 2008; Boggs et al. 2010). We divided the data into calibration (1950 through 1999) and validation (2000 through 2010) sets and used Akaike's information criterion (AIC; Akaike 1973; Burnham and Anderson 2002; Anderson 2008) to select the optimal model to predict aquifer discharge. Model performance was assessed by fitting the model to the calibration set and applying it to the validation set. Final model parameters were estimated by fitting the model to standardized variables, using the entire time series of data.

Time-Series Regression Model

As the response variable occurs in a time series, we used a modification of standard linear regression to account for temporal autocorrelation. In this case, time series diagnostics indicated that correlation between successive observations of discharge was near 1 but that

Table 1
Potential Predictor Variables Used in Regression Models (t Indicates Calendar Year)

Variable	Description	Definition	Year t Value Known at Time of Year t Forecast
BLRiver	Big Lost River Discharge	25 Feb(t) to 24 Feb($t+1$) total	No
BWDiv	Big Wood Canal Diversion	25 Feb(t) to 24 Feb($t+1$) total	No
BWRiver	Big Wood River Discharge	25 Feb(t) to 24 Feb($t+1$) total	No
ET	Reference ET on ESPA (Pocatello)	Irrigation season(t) total	No
MGDiv	Milner-Gooding Canal Diversion	25 Feb(t) to 24 Feb($t+1$) total	No
NSDiv	Northside Canal Diversion	25 Feb(t) to 24 Feb($t+1$) total	No
Precip	Precipitation on ESPA (Aberdeen)	Nov($t-1$) to March(t) Total	Yes
Pump	A&B Irrig. District GW pumping	Irrigation season(t) total	No
Sprinkler	Fraction of land in sprinkler irrigation	1 April(t) value (fit to Contor 2004 data)	Yes
Stor	Reservoir Storage (Palisades+Jackson)	1 April(t) value	Yes
SWE	Soldier Mt. Snow Water Equivalent	1 April(t) value	Yes

Note: Variables are listed in alphabetical order to facilitate reference.

annual changes in discharge were uncorrelated. Thus, the assumption of independence required by regression analysis was met by using a first-order integrated time series model with exogenous predictors (ARIMAX(0,1,0), Shumway and Stoffer 2011). This model has the form

$$y(t) - y(t-1) = \sum_{k=1}^{p} \beta_k \left[x_k(t) - x_k(t-1)\right] + \epsilon_t, \quad (1)$$

where $y(t)$ is the discharge in year t, $x_1(t), x_2(t), \ldots, x_p(t)$ are predictors, β_k is the coefficient on predictor k and ϵ_t is an independent and identically distributed normal random variable. Defining the annual increments in response and predictor, respectively, by $\Delta y(t) = y(t) - y(t-1)$ and $\Delta x_k(t) = x_k - x_k(t-1)$, Equation 1 reduces to

$$\Delta y(t) = \sum_{k=1}^{p} \beta_k \, \Delta x_k(t) + \epsilon_t, \quad (2)$$

which expresses annual increments in the response as a linear function of annual increments in the predictors. Equation 2 is equivalent to standard linear regression of Δy on Δx_k except that there is no intercept term. The "null" model (no predictors) is a simple random walk, and if the only predictor is time itself, the model represents a random walk with constant trend (Shumway and Stoffer 2011).

After calculating the estimated regression coefficients $\widehat{\beta}_k$ by maximum likelihood, the predicted value of the response is

$$\widehat{y}(t) = y(t-1) + \sum_{k=1}^{p} \widehat{\beta}_k \left(x_k(t) - x_k(t-1)\right). \quad (3)$$

Equation 3 is known as a "one-step-ahead" prediction equation, which predicts the value of the response in year t, given the value $y(t-1)$ of the response in the previous year and the values of the predictors $x_k(t)$ in the current and previous years.

To accommodate potential attenuation and lag in the statistical model, we considered the possibility that a predictor variable $x_k(t)$ included in regression model (1) is a simple arithmetic moving average of window width j_k of an observed quantity $z_k(t)$ and is also possibly lagged by i_k time units. The resulting increment in the predictor reduces to

$$\Delta x_k(t) = \frac{1}{j_k}\left[z_k(t - i_k) - z_k(t - j_k - i_k)\right], \quad (4)$$

which is the mean j-year trend in the quantity z_k, lagged i time steps. We selected i_k and j_k for each quantity z_k by choosing i_k and j_k to maximize the correlation between $y(t) - y(t-1)$ and $\frac{1}{j_k}\left[z_k(t - i_k) - z_k(t - j_k - i_k)\right]$. Once the optimal moving-average window j and lag i were determined for each predictor, Equation 4 was used to compute the values of the variable $\Delta x_k(t)$ used in Equation 2.

