USGS
science for a changing world

Prepared in cooperation with the Idaho Department of Environmental Quality and the Bureau of Reclamation

Estimating Locations of Perennial Streams in Idaho Using a Generalized Least-Squares Regression Model of 7-Day, 2-Year Low Flows

Scientific Investigations Report 2009–5015

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

Cover: Middle fork of the Payette River near Boiling Springs, Idaho, looking upstream.
(Photograph taken by Molly S. Wood, U.S. Geological Survey, July 2007.)

Estimating Locations of Perennial Streams in Idaho Using a Generalized Least-Squares Regression Model of 7-Day, 2-Year Low Flows

By Molly S. Wood, Alan Rea, Kenneth D. Skinner, and Jon E. Hortness

Prepared in cooperation with the Idaho Department of Environmental Quality and the Bureau of Reclamation

Scientific Investigations Report 2009–5015

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

U.S. Department of the Interior
KEN SALAZAR, Secretary

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

U.S. Geological Survey, Reston, Virginia: 2009

Suggested citation:
Wood, M.S., Rea, Alan, Skinner, K.D., and Hortness, J.E., 2009, Estimating locations of perennial streams in Idaho using a generalized least-squares regression model of 7-day, 2-year low flows: U.S. Geological Survey Scientific Investigations Report 2009-5015, 26 p.

Contents

Figures

Tables

Conversion Factors, Datums, and Abbreviations and Acronyms

Conversion Factors

Multiply	By	To obtain
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
inch (in.)	2.54	centimeter (cm)
foot (ft)	0.3048	meter (m)
meter (m)	3.281	foot (ft)
mile (mi)	1.609	kilometer (km)
square mile (mi^2)	2.590	square kilometer (km^2)

Datums

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

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

Conversion Factors, Datums, and Abbreviations and Acronyms—Continued

Abbreviations and Acronyms

Abbreviations and Acronyms	Definition
$7Q_2$	7-day, 2-year low flow
BURP	Beneficial Use Reconnaissance Program
DEM	Digital Elevation Model
EMAP	Environmental Monitoring and Assessment Program
ESRI	Environmental Systems Research Institute
GIS	Geographic Information System
GLS	generalized least-squares
HUC	Hydrologic Unit Code
IDEQ	Idaho Department of Environmental Quality
NED	National Elevation Dataset
NHD-Hi_Res	National Hydrography Dataset High-Resolution
PDSI	Palmer Drought Severity Index
RF3	River Reach File 3
SNF	Suislaw National Forest
USEPA	U.S. Environmental Protection Agency
USGS	U.S. Geological Survey

This page intentionally left blank.

Estimating Locations of Perennial Streams in Idaho Using a Generalized Least-Squares Regression Model of 7-Day, 2-Year Low Flows

By Molly S. Wood, Alan Rea, Kenneth D. Skinner, and Jon E. Hortness

Abstract

Many State and Federal agencies use information regarding the locations of streams having intermittent or perennial flow when making management and regulatory decisions. For example, the application of some Idaho water quality standards depends on whether streams are intermittent. Idaho Administrative Code defines an intermittent stream as one having a 7-day, 2-year low flow ($7Q_2$) less than 0.1 ft^3/s. However, there is a general recognition that the cartographic representation of perennial/intermittent status of streams on U.S. Geological Survey (USGS) topographic maps is not as accurate or consistent as desirable from one map to another, which makes broad management and regulatory assessments difficult and inconsistent. To help resolve this problem, the USGS has developed a methodology for predicting the locations of perennial streams based on regional generalized least-squares (GLS) regression equations for Idaho streams for the $7Q_2$ low-flow statistic. Using these regression equations, the $7Q_2$ streamflow may be estimated for naturally flowing streams in most areas in Idaho. The use of these equations in conjunction with a geographic information system (GIS) technique known as weighted flow accumulation allows for an automated and continuous estimation of $7Q_2$ streamflow at all points along stream reaches. The USGS has developed a GIS-based map of the locations of streams in Idaho with perennial flow based on a $7Q_2$ of 0.1 ft^3/s and a transition zone of plus or minus 1 standard error. Idaho State cooperators plan to use this information to make regulatory and water-quality management decisions.

Originally, $7Q_2$ equations were developed for eight regions of similar hydrologic characteristics in the study area, using long-term data from 234 streamflow-gaging stations. Equations in five of the regions were revised based on spatial patterns observed in the initial perennial streams map and unrealistic behavior of the equations in extrapolation. The standard errors of prediction for the final equations ranged from a minimum of +75.0 to -42.9 percent in the central part of the study area to a maximum of +277 to -73.5 percent in the southern part of the study area. The equations are applicable only to unregulated, naturally-flowing streams and may produce unreliable results outside the range of explanatory variables used for equation development. Extrapolation outside the range of available data was necessary, however, to predict perennial flow initiation points and transition zones along stream reaches.

The map of perennial streams was evaluated by comparing predicted stream classifications with four independent datasets, including field observations by other government agencies. Overall, 81 percent of the comparison data points agreed with the USGS perennial streams model. Regions with the highest number of disagreements had a high percentage of mountainous and forested area with potential mountain front recharge zones, and regions with the highest agreements had a high percentage of low gradient, low elevation area. As a whole, the USGS model predicted a higher number of perennial streams than predictions made with the independent datasets. Some disagreements were due to poor site location coordinates, timing of the comparison site visits during unusually wet or dry years, discrepancies in classification criteria, and variable ground water contributions to flow in some areas.

The Idaho Department of Environmental Quality Beneficial Use Reconnaissance Program (BURP) dataset is considered the most representative dataset for comparison because it covered a range of climate conditions and the number of sites visited were consistent from year to year during the study period. Eighty-five percent of BURP comparison data points agreed with the USGS perennial streams model. Although site-specific flow data may be needed to correctly classify streams in some areas, this information rarely is available and is not always practical to obtain. The USGS perennial streams map was determined to be a spatially-consistent and accurate estimate of the locations of perennial streams on a broad scale.

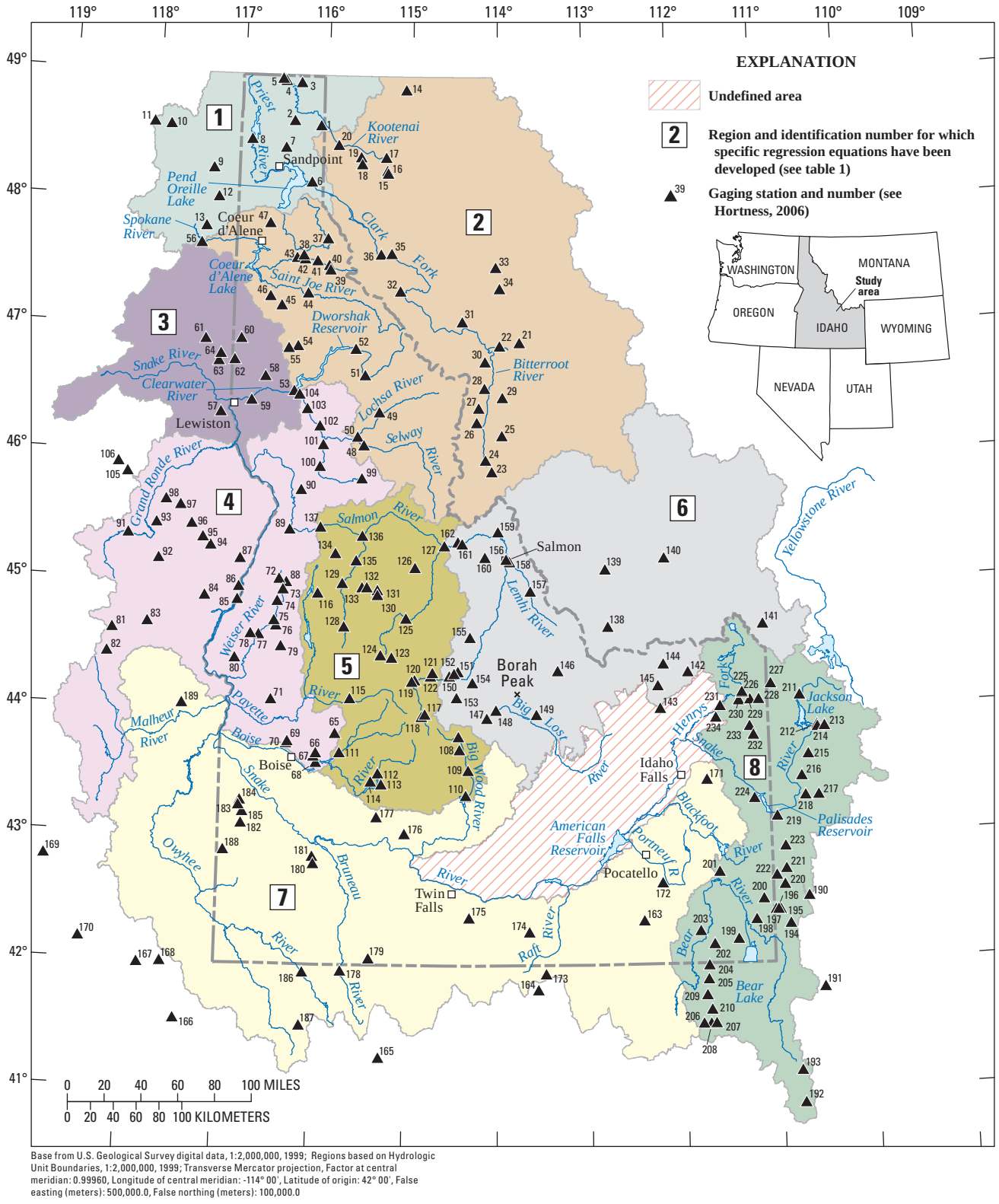
Introduction

The perennial or intermittent status of a stream has bearing on many regulatory requirements. For example, the classification of flow as perennial or intermittent in a stream reach is considered in the application of Total Maximum Daily Load requirements for Idaho. Additionally, permanence of flow has bearing on monitoring and methods of assessing water quality, application of water quality standards, and determination of appropriate use designations. U.S. Geological Survey (USGS) topographic maps commonly are used to determine the perennial or intermittent status of stream. However, there is a general recognition that the cartographic representations of perennial and intermittent streams on USGS topographic maps are not as accurate or consistent as desirable from one map sheet to another. As a result, the USGS, in cooperation with the Idaho Department of Environmental Quality (IDEQ) and Bureau of Reclamation, is attempting to better define the perennial and intermittent status of streams in Idaho.

Idaho Administrative Code defines an intermittent stream as one having a hydrologically-based, unregulated 7-day, 2-year low flow ($7Q_2$) of less than 0.1 ft^3/s (State of Idaho, 2006). The $7Q_2$ is the annual minimum mean streamflow over 7 consecutive days that has a 50 percent probability of not being exceeded in any one year. The USGS has developed regional regression equations for Idaho streams for several low-flow statistics, including the $7Q_2$ statistic (Hortness, 2006), using long-term streamflow-gaging station data. Using these regression equations, the $7Q_2$ streamflow may be estimated for naturally flowing streams in most areas of Idaho, based on estimates of certain basin and climatic characteristics. The study area and hydrologic regions for which equations were developed are shown in figure 1. The boundary of region 6 was extended in the Hortness (2006) study to include streamflow-gaging station data in Montana; however, for this study, the equations were applied only within the Idaho State boundary in region 6. The undefined region shown in figure 1 represents the eastern Snake River Plain, which includes several dams, major irrigation diversions, springs, and drainages and channel bottoms with high infiltration rates. Due to these conditions, flows in this area could not be characterized using a regional regression approach and are not included in this analysis.

The best sources of information available regarding the perennial or intermittent status of streams often are the USGS 1:24,000-scale topographic maps for the area in question: perennial streams are shown on these maps as solid blue lines, and intermittent streams are shown as dot-dashed lines. Streams from these maps have been digitally captured into the National Hydrography Dataset High-Resolution (NHD Hi-Res). According to the draft standards for the NHD Hi-Res, an intermittent stream "contains water for only part of the year, but more than just after rainstorms and at snowmelt", and a perennial stream "contains water throughout the year, except for infrequent periods of severe drought" (U.S. Environmental Protection Agency and U.S. Geological Survey, 1999). Similar definitions were applied in developing the NHD Medium-Resolution, which generally was based on 1:100,000-scale USGS topographic maps.