As an illustration of an averaged and lagged variable, consider annual precipitation. Annual increment in Box Canyon discharge was most strongly correlated with increment in annual precipitation at lag 1 and moving-average window width 5. Thus, for example, the value of the annual precipitation variable for water year 2000 was one fifth of the difference between precipitation in water year 1999 and precipitation in water year 1994. This quantity is the mean annual trend in precipitation over the 1994–1999 period, and its effect on discharge was reflected most strongly in water year 2000. An optimal correlation between discharge and recharge variables that occurs at a moving-average window greater than 1 is the statistical realization of attenuation in the effect of a recharge variable on discharge. Because the time scale of attenuation varies with spatial and temporal properties of the recharge (Boggs et al. 2010), we expected that some moving-average windows would result in higher correlations with discharge than others.

Effects of Information Unknown at the Time of Forecast

Ideally, we would like to be able to predict discharge for the upcoming year with knowledge of only the previous year's discharge and values of the predictors known at the time of the forecast ("historic data"). This would require all predictors to have a lag of at least 1, excepting those such as SWE whose current-year value is known at the time of the forecast (Table 1). In this modeling procedure, which we call "Model 1," we used only information about the predictor variables that would be available at the time of the early-April forecast. This required some of the predictors to be lagged at least one year, even when the optimal lag from the correlation analysis was 0. However, analytical models indicate that annual periodic recharge sources relatively close to the Thousand Springs Complex (approximately 32 miles; 50 km) impact spring discharge in the same year the recharge occurs (Boggs et al. 2010). Thus, we performed a second model-fitting and validation procedure using some information that would not be known at the time of the forecast. In this procedure, the optimal lag was used for all predictors, even if lag-0 values would not be known at the time of the forecast. In particular, this allowed inclusion of diversion, streamflow, pumping, and ET variables for the upcoming water year, which are not known at the time of the forecast. We refer to this procedure as "Model 2," which would be useful for forecasting only if the predictors unknown at the forecast time were themselves forecast. However, this model is useful for assessing whether recharge-discharge relationships predicted by analytical models are reflected in observed data.

Candidate Models

We proposed a priori a set of candidate models (Table 2), based on the water budget and results of analytical models (Boggs et al. 2010). We used the same set of candidate models in both the Model 1 and Model 2 procedures. Thus, the only difference between the two procedures was whether or not the values of all predictors would be known at the time of the forecast. Several

Table 2
Predictors Used in Candidate Models

Model	Predictor(s)							
	BWRiver	ET	MGDiv	NSDiv	Precip	Pump	Stor	SWE
A	•	•	•	•	•	•	•	•
B	**No predictors (null model, random walk)**							
C	**Temporal trend only (random walk with drift)**							
D	•	•	•	•	•	•	•	
E	•		•	•	•	•	•	•
F	•	•	•	•	•	•		•
G	•	•	•	•	•		•	•
H	•	•	•	•		•	•	•
I	•		•	•				
J	•			•				
K	•		•					
L			•	•				
M	•		•	•				•
N	•	•	•	•				
O	•		•	•			•	
P	•		•	•		•		
Q	•		•	•	•			
R	•		•	•			•	•
S	•	•	•	•			•	•
T	•		•	•		•	•	•
U	•		•	•	•		•	•

of the original predictors considered were eliminated from inclusion in candidate-set models based on results of the correlation analysis, including Big Wood Canal diversions (correlations were low compared to other diversion variables; two other canal diversion variables are included in the analysis), Big Lost River discharge (highly correlated with Big Wood River discharge and farther from the Thousand Springs area), and conversion from flood to sprinkler irrigation (correlations were extremely low; in addition, the conversion from flood to sprinkler conversion occurred rapidly compared to our forecasting period). This left eight potential predictor variables were Big Wood River discharge (BWRiver), Northside (NSDiv) and Milner-Gooding (MGDiv) canal diversions, A&B District pumping (Pump), Soldier Mountain SWE (SWE), Aberdeen precipitation (Precip), reservoir storage (Stor), and calculated reference ET (ET). Because they make up the two largest recharge components of the ESPA water budget, Big Wood River discharge and at least one of the canal diversion variables were included in each candidate model with two exceptions. We included a null model (random walk) and a simple trend model (random walk with drift) for comparison purposes.

Model Selection

Once a set of candidate models was defined, we used AIC (Akaike 1973; Burnham and Anderson 2002), for model discrimination. In the case of normally distributed errors,

$$\mathrm{AIC} = n\left[\log\left(\widehat{\sigma}^2\right) + 1 + \log(2\pi)\right] + 2K,$$

where n is the sample size, $\widehat{\sigma}^2$ is the maximum-likelihood estimate of residual variance (residual sum-of-squares divided by n), and K is the total number of parameters in the model, including the intercept and $\widehat{\sigma}^2$. The AIC is an approximately unbiased estimate of Kullback-Leibler information (Kullback and Leibler 1951) of a fitted model (Hurvich and Tsai 1989). When $\frac{n}{K} < 40$, AIC becomes negatively biased (Hurvich and Tsai 1989), so we used the small-sample corrected version of AIC (AICc), given by

$$\mathrm{AIC_c} = n\left[\log\left(\sigma^2\right) + 1 + \log(2\pi)\right] + 2K + \frac{2K(K+1)}{n-K-1} \tag{6}$$

The AICc and ΔAICc were computed for each of the models in the candidate set, where ΔAICc is the difference in AICc values between each model and the model with the lowest AICc value (Sugiura 1978; Hurvich and Tsai 1989; Burnham and Anderson 2002). These ΔAICc values were used to compute the relative evidence weight w_i for each model, given by

$$w_i = \frac{\exp\left(-\frac{\Delta\mathrm{AICc_i}}{2}\right)}{\sum_j \exp\left(-\frac{\Delta\mathrm{AICc}_j}{2}\right)}, \tag{7}$$

where $\Delta\mathrm{AICc}_i$ is the ΔAICc value for model i and the sum in the denominator is taken over all models j in the candidate set (Burnham and Anderson 2002; Anderson

2008). The models are then ranked in decreasing order of model weight. The relative evidence weight can be interpreted as the probability that the model is the best among the candidate set of models. Prior to carrying out the model selection procedure, we analyzed residuals from the model that contained all potential predictors, as recommended by Burnham and Anderson (2002). We also analyzed residuals from the top model after selection with AIC. In both models, residuals were slightly left-skewed. However, departure from normality was not large enough to affect overall results, given the sample size of 50 years in the calibration set.

Results

Lags and Moving-Average Windows

We generated a matrix of correlations between the response and each predictor at each combination of lag times ranging from 0 to 10 years and moving-average windows ranging from 1 to 10 years (Figure 4). Generally, the strongest correlations of appropriate sign (that is, positive with respect to recharge, negative with respect to discharge such as pumping) occurred at a lag of 0. Exceptions were BWDiv, Precip, and Sprinkler (Table 3). In general, the moving-average windows that resulted in the strongest correlations were longer for climate-related variables such as tributary flow, SWE, precipitation, and ET and shorter for variables driven by water management and use such as diversions, pumping, and storage.

Model 1: Information Known at Time of Forecast

Five models accounted for a cumulative weight of 0.94 (Table 4), as fit to the training subset. In the discussion section, we interpret predictive ability of the various explanatory variables based on relative weights and particular combinations of predictors in these models. Here, however, we assess model performance by using the single model with the lowest AICc score to calculate one-step-ahead predictions of Box Canyon Spring discharge for years 2000 through 2010 (the "validation" set, Figure 5). The Nash-Sutcliffe efficiency (Nash and Sutcliffe 1970) for the model as fit to the training set was 0.95, but the Nash-Sutcliffe efficiency for predictions made on the validation set was only 0.29. The Nash-Sutcliffe efficiency is the generalization of the standard R^2 measure in linear regression modeling. The two measures are identical for linear models, in which their common value is the fraction of total sum-of-squares in the response variable explained by the regression model. For nonlinear models, including ARIMA time series models, the total sum-of-squares is not equal to the sum of residual and model sum-of-squares, so the Nash-Sutcliffe efficiency may be negative. However, it is always less than 1, and thus when its value is positive, it can be roughly interpreted in the same way as R^2, namely as the fraction of variability in the response variable explained by the model. The final parameters and standard errors for Model 1 are reported in Table 5.