Although the information represented on USGS topographic maps generally was field verified at the time of map compilation, it was not always possible to verify the perennial or intermittent status of every stream. Additionally, the various topographic maps were compiled over a period of many decades, using varying technologies, cartographers, and standards. Therefore, adjacent topographic maps might have been developed at different times, using different techniques, resulting in mapped streams that change perennial or intermittent status or density where they cross map quadrangle boundaries. Differing techniques and standards also were used and mistakes sometimes were made in the process of digitally capturing the topographic map information and incorporating it into the NHD, which was compiled under 8-digit Hydrologic Unit Codes (HUCs). As a result, streams in the NHD often show changes in patterns at HUC boundaries. The spatial distribution of streams identified as perennial according to the NHD in an area of northern Idaho is shown in figure 2. Several non-realistic density differences are visible following 1:24,000-scale quadrangle and HUC boundaries.



Base from U.S. Geological Survey digital data, 1:2,000,000, 1999; Regions based on Hydrologic Unit Boundaries, 1:2,000,000, 1999; Transverse Mercator projection, Factor at central meridian: 0.99960, Longitude of central meridian: -114° 00', Latitude of origin: 42° 00', False easting (meters): 500,000.0, False northing (meters): 100,000.0

Figure 1. Location of study area, eight regions, and streamflow-gaging stations in Idaho and surrounding areas used to estimate low-flow frequency statistics for unregulated streams (adapted from Hortness, 2006).

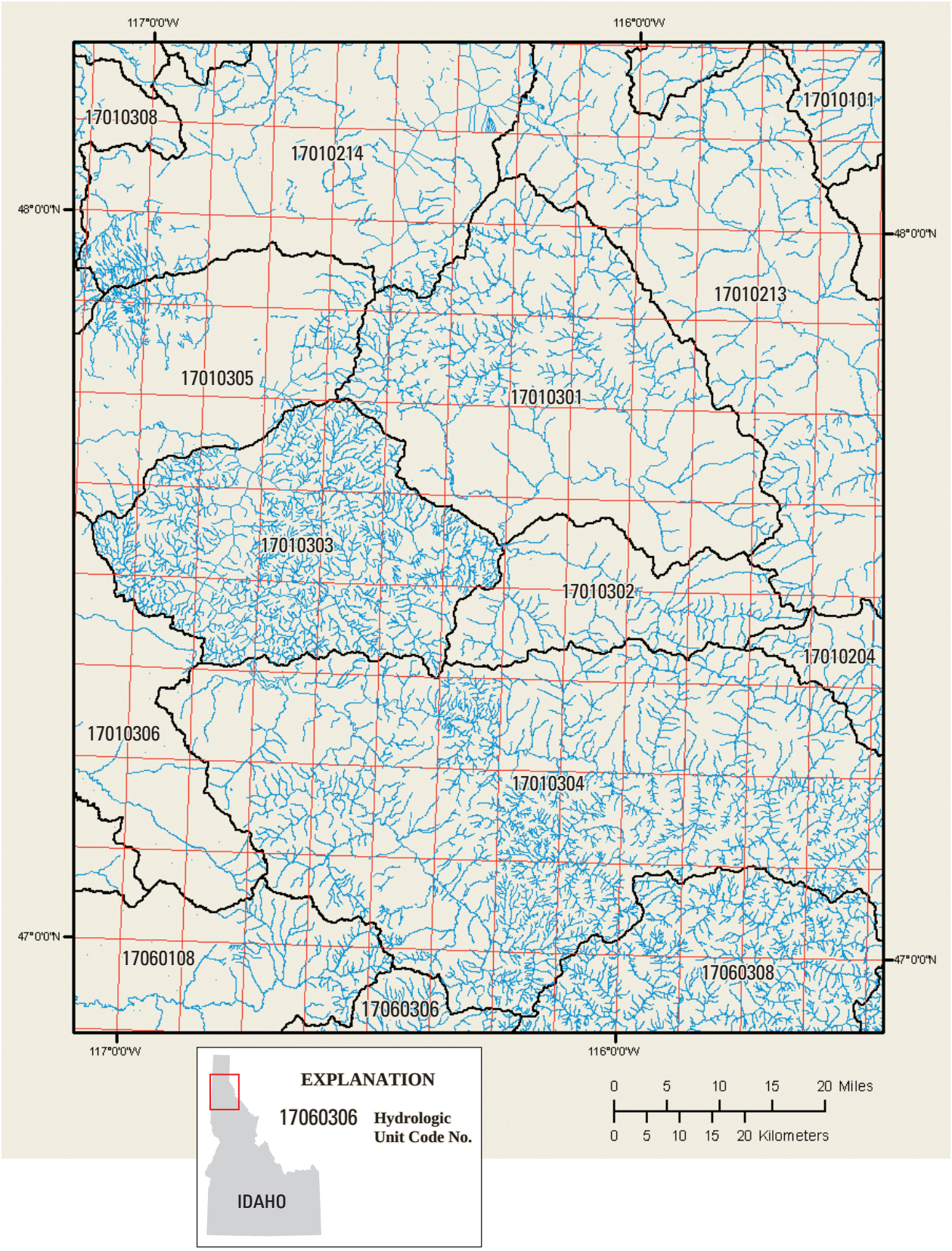


Figure 2. Screen capture of spatial distribution of perennial streams (blue), Hydrologic Unit Code boundaries (black), and 1:24,000-scale quadrangles (red) from National Hydrography Dataset High-Resolution in northern Idaho.

Purpose and Scope

The purpose of this report is to describe the procedures used to create a geospatial model of perennial streams in Idaho and to assess the validity of the model. A by-product of the effort to model perennial streams is an improved suite of geospatial datasets derived from 10-m resolution Digital Elevation Models (DEMs). A full description of the geospatial datasets generated and links to Federal Geographic Data Committee-compliant metadata are provided in Rea and Skinner (2009).

This report describes some shortcomings of previous datasets of perennial streams in Idaho and shows qualitative improvements in the spatial patterns of modeled perennial streams. Quantitative analysis of these data improvements by comparison to field data also is presented.

Description of Study Area

The study area comprises the entire state of Idaho and parts of Washington, Oregon, Nevada, Utah, Wyoming, and Montana where hydrologic basins cross state boundaries (fig. 1). The area is divided into eight hydrologic regions and one undefined region, based on geographic features and a cluster analysis of streamflow data, as described in Hortness and Berenbrock (2001) and Hortness (2006). Terrain is highly varied across the study area, with rugged, mountainous areas in the central and northern areas and gently sloping plains, hills, and canyons in the south and western panhandle of Idaho. Elevations range from 738 ft in Lewiston (western panhandle) to 12,655 ft at Borah Peak (central mountains). Geologic features generally consist of igneous, metamorphic, and sedimentary rocks ranging in age from Precambrian to Holocene (Bond, 1978). The southern and western parts of the study area are dominated by basalts, and the granitic Idaho batholith is the major geologic feature in the central and far northern areas (Ross and Savage, 1967). Glacial erosion and deposition have molded the landscape in mountainous areas and in the northern tip of Idaho's panhandle (Link and Welhan, 2002).

Annual precipitation varies widely with topography, ranging from less than 10 in. in south-central Idaho to 60 to 70 in. in the central mountains (Molnau, 1995). Generally, the source of moisture for precipitation is the Pacific Ocean, although many summer thunderstorms in eastern Idaho are generated by moisture-laden air moving north from the Gulf of Mexico and the Caribbean (National Oceanic and Atmospheric Administration, 1985). About 60 percent of annual precipitation falls during winter and early spring (October through March) as a result of orographic effects (Molnau, 1995); however, summer thunderstorms can contribute significantly to total annual precipitation in the southeastern part of the study area. Highest streamflows generally are observed in April, May, June, and July as a result of snowmelt and rain-on-snow storm events. Baseflow in most unregulated streams dominates flow during August through March, and annual minimum streamflows typically occur in October through January (Hortness, 2006).

Idaho Definition of Perennial Streams

The State of Idaho defines intermittent and ephemeral streams, but not perennial streams, in its water quality rules. Intermittent waters are defined as "a stream, reach, or water body which naturally has a period of zero flow for at least 1 week during most years" (State of Idaho, 2006). Ephemeral waters are on the drier end of the spectrum of flow from intermittent waters, and perennial waters are, by default, all streams with flow greater than meets the definition of intermittent waters. Where flow records are available, a stream with a hydrologically-based, unregulated $7Q_2$ flow of less than 0.1 ft^3/s is considered intermittent. Streams with natural perennial pools containing significant aquatic life uses are not intermittent (State of Idaho, 2006). Because the presence of natural perennial pools is a site-specific determination, this project focuses only on the unregulated $7Q_2$ flow less than the 0.1 ft^3/s criterion for perennial stream classification.

Related Studies

A number of investigations have been conducted to classify or map intermittent and perennial streams, or both, often based on state-specific definitions and in response to observed deficiencies in existing maps. Many approaches rely on site-specific field determinations to classify streams, which may be time consuming and costly but accurate if proper and consistent techniques are applied. Recent approaches have relied on statistical techniques and GIS or remotely-sensed data to predict the locations of intermittent and perennial streams.

A comprehensive study involving field techniques was completed by the State of North Carolina. The North Carolina technique requires trained field technicians to score and classify a stream site based on hydrologic, geomorphologic, and biologic metrics (North Carolina Division of Water Quality, 2005). The technique is designed to classify a site based on long-term indicators of perennial or intermittent flow, not simply on the presence of flow at the time of visit. The U.S. Environmental Protection Agency (USEPA) has developed a similar field operations manual for assessing hydrologic permanence in headwater streams (Fritz and others, 2006). The U.S. Army Corps of Engineers and USEPA currently are combining elements of these methods into a guidance document for classifying streams in Oregon (Brian Topping, U.S. Environmental Protection Agency, oral and written commun., 2007). A similar technique has been adopted by the State of Virginia, Fairfax County (Fairfax County, Virginia, Stormwater Planning Division, 2004), although their technique has been adapted to account for seasonal variations and to include anecdotal or historical information in the final scoring of the site.

Jaeger and others (2007) conducted a study in southwest Washington State to field map channel heads and perennial flow initiation points in streams in three test watersheds. They designated areas with continuous and discontinuous flows along stream reaches over a range of hydrologic conditions from February to September 2003. The perennial flow initiation point was identified as the farthest upslope location with flow during baseflow conditions (that is, summer). They determined that the relation between drainage area and perennial flow initiation varied by lithology. For example, the initiation point migrated seasonally to a much greater extent when underlying lithology was sandstone rather than basalt; as a result, the relation between drainage area and perennial flow for all lithologic types was not well-defined. From a practical standpoint, Jaeger and others (2007) recommended establishing minimum drainage areas for each lithology (for example, 0.004 mi^2 for sandstone and 0.0008 mi^2 for basalt) as a conservative management strategy for identifying and mapping perennial flow initiation points on a broader scale.

Clarke and others (2008) combined regional GIS data and field observations to delineate a high-resolution, synthetic stream network and hydrogeomorphic attributes in coastal Oregon. The resulting map and dataset have served various uses—from management of riparian forests to habitat evaluation for salmonids. The probability of perennial flow in a stream was one of many stream attributes modeled. The probability was based on drainage area, developed using a dataset from the Siuslaw National Forest (SNF) where field technicians determined the upper limit of perennial flow in selected streams during baseflow conditions over a 2-year period. Clarke and others (2008) assumed that, in drainage areas between 0.004 and 0.14 mi^2, probability of perennial flow in a stream reach will follow a cumulative distribution function developed using the SNF field dataset. For example, a 70 percent probability of perennial flow is assigned to stream reaches with a drainage area of 0.015 mi^2 because 70 percent of the stream reaches in the field dataset at that drainage area had perennial flow. These probabilities were used to generate a map of synthetic perennial streams.

Bent and Steeves (2006) used a statistical and GIS-based approach to predict the locations of perennial streams in Massachusetts. A logistic regression equation was developed by relating field observations of the presence or absence of flow at 351 unregulated stream sites with selected basin characteristics such as drainage area, areal percentage of sand and gravel deposits, areal percentage of forest cover, and whether the site was in the eastern or western region of the state. Sites with drainage areas greater than 2.00 mi^2 were assumed to flow perennially and were not used in the logistic regression equation. Bent and Steeves (2006) then used a GIS automated mapping procedure to calculate the explanatory variables in the logistic regression along a stream reach and determine where the transition from intermittent to perennial flow occurs, based on a probability threshold. They determined that the logistic regression equation correctly classified intermittent/perennial flow at 80.3 percent of streams visited for verification with drainage areas from 0.00 to 10.99 mi^2, but that USGS topographic maps portrayed the correct stream classifications at only 69.2 percent of the same verification sites. Olson and Brouillette (2006) developed a similar logistic regression equation to predict the locations of intermittent streams in Vermont. The equation was developed by relating field observations from 682 unregulated stream sites to basin characteristics such as drainage area, elevation, ratio of basin relief to basin perimeter, and areal percentage of well- and moderately well-drained soils. Their procedure correctly classified intermittent/perennial flow at 85 percent of streams visited for verification in Vermont.

Many of these efforts to predict the locations of intermittent and perennial streams have demonstrated the usefulness of larger-scale, statistical and GIS-based techniques combined with field verification data over a range of hydrologic conditions. Studies based solely on field visits often are hampered by time limitations. As a result, a range of hydrologic conditions is not always captured in the study, and therefore, final stream classifications may be biased.

Approach

The USGS has developed a GIS-based map of the locations of streams in Idaho with perennial flow based on generalized least-squares (GLS) regression models of $7Q_2$ flows and continuous parameter estimation of $7Q_2$ at ungaged locations. A transition zone of plus or minus 1 standard error was modeled around the $7Q_2$ cutpoint of 0.1 ft^3/s to account for statistical, climatic, and hydrologic variability. Flow downstream of the transition zone is considered perennial, and flow upstream of the transition zone is considered intermittent.

Development of $7Q_2$ Regression Models

The USGS operates a network of streamflow-gaging stations in Idaho that provides data for various purposes, and low-flow statistics can be calculated from the streamflow data collected at these locations. Because streamflow-gaging stations cannot be located at all sites where streamflow information is needed, other methods are used to estimate streamflow statistics for these ungaged sites. One of the most common methods is to develop regression equations that relate streamflow statistics to selected basin characteristics. With certain limitations, this allows for the estimation of streamflow statistics at ungaged stream locations throughout the State.

During the equation development process, the study area was divided into eight separate geographic regions of similar characteristics. Data from a total of 234 streamflow-gaging stations were included in the analysis. This included all streamflow-gaging stations in Idaho and those in adjacent States within an approximate 80-mi buffer surrounding Idaho with 10 or more years of record through water year 2003, which exhibited little or no sign of trends, and were unaffected

by regulations, diversions, or both. More than 50 basin characteristics were obtained for each of the 234 streamflow-gaging stations included in the study. The basin characteristics were obtained using GIS techniques from digital datasets such as DEMs, precipitation grids, and land-use grids. Several basin characteristics were removed from consideration after a review of the correlation plots of the data. Generally, if two basin characteristics correlated well with each other, the characteristic least difficult to obtain was kept and the other was removed. Other characteristics were removed because of missing data or difficulty in obtaining the data.

Relevant low-flow frequency statistics (including $7Q_2$) were computed for the 234 streamflow-gaging stations. Low-flow frequency statistics are determined using the annual minimum mean flows for any given number of days (N-day low flows) during an annual period. The mean flow for each N-day period throughout the annual period is computed, and the minimum value is selected. The series of annual minimum N-day values are then fit to a log-Pearson Type III distribution to determine the recurrence intervals. The annual period referred to as a climatic year (April 1 through March 31) is often used in low-flow analyses because the annual low-flow period in most parts of the country occurs during the late summer and autumn months. Use of the climatic year allows for inclusion of the entire low-flow period in the same year, whereas use of the traditional water year (October 1 through September 30) may artificially separate the low-flow period into two years.

Multiple linear regression techniques then were used to develop equations for each region that related each low-flow frequency statistic to one or more basin characteristics. A GLS technique that weights the station data to compensate for spatial correlation and differences in record length was used to obtain the equations. The analyses resulted in the development of equations to estimate low-flow frequency statistics for unregulated streams in each of the eight regions in the State. These equations are published and described in more detail in Hortness (2006).

Development of DEMs and Continuous Parameter Grids

The datasets used in this project primarily are derived from the 10-m resolution DEMs produced by the USGS and the 1:24,000-scale NHD. The source DEM was obtained from the 1/3-arc-second resolution National Elevation Dataset (NED) and was projected into the Idaho Transverse Mercator projection based on the North American Datum of 1983 for each 8-digit HUC in Idaho. The NHD vector streams were integrated into the raster NED data, using a process often referred to as stream burning (Saunders, 2000). This process uses the AGREE computer program developed by Hellweger and Maidment (1998). The AGREE program corrects for DEM flow path displacement errors when delineating catchments. The AGREE process and additional manipulation of NED DEM and NHD datasets are presented in more detail in Rea and Skinner (2009).

A continuous parameter grid of 10 by 10 m cells was established for the perennial streams map. In a continuous parameter grid, each cell contains the value of some parameter measured for the entire drainage area upstream of that cell. The flow-accumulation grid described by Jenson and Domingue (1988) is the most basic continuous parameter grid. Each cell in a flow-accumulation grid contains the number of cells upstream of that cell. The upstream drainage area may be determined by multiplying the number of cells by the area of a cell. This area should be adjusted by adding the area of the cell of interest because the flow accumulation grid does not include this cell.

Values of most basin characteristics, such as mean elevation, were assigned to every cell to create a "weight" grid. A flow accumulation function in ARC/INFO (Environmental Systems Research Institute, 1999) sums values from the weight grid for each cell and accumulates cell values as it moves downgradient. In contrast, if no weight grid is assigned, the flow accumulation function simply totals the number of upstream cells, a process called "unweighted flow accumulation." The weighted flow accumulation value, divided by the unweighted flow accumulation value, gives the mean value of the basin characteristic in the weight grid upstream of any grid cell.

This concept may be used with continuous-value parameters, such as elevation or precipitation, and with categorical-value parameters. For example, using a grid containing 1 for every cell categorized as forested in a land-cover dataset, and 0 for every other cell, the fraction of the watershed area draining to a particular forested grid cell can be computed. The development of the continuous parameter grid is described in more detail in Rea and Skinner (2009).

Computation of Continuous $7Q_2$ Estimates

Most basin characteristics can be calculated on a continuous basis for every grid cell using the method described in the previous section and in Rea and Skinner (2009). If all parameters of a regression equation can be computed in this manner, then the result of the regression equation also may be computed continuously for every grid cell; therefore, the equations developed by Hortness (2006) were used to compute preliminary estimates of the $7Q_2$ statistic for every grid cell in the study area. A map of streams with perennial flow then was developed by comparing these estimates to the state standard of intermittent flow as less than 0.1 ft^3/s. A transition zone also was modeled corresponding to flow estimates of 0.1 ft^3/s plus and minus one standard error. It should be noted that although standard error is valid only for the range of predictor variables, it was used to model the transition zone in the range of extrapolation because no other statistical measure was available. The transition zone

represents the part of the stream that, given annual climate and hydrologic fluctuations as well as inherent statistical uncertainty, is assumed to contain the point at which a stream changes from perennial to intermittent. In other words, flow in reaches upstream of the top of the transition zone is considered intermittent, and flow in reaches downstream of the bottom of the transition zone is considered perennial. Within the transition zone, flow could be perennial or intermittent. The preliminary map generated some anomalous stream patterns that necessitated some revisions to the $7Q_2$ equations developed by Hortness (2006).

Model Revisions Based on Spatial Patterns

Statistical regression models are developed using measured data. Estimates of prediction error can be developed for data that fall inside the range of values used to develop the regression model. However, outside the range of the values used to develop the regression model, the bounds of statistical confidence are unknown. Regression analyses often yield equations that provide valid results in the range of input values, but may produce invalid or unrealistic estimates when extrapolating beyond the range of input values. Therefore, common statistical principles dictate that regression equations should not be used in extrapolation, but rather should be used only for interpolation (that is, within the range of measured values of the explanatory variables used in the regression analysis).

The purpose of the project was to distinguish between perennial and intermittent streams in Idaho; however, most streamflow measurements used to develop regression equations come from streamflow-gaging stations that are rarely placed on small, intermittent streams. As a result, the sample data typically do not represent the size of streams of interest in this application, and the regression equations must be applied in extrapolation. This was referred to as extreme extrapolation because the equations for $7Q_2$ were applied to small watersheds (as small as a single 10 m by 10 m grid cell). Standard statistical measures are not always useful for assessing the validity of the model in extrapolation. Regression equations must be evaluated as to whether they are well behaved in extreme extrapolation for small streams. The spatial behavior of an equation can be evaluated based on the pattern of perennial streams simulated by the model. Regression models that simulated unrealistic stream network patterns were determined to be badly behaved or ineffective for this application.

If a regression equation includes only one explanatory variable (such as drainage area), regression estimates for a large number of values in the range of interest can be generated to determine how well the regression equation functions. A Monte-Carlo analysis often is used to generate simulated parameter values when multiple variables are involved. This analysis involves estimating the statistical distribution of the observed variables and randomly generating parameter values using this distribution for the range of interest. The combination of large numbers of randomly generated multi-parameter values then can be used to examine the behavior of the regression equation in extrapolation; however, randomly generated parameter values were not needed for this analysis. The complete set of parameter values for every grid cell in the domain of interest can be computed; therefore, the regression equation results can be evaluated for every combination of parameter values that exists in the modeled area, which produces a map of the spatial behavior of the equation.

As a test of regression equation behavior in extrapolation, continuous parameter grids were evaluated for the $7Q_2$ equation for region 8 published by Hortness (2006), and given here as:

$$7Q_2 = 3.86\, A^{0.930}\, BS^{-0.648}, \quad (1)$$

where:

$7Q_2$ is the 7-day, 2-year low flow, in ft^3/s
A is drainage area, in mi^2, and
BS is average basin slope, in percent.

This equation was determined to be not well behaved in extreme extrapolation. The minimum drainage area in region 8 used to develop this regression equation was 6.6 mi^2, and the minimum basin slope was 6.15 percent. In the continuous parameter grid, the minimum drainage area is one 10 by 10 m grid cell, or 3.86×10^{-5} mi^2. Model-generated values for basin slope for very small basins can be zero, which causes a numerically invalid result; therefore, slope is a poor predictor variable for estimating flow characteristics of very small basins, unless it is censored to remove zero values.

Alternative regression models were evaluated due to the poor spatial behavior of some of the previously developed equations. Regression models that produced stream network patterns that clearly were not reasonable in comparison with topographic maps were rejected in favor of models that produced more reasonable patterns. Methods described in Hortness (2006) were followed when developing revised regression equations.

Certain types of categorical variables such as forested area tended to produce unrealistic patterns. In small drainage areas, categorical variables tended to have extreme values of either zero or 100 percent. These extremes are rarely if ever present in the large basins upstream of streamflow-gaging stations and usually would be outside the range of data values used in the regression analysis. For this reason, categorical variables such as forested area were avoided where possible.

The final regression equations developed for modeling perennial streams in Idaho are shown in table 1. The ranges of basin characteristics used to develop the final regression equations are shown in table 2. Revised regression models were selected for five of the eight regions; the original equations by Hortness (2006) were retained for regions 2, 4, and 6. These regression equations were applied to create revised continuous grids of $7Q_2$ estimates for the eight regions in the study area.

The scatterplot in figure 3 shows $7Q_2$ values plotted against drainage area for the (*A*) original and (*B*) revised regression models in small drainages in region 8. Note that basin slope and mean annual precipitation, although parameters in the original and revised equations, respectively, are not shown on the plot. The minimum drainage area used to develop the regression equations, 6.6 mi^2, is shown in figure 3 as a vertical line. Although some spatial patterns can be attributed to variability in additional explanatory variables, it is evident that the original model does not produce stable linear results in extreme extrapolation for small drainage areas. The revised model simulates linear results in extreme extrapolation for small drainage areas better than the original model.