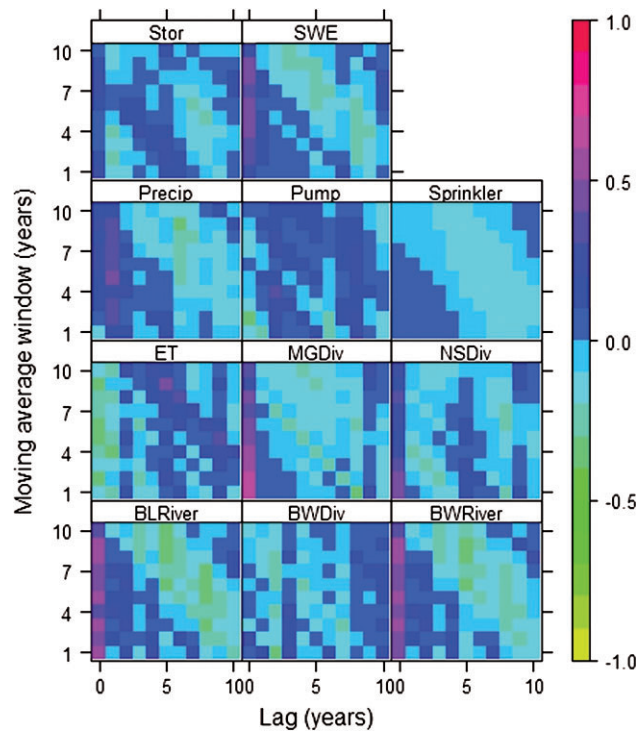

Figure 4. Correlation between Box Canyon Spring discharge and the predictors as a function of lag time and moving-average window.

Table 3
Predictor-Variable Lag Time and Moving-Average Window (MAW) That Produced Optimum Correlation with Box Canyon Spring Discharge

Predictor	Lag	MAW	Correlation of Maximum Absolute Value and 95% Confidence Intervals
BLRiver	0	8	0.54 ± 0.21
BWDiv	3	4	0.25 ± 0.25
BWRiver	0	5	0.55 ± 0.21
ET	0	5	-0.38 ± 0.25
MGDiv	0	2	0.59 ± 0.20
NSDiv	0	1	0.45 ± 0.23
Precip	1	5	0.42 ± 0.23
Pump	0	2	-0.30 ± 0.22
Sprinkler	8	2	-0.20 ± 0.23
Stor	0	1	0.26 ± 0.25
SWE	0	9	0.46 ± 0.23

Note: Predictors are listed in alphabetical order to facilitate reference.

Model 2: Unknown Information Included

For Model 2, we chose to use a slightly different procedure than that used for Model 1 for selecting the lag and moving-average window for each variable. For Model 2, we chose the lag and moving-average window for each variable by starting with the lag that produced the highest correlation with Box Canyon Spring (this was lag 0 for all predictors except direct precipitation, which had optimal correlation at lag 1). Within this optimum

Table 4
Model Comparison for Model 1, After Removing Models Containing Pretending Variables and Models Accounting for Less Than 1% of Model Weight

Model and Predictors	K	log(L)	AICc	ΔAICc	w_i	$\sum w_i$
U: BWRiver,MGDiv,NSDiv,Precip,Stor,SWE	7	−484.6	985.9	0.00	0.34	0.34
E: BWRiver,MGDiv,NSDiv,Precip,Pump,Stor,SWE	8	−483.4	986.2	0.36	0.28	0.62
F: BWRiver,ET,MGDiv,NSDiv,Precip,Pump,SWE	8	−484.3	988.1	2.26	0.11	0.73
A: BWRiver,ET,MGDiv,NSDiv,Precip,Pump,Stor,SWE	9	−482.9	988.2	2.36	0.10	0.83
G: BWRiver,ET,MGDiv,NSDiv,Precip,Stor,SWE	8	−484.4	988.3	2.42	0.10	0.94

K, number of fitted parameters; log(L), log-likelihood; AICc, small-sample Akaike's Information Criterion; w_i, model weight; $\sum w_i$, cumulative model weight.

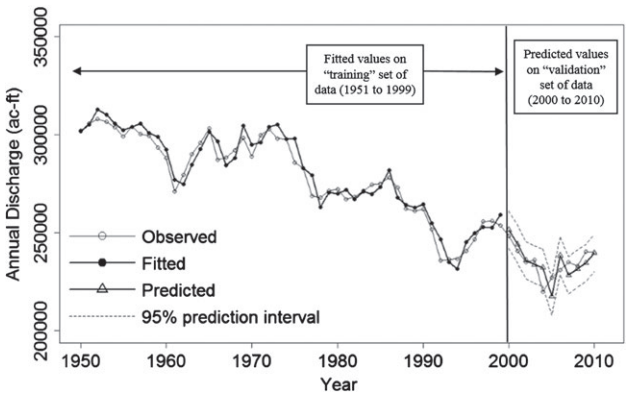

Figure 5. Observed and predicted (Model 1) Box Canyon Spring discharge for the "validation" set of data (years 2000 through 2010); fitted values for the "training" set of data (1951 through 1999) are also shown.