Screen captures of the stream network derived using the (*A*) original (Hortness, 2006) and (*B*) revised perennial streams models for a small area in region 8 are shown in figure 4. The grid cells with $7Q_2$ estimates greater than or equal to 0.1 ft^3/s are shown in pink transparency over a USGS 1:24,000-scale topographic base map. On the original map (fig. 4*A*), unreasonably dense and discontinuous drainage patterns appear that do not correspond to locations that appear to be stream channels based on topographic maps and knowledge of the area. This illustrates that regression equations that are not well behaved (in extreme extrapolation) may be detected by looking for unrealistic spatial patterns in the drainage network derived from application of the regression equation in a continuous manner. The spatial pattern for the revised model (fig. 4*B*) is much more reasonable than the original model (Hortness, 2006). However, the original model would still be preferred for predicting streamflow statistics in streams where extrapolation is not required because the standard error of prediction is better for the original model than the revised model.

The results of the revised regression models for region 8 in a broader spatial context are shown in figure 5. Figure 5*A* shows the results of the original model by Hortness (2006) for one 8-digit HUC (16010201). Figure 5*B* shows the results of the revised model for the same area.

The standard error of prediction for each regional regression equation is shown in table 1. The standard error of prediction is a measure of the overall model error as well as the sample error and is a good indicator of the overall predictive ability of the model within the range of input variables (Pope and others, 2001); however, it is not a suitable measure of the model error in the range of extrapolation because it is calculated based on measured values. The values presented in $\log_{10}$ format represent the errors in the log-transformed equations, and the percentage values represent the range of errors for the untransformed equations, as presented in table 1. The values were determined using error transformation equations presented in Riggs (1968). For example, the percent range for standard error of prediction for region 1 is +86.7 to -46.4 percent. In this region, the error range for a $7Q_2$ value of 1.0 ft^3/s would be 0.54 to 1.87 ft^3/s. The highest standard errors of prediction are observed in regions 3 and 7. These regions encompass the Columbia Plateau, Snake River Plain, and Owyhee Uplift physiographic regions (not labeled in fig. 1), and are characterized by lower elevations, smaller terrain relief, and less rainfall compared to the remaining regions. Large standard errors of prediction also were observed for the original Hortness (2006) regression equations for regions 3 and 7. Hortness (2006) speculated that the large errors in region 3 likely were due to few streamflow-gaging stations available for developing the regression model. Region 7 covers a large part of the study area, and Hortness and Berenbrock (2001) noted a high degree of natural variability in streamflow in region 7 that could contribute to a high error of prediction. Revising the regression model did not improve the standard error of prediction but substantially improved spatial patterns observed in extrapolation.

Table 1. Regression equations for the low-flow statistic, $7Q_2$, for unregulated streams in regions 1–8, Idaho.

[Locations of regions are shown in figure 1. **Standard error of prediction:** Values determined using error transformation equations in Riggs (1968). **Abbreviations**: $7Q_2$, 7-day 2-year low flow, in cubic feet per second; *A*, drainage area, in square miles; *BS*, basin slope, in percent; *E*, mean elevation, in feet; *P*, mean annual precipitation, in inches; *R*, basin relief, in feet; *S*50, slopes greater than 50 percent in percentage of drainage area; *V*, surficial volcanic rocks, in percentage of drainage area; *W*, water in percentage of drainage area]

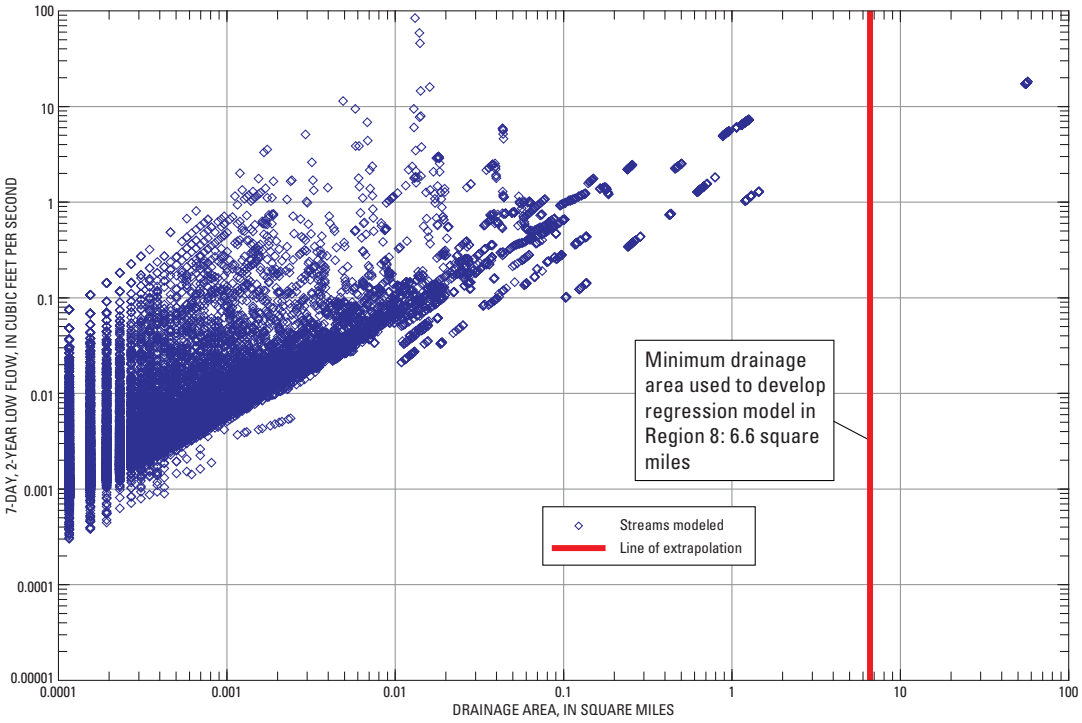
Region	$7Q_2$ regression equation		Standard error of prediction	
	Original (Hortness, 2006)	Revised	$\log_{10}$	Percent
1	$7Q_2 = 0.731A^{0.613}(W+1)^{0.907}$	$7Q_2 = 0.300A^{0.879}$	0.271	+86.7 to -46.4
[1]2	$7Q_2 = 0.000153A^{1.04}P^{1.92}$	No change	0.253	+78.9 to -44.1
3	$7Q_2 = 0.00953A^{0.392}(R/1{,}000)^{3.36}$	$7Q_2 = 0.0214A^{0.971}$	0.277	+236 to -70.2
[1]4	$7Q_2 = 0.0000215A^{1.04}P^{2.41}$	No change	0.341	+119 to -54.4
5	$7Q_2 = 0.000453A^{1.04}(V+1)^{1.51}$	$7Q_2 = 0.334A^{0.963}$	0.249	+77.3 to -43.6
[1]6	$7Q_2 = 0.000133A^{1.05}P^{2.10}$	No change	0.243	+75.0 to -42.9
7	$7Q_2 = 0.0329A^{0.678}(S50+1)^{0.796}$	$7Q_2 = 0.0000115A^{0.837}(E/1000)^{4.658}$	0.577	+277 to -73.5
8	$7Q_2 = 3.86A^{0.930}BS^{-0.648}$	$7Q_2 = 0.00181A^{0.981}P^{1.515}$	0.331	+116 to -53.4

[1] Denotes region with original Hortness (2006) equation.

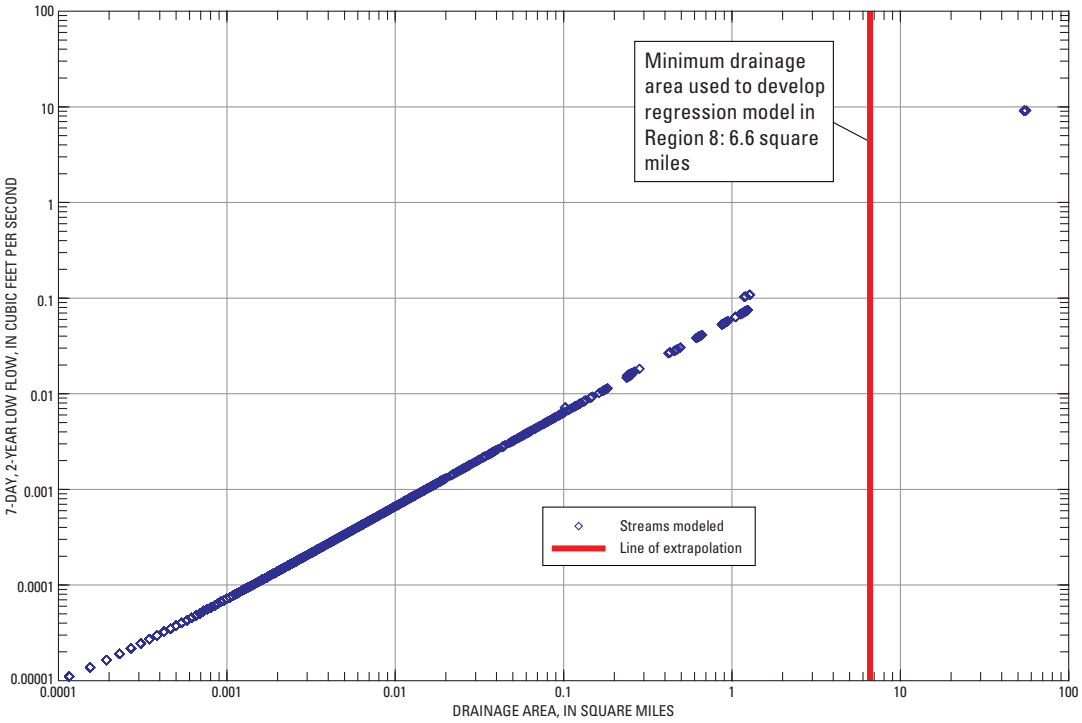
Table 2. Range of values of basin characteristics used to estimate the low-flow statistic, $7Q_2$, for unregulated streams in regions 1–8, Idaho.

[Locations of regions are shown in figure 1. **Equation variables:** *A*, drainage area, in square miles; *E*, mean elevation, in feet; *P*, mean annual precipitation, in inches. **Abbreviations:** $7Q_2$, 7-day 2-year low flow, in cubic feet per second; NA, not applicable—variable not used in regression equation]

	Equation variables		
	A	*E*	*P*
		Region 1	
Maximum	1,011.0	NA	NA
Minimum	12.5	NA	NA
		Region 2	
Maximum	2,442.5	NA	69.4
Minimum	3.0	NA	24.8
		Region 3	
Maximum	674.9	NA	NA
Minimum	17.6	NA	NA
		Region 4	
Maximum	5,507.9	NA	65.6
Minimum	4.0	NA	15.9
		Region 5	
Maximum	12,228.0	NA	NA
Minimum	19.3	NA	NA
		Region 6	
Maximum	6,236.7	NA	42.3
Minimum	6.4	NA	15.3
		Region 7	
Maximum	535.3	8,463.4	NA
Minimum	7.4	2,984.4	NA
		Region 8	
Maximum	874.8	NA	56.0
Minimum	6.6	NA	14.2