Table 5
Final Parameter Estimates (Based on Standardized Variables) and Standard Errors for Model 1

Predictor	Coefficient Estimate (−)	Standard Error (−)	Coefficient of Variation
BWRiver	−0.012	0.041	−3.417
MGDiv	0.428	0.152	0.355
NSDiv	0.241	0.092	0.382
Precip	0.155	0.050	0.323
Stor	0.066	0.030	0.455
SWE	0.238	0.080	0.336

Note: Coefficient estimates reported in this table were generated using dimensionless response and predictor variables obtained by subtracting the mean and dividing by the standard deviation.

lag, we chose the shortest moving-average window that (1) produced a correlation that was significantly different than 0 (of correct sign), and (2) produced a correlation that was not significantly different from the strongest correlation produced by any moving-average window at that particular lag. This procedure produces more parsimonious models, while accounting for uncertainty in the correlation estimates. Specifically, among moving-average windows that produced correlations that did not different significantly from one another, we choose the shortest moving-average window. Using this procedure, we used a moving-average window of 1 for variables BWRiver, MGDiv, NSDiv, Stor, and SWE, and a moving-average window of 2 for variables ET, Precip, and Pump.

The residuals from the model with the lowest AICc score, as fit to the training set, had first-order autocorrelation, indicating that our procedure for calculating predictions should change accordingly. Therefore, for Model 2, we modified our model from an ARIMAX(0,1,0) model to an ARIMAX(1,1,0) model, as given by

$$y_t = y_{t-1} + \sum_{k=1}^{p} \beta_k \left(x_k(t) - x_k(t-1)\right) + \alpha \left[(y_{t-1} - y_{t-2}) - \sum_{k=1}^{p} \beta_k \left(x_k(t-1) - x_k(t-2)\right) \right] \quad (8)$$

where α is the first-order autocorrelation coefficient, and all other variables are the same as in Equation 1. Six models accounted for a cumulative weight of 0.83 (Table 6), as fit to the training subset. Using Equation 8 and the model with the lowest AICc score, we calculated two-step-ahead predictions of Box Canyon Spring discharge for years 2000 through 2010 (the "validation" set, Figure 6), again deferring interpretation of predictive ability of the various explanatory variables to the discussion section. The Nash-Sutcliffe efficiency for the model as fit to the training set was 0.96, and the Nash-Sutcliffe efficiency for model predictions made on the validation set was 0.57, compared with 0.29 for Model 1. We calculated final parameter estimates and unconditional standard errors for Model 2 (Table 7).

Discussion

Predictors that reflect environmental conditions either (1) recharge the ESPA at a relatively large distance from the spring discharge location (e.g., SWE), or (2) recharge relatively uniformly over a large area (e.g., precipitation). The fact that these environmental predictors generally have stronger correlations with spring discharge at larger moving-average windows compared to predictors that are reflective of irrigation impacts is consistent

Table 6
Model Comparison for Model 2, After Removing Models Accounting for Less Than 1% of Model Weight

Model and Predictors	K	log(L)	AICc	ΔAICc	w_i	$\sum w_i$
O: BWRiver,MGDiv,NSDiv,Stor	6	−478.0	970.1	0.00	0.37	0.37
S: BWRiver,MGDiv,NSDiv,ET,Stor,SWE	8	−476.4	972.3	2.20	0.12	0.49
R: BWRiver,MGDiv,NSDiv,Stor,SWE	7	−477.9	972.5	2.46	0.11	0.60
D: BWRiver,MGDiv,NSDiv,ET,Precip,Pump,Stor	9	−475.3	973.0	2.98	0.08	0.69
T: BWRiver,MGDiv,NSDiv,Pump,Stor,SWE	8	−476.8	973.1	3.07	0.08	0.77
H: BWRiver,MGDiv,NSDiv,ET,Pump,Stor,SWE	9	−475.5	973.5	3.43	0.07	0.83

K, number of fitted parameters; log(L), log-likelihood; AICc, small-sample Akaike's Information Criterion; w_i, model weight; $\sum w_i$, cumulative model weight.