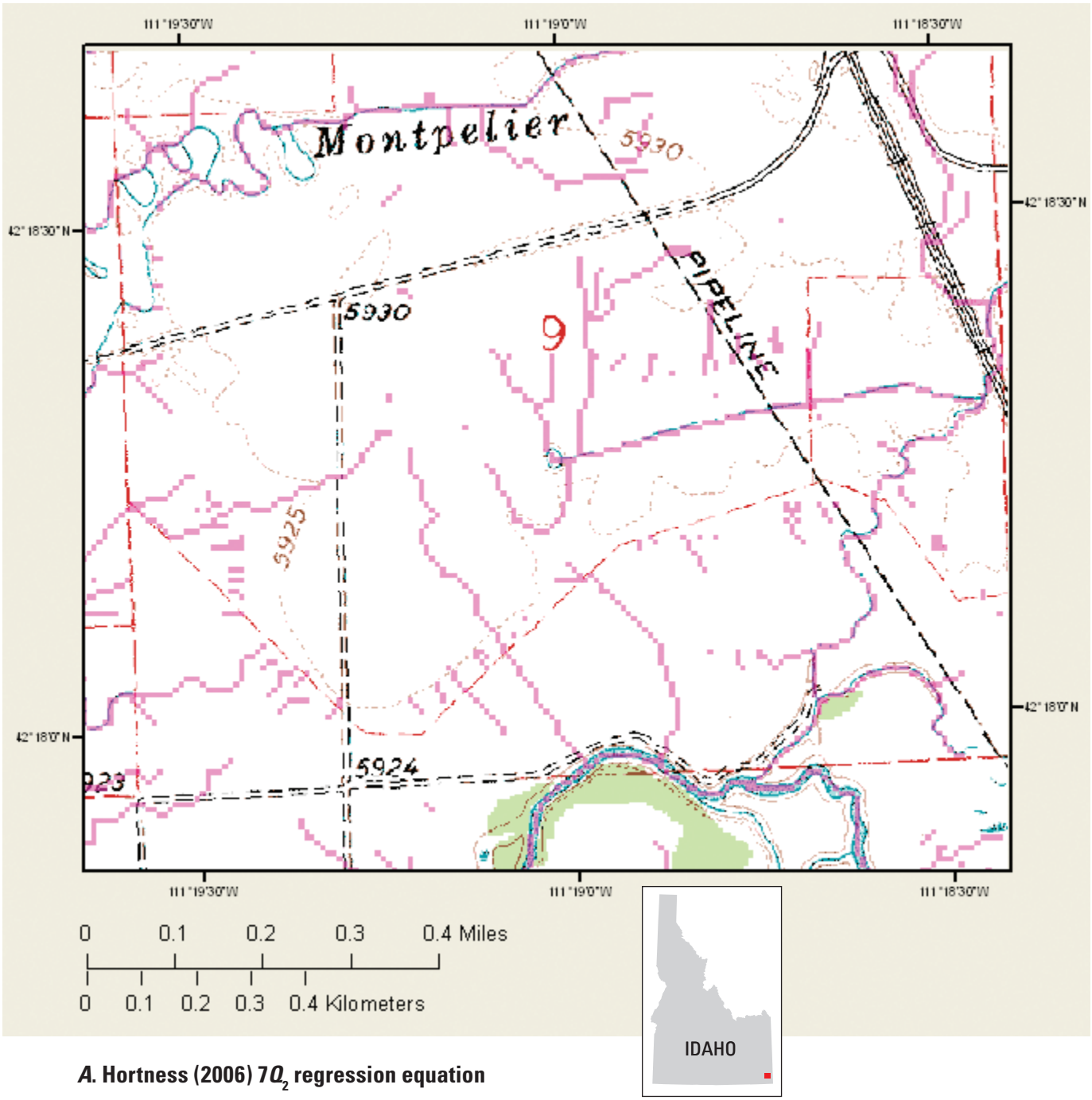


A. **From Hortness (2006) regression equation**



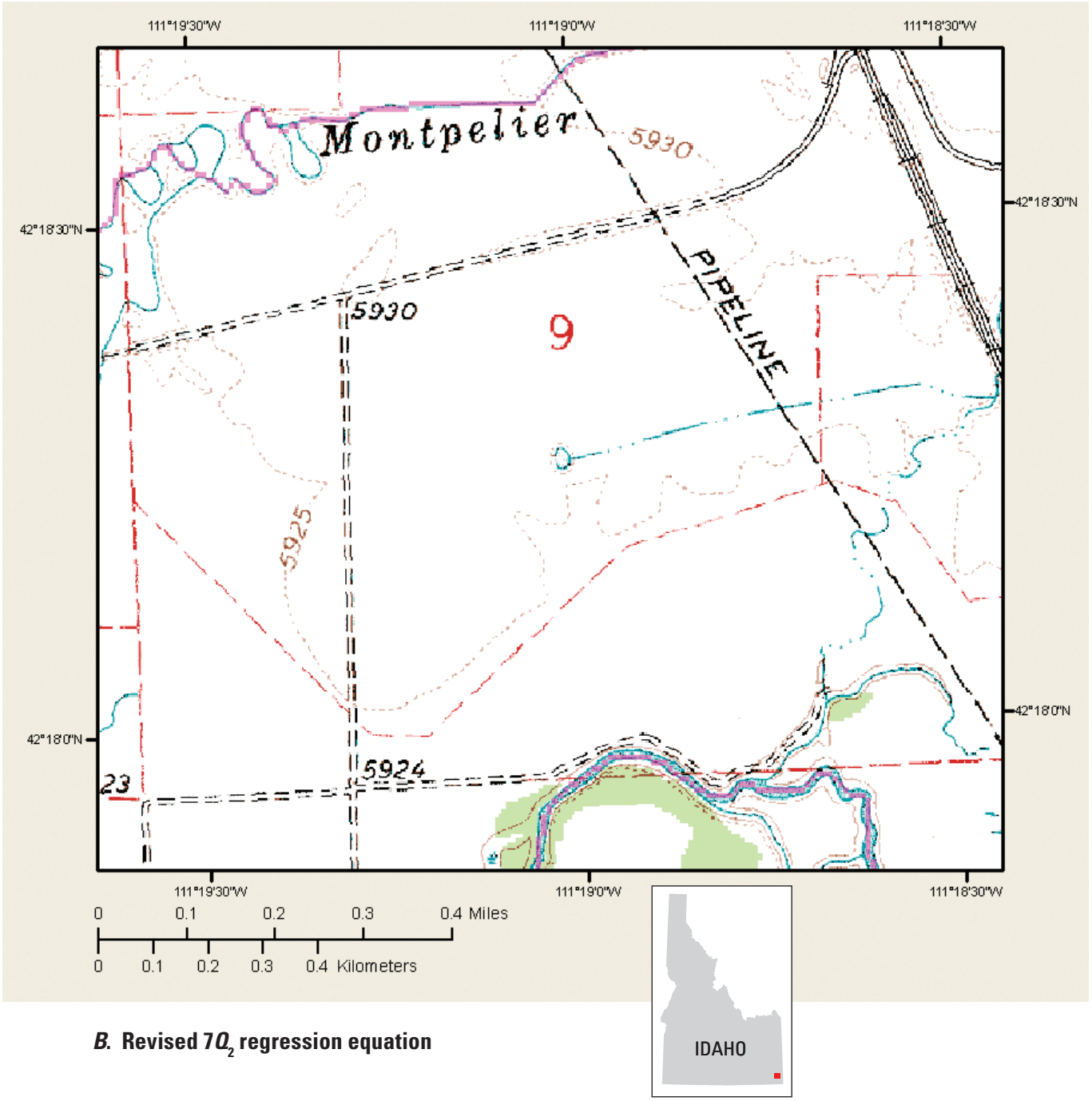
B. **Revised regression equation**

Figure 3. Improvement of $7Q_2$ flow estimates in extreme extrapolation for region 8 from (*A*) the Hortness (2006) regression equation to (*B*) the revised regression equation for Idaho.



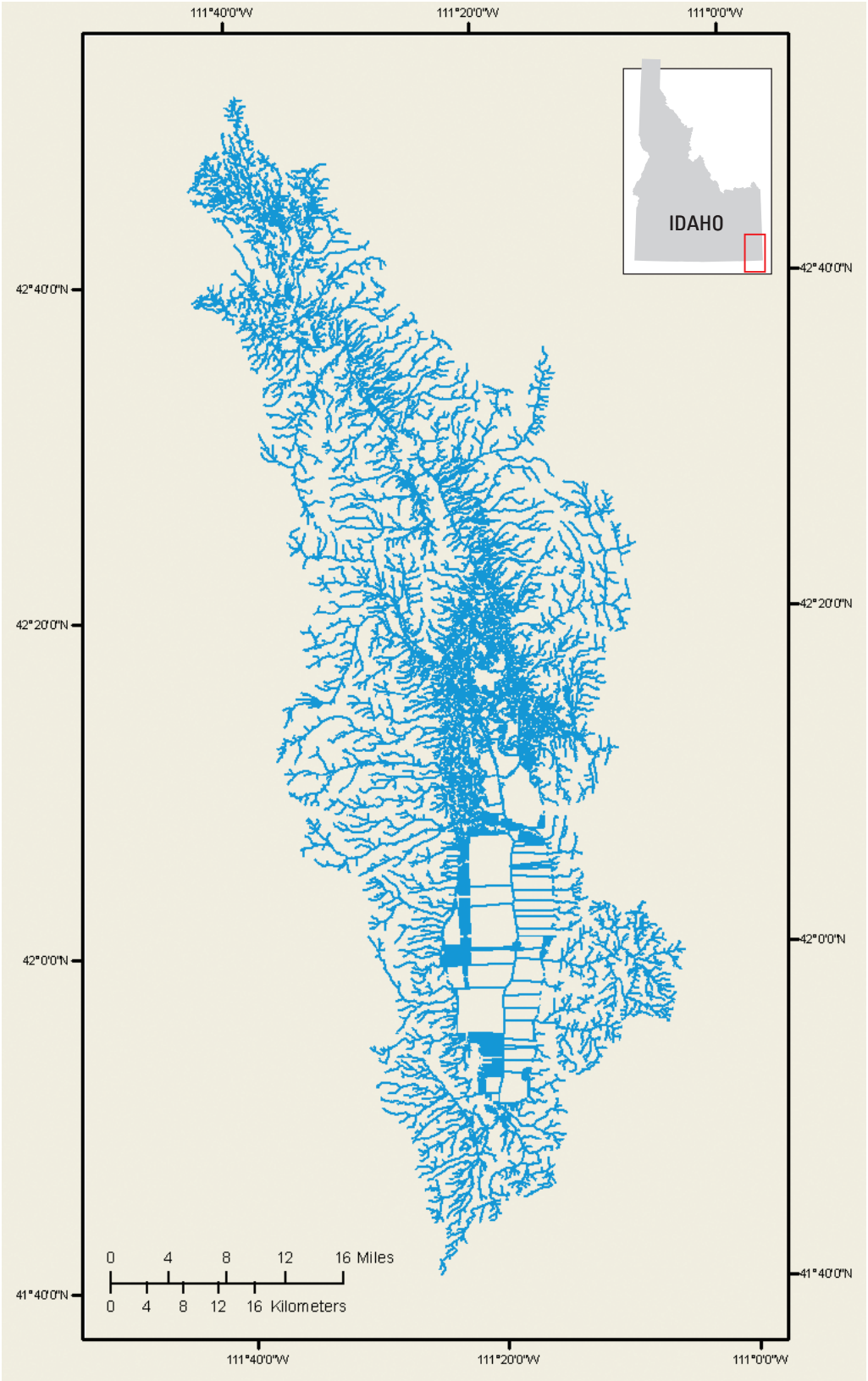
***A*. Hortness (2006) $7Q_2$ regression equation**

Figure 4. Screen captures from U.S. Geological Survey perennial streams model showing spatial pattern of stream network (pink pixels) in a selected area in region 8 derived from (*A*) Hortness (2006) $7Q_2$ regression equation and (*B*) revised $7Q_2$ regression equation for Idaho.