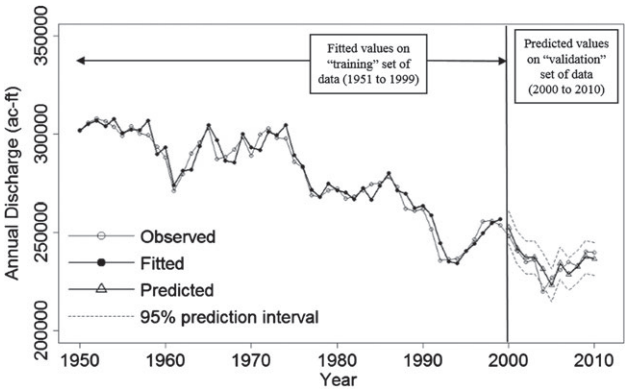

Figure 6. Observed and predicted (Model 2) Box Canyon Spring discharge for the "validation" set of data (years 2000 through 2010); fitted values for the "training" set of data (1951 through 1999) are also shown.

Table 7
Final Parameter Estimates (Based on Standardized Variables) and Standard Errors for Model 2

Predictor	Coefficient Estimate (−)	Standard Error (−)	Coefficient of Variation
Autocorrelation term	0.271	0.128	0.472
BWRiver	0.086	0.016	0.186
MGDiv	0.096	0.024	0.250
NSDiv	0.080	0.033	0.413
Stor	0.100	0.025	0.250

Note: Coefficient estimates reported in this table were generated using dimensionless response and predictor variables obtained by subtracting the mean and dividing by the standard deviation.

with analytical model results, which indicate that (1) there is a direct relationship between the duration of spring discharge impact and the distance between the recharge and discharge location, (2) ESPA recharge from sources extending over the entire aquifer affect the spring discharge for more than one year, and (3) impacts of ESPA recharge associated with the irrigation season near the Thousand Springs area contribute to spring discharge variability for no more than one water year (Boggs et al. 2010).

Model 2 had much greater predictive capability than Model 1, and model coefficients for the strongest predictors were more precisely estimated in Model 2 (Tables 5 and 7). However, of the four variables—BWRiver, MGDiv, NSDiv, and Stor—that appeared in all of the top models in the second modeling procedure, only the value of carryover reservoir storage (Stor) is known at the time of the early-April forecast. Therefore, a useful forecast of tributary discharge and canal diversions for the upcoming runoff/irrigation season is needed to use Model 2 to predict spring discharge for the upcoming year.

The AICc model ranking is an effective way to evaluate the relative strength of predictors. Focusing on Model 2, it is clear that BWRiver, NSDiv and MGDiv, and Stor have strong predictive capability. These variables represent ESPA recharge in the form of flow entering the aquifer from tributary basins, ESPA recharge associated with irrigation, and water conditions from the year prior to the forecast, respectively. Recharge associated with two different irrigation entities (Northside and Milner-Gooding) are represented in the models. Using recharge associated with both irrigation entities leads to a more efficient forecast of spring discharge than either one alone, reflecting the difference in water rights for each. The Northside irrigation entity relies heavily on storage and the Milner-Gooding irrigation entity relies primarily on natural flow. However, when only two of the variables BWRiver, MGDiv and NSDiv were included without the third (Models J, K, and L), model weights were essentially zero, indicating that these three variables have high predictive power only when they appear together. It is also important to observe that when these three are the only predictors used (Model I), model weight is also zero, indicating that additional information beyond that contained in these three variables is needed.

Among variables that provide additional information, Stor is the only predictor present in each of the top ten models in Model 2, which account for 98% of the model weight. The storage variable primarily reflects carryover storage from the previous year (thus supply for the upcoming year), because Palisades Reservoir and Jackson Lake fill after April 1 from runoff, whereas reservoirs in the lower reaches of the ESPA fill during the winter and are drafted consistently each year. Apparently, the storage variable reflects overall water conditions from the previous year (the year prior to the forecast) in ways that

the other predictors do not. SWE (variable SWE) appears to be important as well, as it appears in the second and third best model of Model 2. The SWE variable reflects water availability for the forecast year. The cumulative AICc weight of models containing either or both of the predictors Stor and SWE was 0.6, whereas the cumulative AICc weight of models containing the other three variables used in our analysis was 0.2, 0.3, and 0.4 for Precip, Pump, and ET, respectively. The variable Precip has relatively poor predictive ability compared with Stor and SWE because the majority of water supply to the ESPA originates as snowpack rather than as precipitation directly over the aquifer. Furthermore, water supply and availability, as reflected by Stor and/or SWE, are more important predictors than the water-use variables Pump and ET.