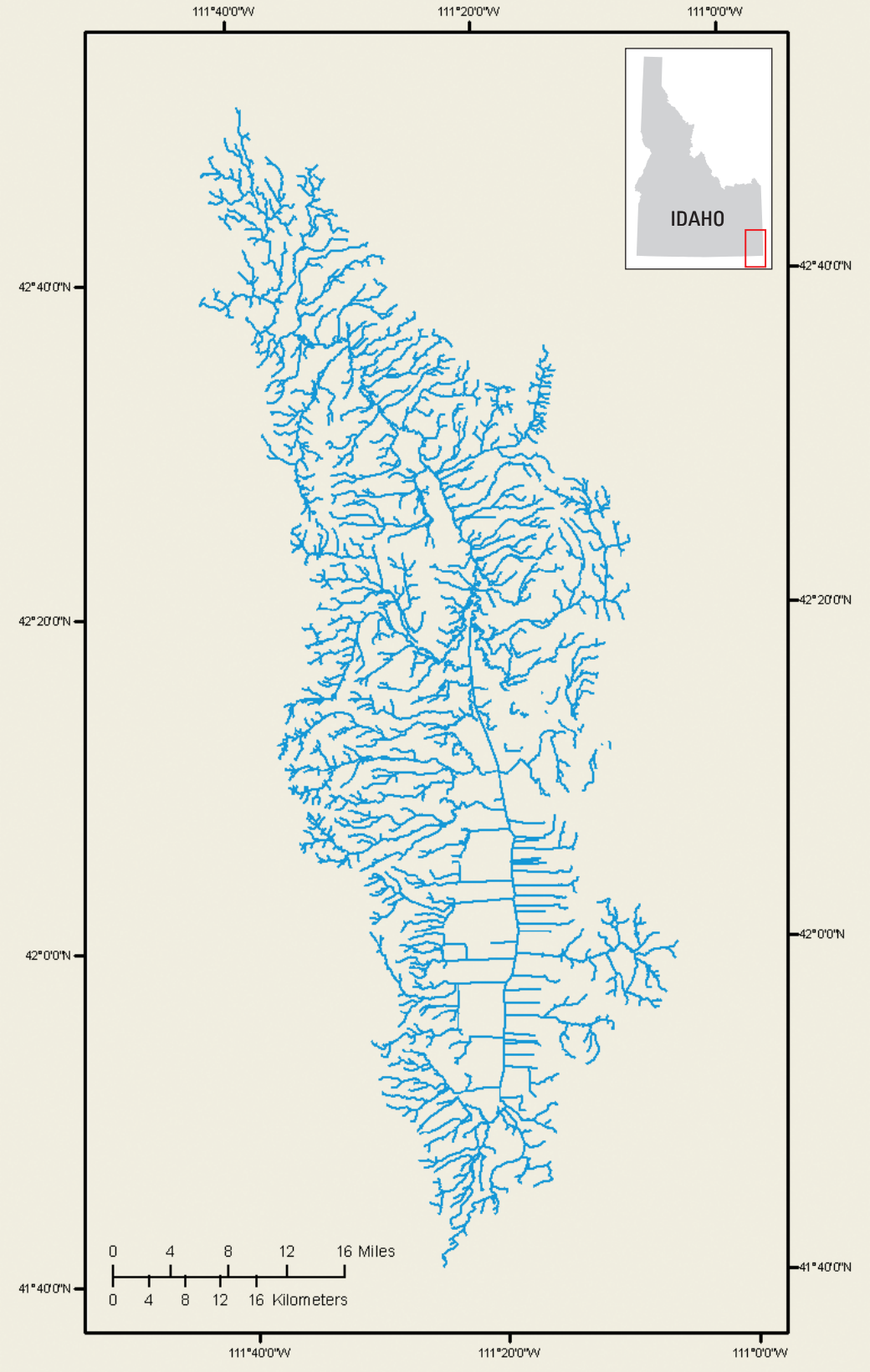
B. **Revised $7Q_2$ regression equation**

Figure 4. Continued.



***A*. Hortness (2006) $7Q_2$ regression equation**

Figure 5. Screen captures from U.S. Geological Survey perennial streams model showing spatial pattern of stream network for Hydrologic Unit Code 16010201 (in region 8) derived from (*A*) Hortness (2006) $7Q_2$ equation and (*B*) revised $7Q_2$ regression equation for Idaho.



B. **Revised $7Q_2$ regression equation**

Figure 5. Continued.

Final Perennial Streams Model

A suite of geospatial datasets has been compiled for estimating streamflow statistics in Idaho, to support statistical modeling of perennial streams. The supporting datasets and metadata are described in detail in Rea and Skinner (2009).

The final map of estimated perennial streams consists of "synthetic" lines or vector representations of streams from the gridded model and is provided in Environmental Systems Research Institute (ESRI) shapefile format. The map and associated files and metadata are available for download at: http://water.usgs.gov/lookup/getspatial?ds412_syntheticperennialstreams.

Although the synthetic streams derived from the regression models provides a consistent, connected representation of the stream network, many users need the results referenced to a commonly recognized hydrography framework, such as the NHD. Therefore, the modeled stream lines were merged with the 1:100,000-scale NHD dataset. Details of this process are beyond the scope of this report but are described in detail in Rea and Skinner (2009).

Although the full map of perennial streams in Idaho is too dense to display in this report, figure 6 shows a screen capture of the perennial streams map in the same area shown in figure 2. Most of the area shown is in region 2, and the original regression equation by Hortness (2006) was applied. The pattern of perennial streams shown is more reasonable than that derived from the NHD Hi-Res and shown in figure 2. However, HUC 17010305, in the northwest area of figures 2 and 6, is known as the Rathdrum Prairie and is an area of very porous soils with very few surface-water streams. Based on the authors' knowledge of the area, the streams shown in figure 2 for HUC 17010305 are a better representation of the perennial streams than what is shown in figure 6. This is a good example of an area where the surface-water hydrography is dominated by local hydrogeologic effects and does not follow the general trend represented by the regression equations.

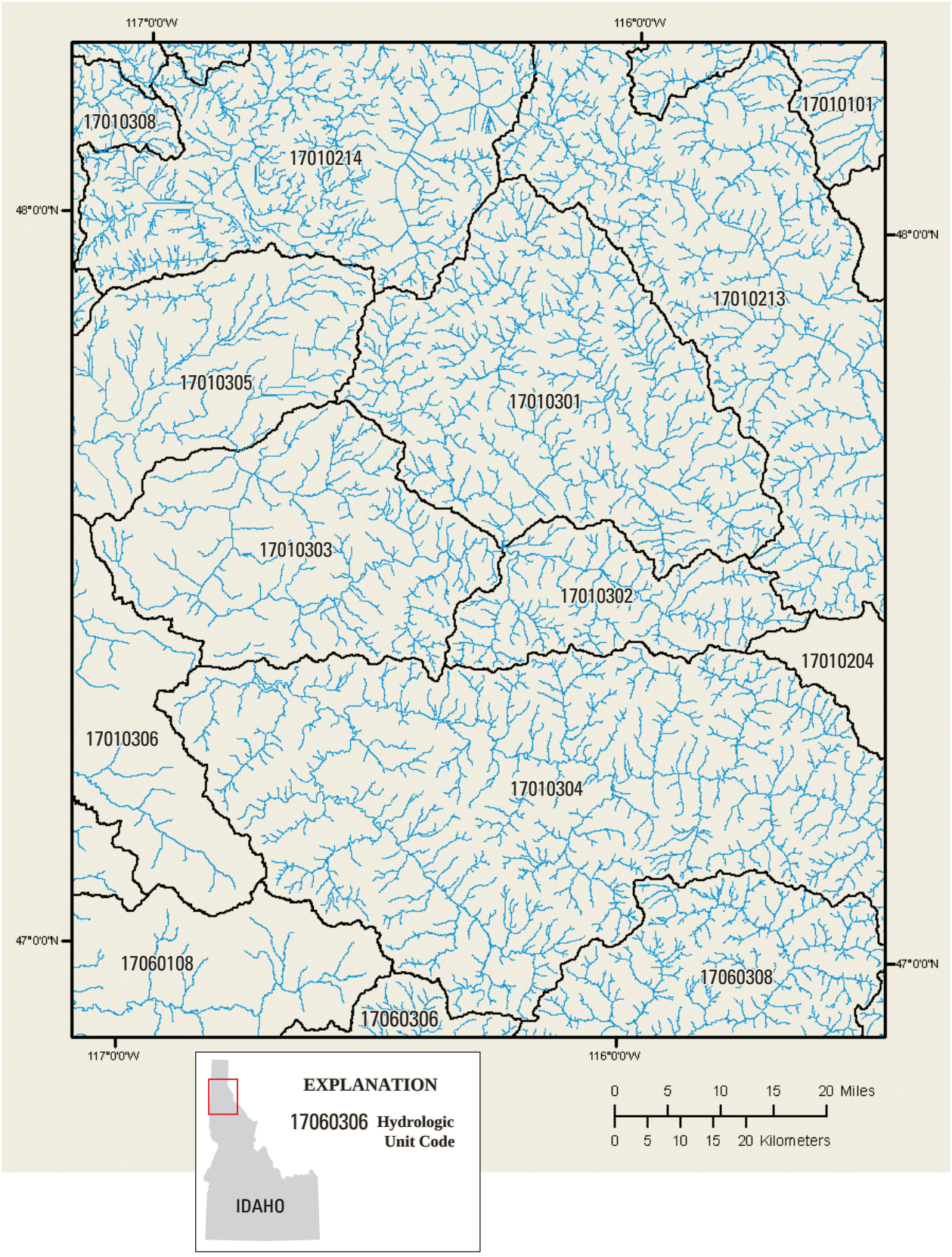


Figure 6. Screen capture of spatial distribution of modeled streams (blue) and Hydrologic Unit Code boundaries (black) from U.S. Geological Survey perennial streams model in northern Idaho.

Evaluation of Perennial Streams Model

The perennial streams model was evaluated by comparing classifications at various stream points on the map with four independent datasets from government agencies. A brief overview of these datasets is provided in table 3. Some stream classifications in the comparison datasets were based on single site visits and may not be representative of conditions throughout the year or over several years. Additionally, classification methods differed slightly among datasets. The locations of comparison points for each of the datasets are shown in figure 7.

The Beneficial Use Reconnaissance Program (BURP) dataset was provided by the IDEQ. The purpose of BURP is to help Idaho regulators meet the requirements of the Clean Water Act by collecting data for determining the existing uses and beneficial use support states of Idaho's water bodies (Idaho Department of Environmental Quality, 2008). Field visits were conducted in summer and autumn each year to collect aquatic community samples and aquatic habitat information. Field crews measured flow at a site if it was not dry. Sites were coded intermittent if dry or measured flow was less than 0.1 ft^3/s; otherwise, the site was coded as perennial.