The regression coefficient for BWRiver was negative and weak in Model 1, counter to expectations. In Model 2, the regression coefficient was positive and very strong. In fact, BWRiver was the strongest predictor in Model 2 (CV = 0.186, the lowest coefficient of variation of any regression coefficient). These statistical observations are consistent with analytical theory, which indicates recharge occurring within a distance of approximately 50 km of Thousand Springs impacts the annual variation in the spring discharge in the same year the recharge occurs (Boggs et al. 2010).

As with any model, there are limitations in applying the statistical forecast of annual Box Canyon Spring discharge. We utilized historic data to develop our time series regression coefficients, and it is clear that current conditions on the ESPA are not consistent with historic conditions, as reflected by nonstationarity in the time series. For example, since the 1950s, water use and management on the ESPA has changed for a variety of reasons, including changes in land use, irrigation practices, and in the flow regime of the Snake River system associated with climate conditions. The lack of stationarity in natural systems has raised concerns over the ability to effectively manage water resources (Milly et al. 2008). We believe these concerns are valid, and that caution and professional judgment must be used when applying time series regression techniques in the face of increasing climatic variability.

The forecasting approach described in this paper may work well for aquifers where there is a sufficient record of aquifer discharge. Data on recharge components is available for most aquifers. However, for many aquifers, there is not a clearly defined discharge point at which flow data are available. For aquifers where the interaction between surface water and groundwater is important, this forecasting procedure is appropriate and applicable. The discharge forecasting procedure may not be appropriate where hydraulic heads are more important than discharge.

Conclusions

We developed a procedure to predict annual aquifer discharge in April, months in advance of the peak irrigation season. Using only information known at the time of the forecast (Model 1 approach) resulted in a model that had excellent ability to match historic data (Nash-Sutcliffe efficiency of 0.95), but only moderately weak predictive capability (Figure 5), with a Nash-Sutcliffe efficiency of 0.29 for predictions on the validation set of data. For Model 1, the cumulative AICc weight was over 0.95 for models containing BWRiver, MGDiv, NSDiv, Precip, and SWE as predictors, although this is not unexpected for the river flow and diversion variables because all except two models contained at least one of these predictors. Only three other predictors were used (Stor, Pump, and ET), and the cumulative AICc weight for models containing these variables was 0.87, 0.53, and 0.35, respectively. Coefficients of variation on the parameter estimates indicated that MGDiv, NSDiv, Precip, SWE, and Stor were the most precisely estimated (CV around 0.5 or lower). Based on coefficients of variation, BWRiver is a weak predictor in Model 1. The "best" model (highest AICc weight) using the Model 1 approach was a model that included all potential predictors except Pump and ET (Table 4).

We developed a second model, one that not only used information known at the time of the forecast but that also incorporated annual values of stream discharge, irrigation diversion, ET, and pumping for the upcoming year (Model 2). Using this approach, the Nash-Sutcliffe efficiency remained high for the model fit to historic data (0.96), but increased to 0.57 for predictions on the validation set of data, nearly a twofold increase compared to the Model 1 approach. The fact that Model 2 explained nearly twice as much variability in discharge verifies analytical findings, namely that recharge from irrigation close to the discharge point has a large effect in the same water year (Boggs et al. 2010).

We included null and temporal trend models in our analysis. In both Model 1 and Model 2, the null model performed better than the trend model, and almost all candidate models performed better than the null and trend models. These results provide statistical evidence that spring discharge does in fact respond to the external predictors on an annual time scale and that annual variability is neither random nor well described by a simple downward trend. Therefore, we conclude that not only can aquifer discharge be reasonably well predicted with knowledge of recharge variables but also that discharge can and does respond to recharge on time scales relevant to water management and planning.

Acknowledgments

The final manuscript was improved based on suggestions of the Editor-in-Chief (Frank W. Schwartz), an associate editor, and one anonymous reviewer. Their reviews are greatly appreciated.

References

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, ed. B.N. Petrov, 267–281. Budapest, Hungary: Akademia Kiado.

Allen, R.G., and C.W. Robison. 2009. *Evapotranspiration and consumptive irrigation water requirements for Idaho: Supplement updating the time series through December 2008, Research Technical Completion Report*. Moscow, Idaho: Kimberly Research and Extension Center, University of Idaho. http://data.kimberly.uidaho.edu/ETIdaho/ (accessed November 9, 2012).

Anderson, D.R. 2008. *Model Based Inference in the Life Sciences*. New York: Springer.

Anderson, S.R. 1991. Stratigraphy of the unsaturated zone and uppermost part of the Snake River Plain aquifer at the Idaho Chemical Processing Plant and Test Reactors Area, Idaho National Engineering Laboratory, Idaho. U.S. Geological Survey Water-Resources Investigations Report 91-4010. 71 p.