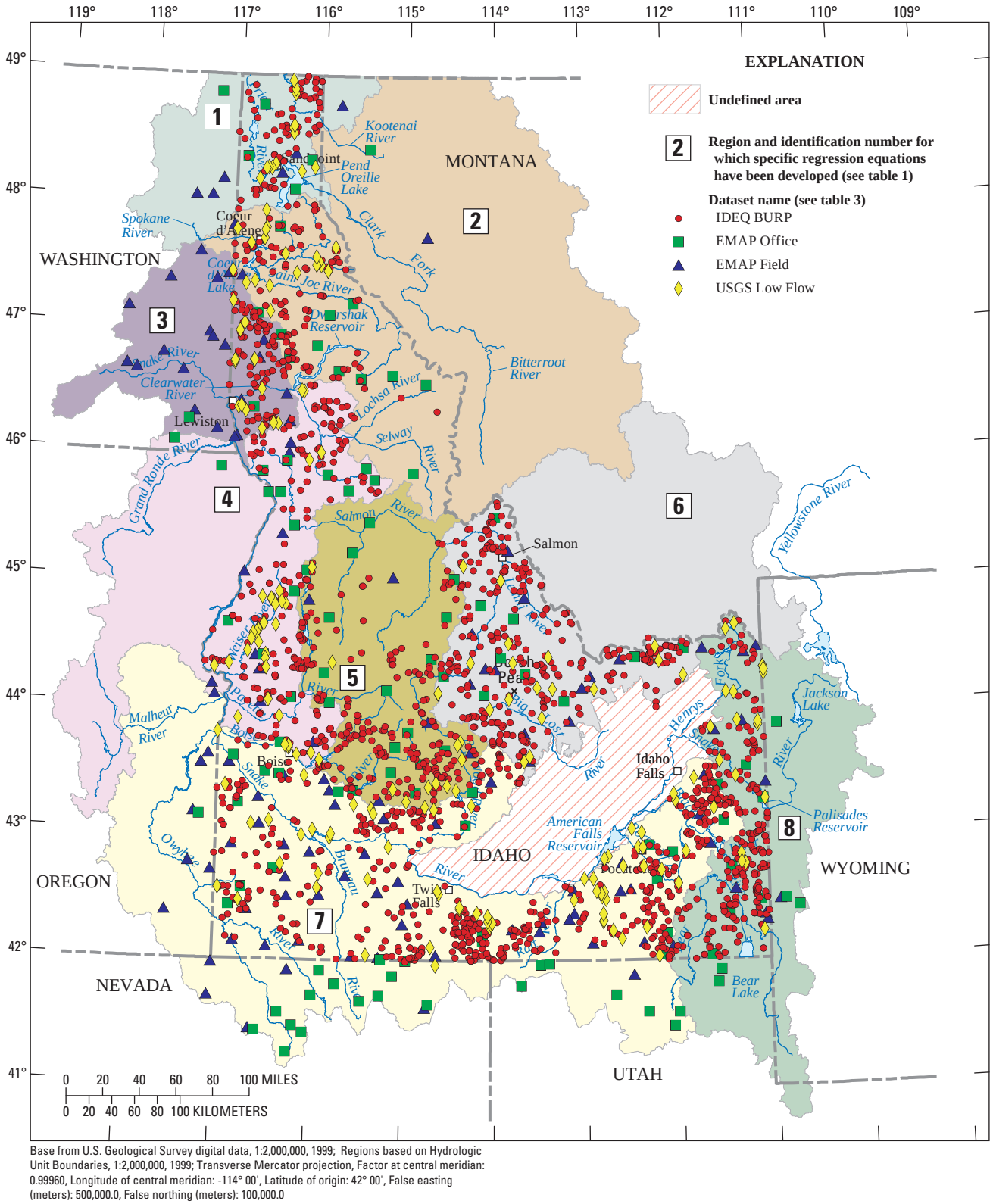
The Environmental Monitoring and Assessment Program (EMAP) datasets were provided by the USEPA. EMAP was developed to monitor and assess the status and trends of national ecological resources with a focus on forecasting future risks to those resources (U.S. Environmental Protection Agency, 2008). EMAP datasets used in this comparison are "EMAP Office" and "EMAP Field". USEPA personnel initially evaluated candidate stream sites using River Reach File 3 (RF3) (U.S. Environmental Protection Agency, 1994) and NHD maps and stream classifications. The EMAP Office dataset included stream sites that were coded as nonperennial on RF3/NHD maps and stream sites that were coded perennial but were not considered candidates for field visits because of limited access or other reasons. Streams in this dataset were then coded as perennial or intermittent using several office-based approaches, including examining maps and GIS coverages and talking to regional and state scientists with knowledge of the area. The EMAP Field dataset included sites that were coded as perennial on RF3/NHD maps and were deemed candidates for further field verification. After field visits, sites were coded perennial if there was enough water present to sample, even if flow was not continuous at the time of the visit. Otherwise, the site was coded intermittent. Field visits were made during the summer/autumn baseflow period, which was expected to be the minimum flow period for any given year.

The USGS Low Flow dataset includes stream sites that are measured periodically as part of the USGS Idaho Water Science Center partial record network used to supplement long-term streamflow-gaging records when generating peak and low flow statistics. Of the master partial record network database, sites were included in the comparison with the USGS perennial streams map if the site was dry, water was pooled, or measured flow was less than 0.1 ft^3/s (to be comparable with IDEQ BURP sites). All sites in this dataset are considered intermittent because they were observed to go dry or nearly dry annually. No attempt was made by the field hydrographer making the streamflow measurement to classify the stream as intermittent or perennial.

Table 3. Description of datasets used for evaluation of the U.S. Geological Survey perennial streams model for Idaho.

[**Abbreviations**: IDEQ BURP, Idaho Department of Environmental Quality Beneficial Use Reconnaissance Program; EMAP, Environmental Monitoring and Assessment Program; USGS, U.S. Geological Survey; ft^3/s, cubic foot per second; GIS, geographic information system; <, less than]

Dataset			Dates of field visits	Classification method	Number of sites
No.	Source	Name			
1	Idaho Department of Environmental Quality	IDEQ BURP	Summer/autumn 1993–2006	Streams were coded perennial if measured flow during field visits was greater than or equal to 0.1 ft^3/s; otherwise, streams were coded intermittent.	1,804
2	U.S. Environmental Protection Agency	EMAP office	No dates; determinations made in 1999	Made using several office-based approaches, including examining maps and GIS coverages and talking to regional biologists and State employees with knowledge of the area. No field visits.	114
3	U.S. Environmental Protection Agency	EMAP field	Summer/autumn 2000–2003	Streams were coded perennial if enough water was present to sample during field visits; otherwise, streams were coded intermittent.	123
4	U.S. Geological Survey	USGS low flow	Throughout year 1945–2001	Field visits made throughout the year, but mostly in summer/autumn. Sites were used for model comparison only if site was dry, water was pooled, or measured flow was <0.1 ft^3/s (to be comparable with BURP sites). Sites selected were considered intermittent.	207



Base from U.S. Geological Survey digital data, 1:2,000,000, 1999; Regions based on Hydrologic Unit Boundaries, 1:2,000,000, 1999; Transverse Mercator projection, Factor at central meridian: 0.99960, Longitude of central meridian: -114° 00', Latitude of origin: 42° 00', False easting (meters): 500,000.0, False northing (meters): 100,000.0

Figure 7. Locations of comparison points from datasets used to evaluate the U.S. Geological Survey perennial streams model for Idaho.

The comparison datasets were overlaid on the USGS perennial streams model. Comparison sites were forced to snap to the nearest USGS modeled stream (perennial or transitional), and coding of the comparison site was compared with the coding of the nearest USGS modeled stream. Sites were considered "agreements" or "disagreements" with the model based on the rules presented in table 4. All sites that were initially coded "disagreements" were checked to verify whether the comparison point snapped to the correct stream. In many cases, the comparison point was on a NHD stream not modeled by the USGS (for example, a true intermittent stream) but snapped to the closest USGS-modeled perennial stream. Once the correct location of the comparison point was verified, these cases were considered agreements with the model, because the USGS model was intended to include only perennial and transitional stream reaches. Comparison data points were removed from the analysis if their coordinates appeared grossly incorrect and no accompanying description was available to determine the correct location of the site. In addition, about 20 percent of the sites that were coded as agreements were checked to ensure that the point snapped to the correct USGS-modeled stream. Codings were revised if necessary.

A classification table (table 5) shows a comparison between the USGS model and comparison datasets. Overall, 81 percent of the comparison data points agreed with the USGS model according to the agreement rules shown in table 4. In table 5, the classification categories that are considered "disagreements" are: Observed intermittent stream—Predicted perennial stream (417 sites) and Observed perennial stream—Not mapped (4 sites). A comparison of each dataset with the USGS perennial streams model is presented by region in table 6. The total number of sites available for comparison and the percentage of total sites that agreed with the USGS model are shown for each dataset.

As a whole, the USGS model compared well with comparison datasets. In most of the disagreements, the USGS model overestimated the amount of perennial streams, meaning that many streams were predicted perennial when comparison datasets classified them as intermittent. Some disagreements were expected due to varying quality of site location coordinates (latitude/longitude), timing of site visits during unusually wet or dry years, discrepancy between 1-day site visits and a 7-day low flow criterion, and localized contributions of ground water to flow in some areas.

Table 4. Agreement rules for comparing independent datasets with the U.S. Geological Survey perennial streams model for Idaho.

[**Stream classification, USGS model:** Perennial, stream reach with $7Q_2$ flow greater than or equal to 0.1 ft^3/s; Transitional, stream reach where $7Q_2$ is 0.1 $ft^3/s \pm 1$ standard error; Not mapped, stream reach not on USGS model and assumed intermittent. **Stream classification, Comparison dataset:** See table 3. **Abbreviations**: USGS, U.S. Geological Survey; $7Q_2$, 7-day 2-year low flow, ft^3/s, cubic foot per second]

Stream classification		Coding
USGS model	Comparison dataset	
Perennial	Perennial	Agree
Transitional	Intermittent	Agree
Transitional	Perennial	Agree
Not mapped	Intermittent	Agree
Not mapped	Perennial	Disagree
Perennial	Intermittent	Disagree

Table 5. Classification table of accuracy of the U.S. Geological Survey perennial streams model relative to comparison datasets for Idaho.

[Shading denotes disagreements between the comparison dataset and the USGS model. **Abbreviation**: USGS, U.S. Geological Survey]

Observations (all comparison datasets)	Predictions from USGS model (number of sites)			Total number of sites
	Predicted perennial stream	Predicted transitional stream	Not mapped (assumed intermittent stream)	
Observed intermittent stream	417	399	171	987
Observed perennial stream	843	414	4	1,261
Total	1,260	813	175	2,248

Table 6. Agreements between the U.S. Geological Survey perennial streams model and comparison datasets by hydrologic region, Idaho and surrounding areas.

[See table 4 for agreement rules followed when determining "Percentage Agreement." **Abbreviations**: USGS, U.S. Geological Survey; BURP, Idaho Department of Environmental Quality Beneficial Use Reconnaissance Program; EMAP, Environmental Monitoring and Assessment Program]

Region	Comparison datasets							
	BURP		EMAP office		EMAP field		USGS low flow	
	Percentage agreement	Number of sites compared	Percentage agreement	Number of sites compared	Percentage agreement	Number of sites compared	Percentage agreement	Number of sites compared
1	80	76	80	5	33	6	8	13
2	87	115	100	12	100	2	32	19
3	89	92	100	3	86	22	40	15
4	83	178	88	16	100	10	75	40
5	81	233	81	16	60	5	8	12
6	84	316	90	10	67	12	50	24
7	87	561	93	41	84	55	63	64
8	82	233	73	11	55	11	45	20
Overall agreement	85	1,804	89	114	78	123	51	207

The USGS perennial streams model was generated based on long-term streamflow-gaging records and, with the inclusion of a transition zone upstream of the perennial stream endpoint, was expected to represent a range of climatic conditions. However, if a comparison dataset was generated solely during an unusually wet or dry period, a high number of disagreements with the model are expected. To evaluate climatic representativeness, the average Palmer Drought Severity Index (PDSI) for Idaho was examined for each year in the comparison dataset. PDSI generally ranges from -6 (extreme drought) and +6 (extreme wet), centering on 0 as a "normal" condition (Palmer, 1965).

The EMAP Office dataset compared better with the USGS model than the EMAP Field dataset (table 6). This was expected because the EMAP Field dataset was based on a single site visit, and the EMAP Office dataset was generated from professional opinion based on overall long-term knowledge of the area and map resources. Additionally, the EMAP Field dataset site visits were conducted during 2000–2003, when the statewide PDSI ranged from -1 to -4 (Cook and Krusic, 2008), an unusually dry period, and a higher number of disagreements were expected.

The highest number of disagreements with the USGS model was observed in the USGS Low Flow Network dataset. Most disagreements probably are due to poor latitude/longitude coordinates during some periods in the dataset and the timing of site visits. Because the dataset spans about 6 decades, global positioning system (GPS) receivers were not available to measure site latitude and longitude for the earlier measurements. Many coordinates were pulled from maps, which may not have been as accurate as GPS, and many of these comparison points did not include a stream name or description. In addition, the accuracy of GPS coordinates from March 1990 to May 2000 was hampered by selective availability, an intentional GPS signal degradation feature implemented by the U.S. Department of Defense for national defense purposes (National Executive Committee for Space-Based Positioning, Navigation, and Timing, 2008). Selective availability introduced slowly changing random errors of as much as 100 m in publicly-available GPS navigation signals (U.S Department of Commerce Technology Administration, 2000). Eight comparison points were measured during this period, and if site coordinates were poor, the location as plotted may be wrong and may indeed be located on a non-modeled (intermittent) part of a stream or entirely on another stream.

Like other comparison datasets, stream classification in the USGS Low Flow Network dataset was made based on a single assessment, not a 7-day duration of flow, and no attempt was made by hydrographer to classify the stream when in the field. Stream classification in this dataset was made solely for comparison with the USGS perennial streams model. Also, most disagreements occurred during years when the statewide PDSI was negative, indicating dry and, in some years, drought conditions. Of the sites that were considered disagreements, 80 percent were visited during years when the PDSI was negative according to Cook and Krusic (2008). About 25 percent of all site visits in this dataset were made in 1977, when the statewide PDSI ranged from -6 to -2 (Cook and Krusic, 2008).