Boggs, K.G., R.W. Van Kirk, G.S. Johnson, J.P. Fairley, and P.S. Porter. 2010. Analytical solutions to the linearized Boussinesq equation for assessing the effects of recharge on aquifer discharge. *Journal of the American Water Resources Association* 46, no. 6: 1116–1132. DOI:10.1111/j.1752-1688.2010.00479.x.

Burnham, K.P., and D.R. Anderson. 2002. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag.

Contor, B. 2004. Delineation of Sprinkler and Gravity Application Systems. Idaho Water Resources Research Institute Report, 2004-005.

Coppola, E.A. Jr., A.J. Rana, M.M. Poulton, F. Szidarovszky, and V.W. Uhl. 2005. A neural network model for predicting aquifer water level elevations. *Ground Water* 43, no. 2: 231–241.

Criss, R.E., and W.E. Winston. 2008. Discharge predictions of a rainfall-driven theoretical hydrograph compared to common models and observed data. *Water Resources Research* 44: W10407. DOI:10.1029/2007WR006415.

Harbaugh, A.W., E.R. Banta, M.C. Hill, and M.G. McDonald. 2000. MODFLOW-2000, the U.S. Geological Survey modular ground-water model-user guide to modularization concepts and the ground-water flow process. USGS Open-File Report 00-92.

Houston, J.F.T. 1983. Ground-water systems simulation by time-series techniques. *Ground Water* 21, no. 3: 301–310.

Hurvich, C.M., and C.L. Tsai. 1989. Regression and time series model selection in small samples. *Biometrika* 76, no. 2: 297–307.

Idaho Department of Water Resources. 2013. Enhanced Snake Plain Aquifer Model Version 2.1 Final Report (Draft). Boise, Idaho. 113 pp.

Idaho Water Resource Board. 2008. Eastern Snake Plain Aquifer (ESPA) Comprehensive Aquifer Management Plan. 41 pp.

Kaczmarek, Z. 1961. Some methodical aspects of ground water forecasting. *International Association of Scientific Hydrology Bulletin* 6, no. 4: 70–74.

Knight, J.H., M. Gilfedder, and G.R. Walker. 2005. Impacts of irrigation and dryland development on groundwater discharge to rivers—A unit response approach to cumulative impacts analysis. *Journal of Hydrology* 303: 79–91. DOI:10.1016/j.jhydrol.2004.08.018

Kuntz, M.A., H.R. Covington, and L.J. Schorr. 1992. An overview of basaltic volcanism of the eastern Snake River Plain, Idaho. In *Regional Geology of Eastern Idaho and Western Montana*, ed. P.K. Link, M.A. Kuntz, and L.B. Platt, 227–267. Geological Society America Memoir 179.

Kullback, S., and R.A. Leibler. 1951. On information and sufficiency. *Annals of Mathematical Statistics* 22: 79–86.

Milly, P.C.D., J. Betancourt, M. Falkenmark, R.M. Hirsch, Z.W. Kundzewicz, D.P. Lettenmaier, and R.J. Stouffer. 2008. Stationarity is dead: Whither water management? *Science* 319, no. 5863: 573–574.

Nash, J.E., and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models: Part I—A discussion of principles. *Journal of Hydrology* 10, no. 3: 282–290.

Poeter, E. 2007. All models are wrong, how do we know which are useful?—Looking back at the 2006 Darcy lecture tour. *Ground Water* 45, no. 4: 390–391.

Shumway, R.H., and D.S. Stoffer. 2011. *Time Series Analysis and Its Applications*, 3rd ed. New York: Springer.

Smith, R.P. 2004. Geologic setting of the Snake River plain aquifer and Vadose zone. *Vadose Zone Journal* 3: 47–58.

Sugiura, N. 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics.* A7: 13–26.

Townley, L.R. 1995. The response of aquifers to periodic forcing. *Advances in Water Resources* 18, no. 3: 125–146.

Whitehead, R.L. 1986. Geohydrologic framework of the Snake River Plain, Idaho and Eastern Oregon. Atlas HA-681.

Zaltsberg, E. 1987. Evaluation and forecasting of ground water runoff in a small watershed in Manitoba. *Hydrological Sciences Journal* 32, no. 1: 69–84.

Zaltsberg, E. 1982. Methods of forecasting and mapping of ground-water tables in the USSR. *Ground Water* 20, no. 6: 675–679.