These measurements were made during abnormally dry years and resulted in an intermittent classification when, during years of normal precipitation, the stream may flow year-round. Mean drainage area for the sites considered disagreements was relatively high (38.5 mi^2) in comparison with sites considered agreements (7.4 mi^2). Several opportunistic measurements were made at sites with drainage areas as large as 517 mi^2 during years of drought, resulting in disagreements with the model. Under normal hydrologic conditions, most of these sites were expected to agree with the model.

The BURP dataset represented a range of climate conditions and included the highest number of comparison sites. Statewide average PDSI ranged from -3.5 to 4.5 during the study period, 1993–2006 (Cook and Krusic, 2008). Eight years during the study period were considered "wet" years (positive PDSI) and six years were considered "dry" years (negative PDSI). The number of sites visited each year was fairly constant during the study period; therefore, the BURP dataset is considered the most representative dataset for comparison, and it agrees well with the model (85 percent agreement overall).

Overall, regions with the highest number of disagreements were 1, 5, 6, and 8, which have a high percentage of mountainous and forested area. These regions also include many areas of rapid transition from mountains to expansive valleys. In these areas, the USGS model over-predicted the number of perennial streams in comparison with the NHD and independent datasets. Highly variable ground water and surface water interactions, including streamflow losses, are documented in these regions in Donato (1998), Kahle and Bartolino (2007), and Skinner and others (2007). Streamflow losses to ground water along mountain fronts, where an interface can form in the surficial geology between residuum and more-permeable colluvium and alluvium, are well documented (Niswonger and others, 2005; Covino and McGlynn, 2007; Foster, 2008). Streams can change from perennial to intermittent at this interface, often called a mountain front recharge zone (Covino and McGlynn, 2007). The USGS model does not account for these site-specific hydrologic conditions and may not adequately represent true conditions in stream reaches where ground water gains and losses are highly variable.

Regions with the highest agreements were 2, 3, 4, and 7; all these regions except region 2 have a high percentage of low gradient, low elevation area and fewer mountain front recharge zones. However, most of the comparison sites in region 2 were in the lower gradient areas around Lake Coeur d'Alene and Lake Pend Oreille. These areas were expected to have more consistent ground water contributions to streamflow and less variable surficial geology, resulting in more consistent low flow patterns and a higher percentage of perennial streams. Surprisingly, the highest standard errors of prediction were for regions 3 and 7 in their respective regression equations; an explanation for these high standard errors is provided in the section titled "Model Revisions Based on Spatial Patterns".

Limitations and Areas for Further Study

The GLS regression equations were generated using streamflow-gaging station data and are applicable for stream sites in the study area with drainage areas ranging from 3.0 to 12,228 mi^2; however, this range varies by region. Ranges for other explanatory variables used to develop the regression equations are presented in table 2. The regression equations were applied outside the range of values in table 2 to predict endpoints of perennial flow along stream reaches. Such extrapolation cannot be avoided because of a lack of streamflow data for small streams. As a result, considerable uncertainty exists in the perennial streams model presented in this report, though agreement with independent datasets was good. The regression models provide overall estimates based on the general trends within each region.

Local factors such as a large spring or a losing reach are not included in these regression models, and these factors may greatly affect flows at any given point. Site-specific flow data, assuming a sufficient period of record, generally would be considered to better represent flow conditions at a given site than flow estimates based on regionalized regression models. However, this information rarely is available and is not always practical to obtain. The map is considered a first-cut, broad scale estimate of the locations of perennial streams that could be used for state-wide planning, assessment, and reporting purposes. The regression equations and perennial streams model presented are applicable only to unregulated, naturally-flowing streams.

Most studies to predict low-flow statistics at ungaged sites have relied on regression equations with basin characteristics such as drainage area, slope, elevation, and land cover. Permanence of flow in small streams is expected to depend on these characteristics as well as on ground water, or baseflow, contributions to the stream throughout the year. Baseflow recession information would be helpful in quantifying low-flow characteristics on a regional basis. Many studies have shown that the use of a baseflow recession constant or streamflow variability predictor greatly reduces the error of estimation of low-flow statistics (Vogel and Kroll, 1992; Yu and others, 2002; Kroll and others, 2004; Eng and Milly, 2007). However, baseflow information typically is gathered from long-term, continuous streamflow-gaging stations. Quantifying and predicting the baseflow recession constant (tau) at ungaged headwater sites, through procedures outlined in Eng and Milly (2007), may be an important element in improving accuracy in low flow statistics and stream classification.

In June 2006, the United States Supreme Court issued a ruling in the consolidated cases Rapanos vs. United States and Carabell vs. United States that changed the interpretation of stream classifications under the jurisdiction of the Clean Water Act. The ruling, called the Rapanos Decision, may eventually supersede many State and Federal stream management and regulatory guidelines that once were applied based on intermittent and perennial classifications. The new classification considers waters jurisdictional if they are

"relatively permanent", and possibly jurisdictional if "non-relatively permanent" when there is a "significant nexus" to an otherwise jurisdictional water (U.S. Environmental Protection Agency and U.S. Army Corps of Engineers, 2007). Streams are considered relatively permanent if flow is continuous on a seasonal basis (for example, 90 days or more). This report describes an example where streamflow information from the USGS streamflow-gaging network has been used to develop regression equations that predict streamflow characteristics at ungaged sites. However, under the Rapanos Decision, water quality standards may be applied to streams despite long durations of zero flow, and few to no long-term streamflow-gaging stations are located on such streams. Therefore, the existing USGS streamflow-gaging network cannot be used alone to extrapolate statistics to the level necessary to classify and map streams under the Rapanos guidance. A new flow-detection monitoring network and prediction methodology are needed to develop a map of relatively permanent streams to assist regulators and water-resource managers in making jurisdictional determinations. The map of perennial streams described in this report would be useful as a reference when selecting locations of flow-detection monitoring sites.

Summary

The U.S. Geological Survey (USGS), in cooperation with the Idaho Department of Environmental Quality (IDEQ) and Bureau of Reclamation, developed a map of streams with perennial flow in eight hydrologic regions in Idaho using generalized least-squares (GLS) regression models for estimating the 7-day, 2-year low flow ($7Q_2$) statistic and applying a 0.1 ft^3/s cutoff criterion. The regression equations and perennial streams model are applicable only to unregulated, naturally-flowing streams. A transition zone was modeled as plus or minus 1 standard error around the 0.1 ft^3/s criterion to account for changes in the perennial flow initiation point due to climatic, hydrologic, and statistical variability. In many cases, it was necessary to apply the regression equations well below the range of values used to develop the regression models. This extreme extrapolation produced unrealistic flow estimates and spatial patterns in the modeled stream network in the headwaters of some regions. As a result, alternative regression models for $7Q_2$ were generated for five of eight regions to produce realistic results in the range of extreme extrapolation. Standard errors of prediction were calculated for the regression equations to provide a measure of overall error and predictive power of the model within the range of input variables. Standard errors of prediction for the regression equations were lowest for region 6 (+75.0 to -42.9 percent) and highest for region 7 (+277 to -73.5 percent). Although the equations published by Hortness (2006) remain the preferred equations to use for generating streamflow statistics within the range of the original input variables, the selected alternative equations are preferred for extrapolating estimates of the $7Q_2$ in small drainage areas for the purpose of determining perennial flow initiation points.

The resulting $7Q_2$ equations were used in conjunction with a continuous parameter grid of basin characteristics to generate a geographic information system-based map of streams in Idaho with perennial flow according to the state definition of intermittent flow as less than 0.1 ft^3/s. The final map of perennial streams consists of synthetic lines or vector representations of streams from the gridded model and is provided in Environmental Systems Research Institute shapefile format.

The perennial streams model was evaluated by comparing classifications at various stream points on the map with four independent datasets from government agencies: IDEQ Beneficial Use Reconnaissance Program (BURP), Environmental Monitoring and Assessment Program (EMAP) Office, EMAP Field, and USGS Low Flow Network datasets. Overall, 81 percent of the comparison data points agreed with the USGS model. As a whole, the USGS model appeared to overestimate the amount of perennial streams in comparison with the independent datasets, meaning that many streams were predicted perennial when comparison datasets classified them as intermittent. Some disagreements between the model and comparison datasets were due to poor site location coordinates, timing of the comparison site visits during unusually wet or dry years, discrepancy between 1-day site visits and a 7-day low flow criterion, localized groundwater contributions to flow in some areas, and uncertainty due to model extrapolation. The IDEQ BURP dataset represented a range of climate conditions, including about equal numbers of wet and dry years, and included the highest number of comparison sites. Therefore, the BURP dataset is considered the most representative dataset for comparison, and it agrees well with the USGS model (85 percent agreement overall). Overall, regions with the highest number of disagreements have a high percentage of mountainous and forested area. The regions also include many mountain front recharge zones where streamflow losses to groundwater typically occur, possibly resulting in a change from perennial to intermittent flow. In these areas, the USGS model over-predicted the number of perennial streams in comparison with the National Hydrography Dataset and independent datasets. The USGS model does not account for these site-specific hydrologic conditions and may not adequately represent true conditions in stream reaches where groundwater gains and losses are highly variable. Regions with the highest agreements have a high percentage of low gradient, low elevation area and fewer mountain front recharge zones, resulting in more consistent low flow patterns that were well represented by the GLS regression equations. Site-specific flow data, assuming a sufficient period of record, generally would be considered to better represent flow conditions at a given site than flow estimates based on regionalized regression models; however, flow data on small headwater streams rarely are available and not always practical to obtain. The developed map is considered to be a first-cut, broad scale estimate of the locations of streams with perennial flow.

Acknowledgments

The authors wish to thank Don Essig with IDEQ for his ongoing interest, patience, and input in this project and for supplying the BURP comparison dataset. Special thanks also go to Barbara Rosenbaum and David Peck with USEPA for supplying the EMAP comparison dataset and to Tom Brennan and Jon Hortness with USGS for generating the USGS partial record network database. In addition, the authors wish to thank Steve Garcia with USGS for his assistance with model evaluation.

References Cited

Bent, G.C., and Steeves, P.A., 2006, A revised logistic regression equation and an automated procedure for mapping the probability of a stream flowing perennially in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2006-5031, 107 p.

Bond, J.G., 1978, Geologic map of Idaho: Idaho Bureau of Mines and Geology, 1 sheet, scale 1:500,000.

Clarke, S.E., Burnett, K.M., and Miller, D.J., 2008, Modeling streams and hydrogeomorphic attributes in Oregon from digital and field data: Journal of the American Water Resources Association, v. 44, no. 2, p. 459–477.

Cook, Edward and Krusic, Paul, 2008, North American Drought Atlas PDSI Reconstructions, Version 2a – Annual Maps, accessed May 8, 2008, at http://www.ncdc.noaa.gov/cgi-bin/paleo/pd08plot.pl

Covino, T.P., and McGlynn, B.L., 2007, Stream gains and losses across a mountain-to-valley transition—Impacts on watershed hydrology and stream water chemistry: Water Resources Research, v. 43, W10431.

Donato, M.M., 1998, Surface-water/ground-water relations in the Lemhi River Basin, East-Central Idaho: U.S. Geological Survey Water-Resources Investigations Report 98-4185, 20 p.

Eng, Ken, and Milly, P.C.D., 2007, Relating low-flow characteristics to the base flow recession time constant at partial record stream gauges: Water Resources Research, v. 43, W01201.

Environmental Systems Research Institute, Inc., 1999, Using ARC GRID with ARC/INFO, Version 2: Redlands, Calif., Environmental Systems Research Institute.

Fairfax County, Virginia, 2004, Perennial stream mapping project: Fairfax County, Virginia, Stormwater Planning Division, accessed April 3, 2008, at http://www.fairfaxcounty.gov/dpwes/watersheds/perennial.htm

Fritz, K.M., Johnson, B.R., and Walters, D.M., 2006, Field operations manual for assessing the hydrologic permanence and ecological condition of headwater streams: Washington, D.C., U.S. Environmental Protection Agency, Office of Research and Development, 151 p.

Foster, S.E., 2008, Groundwater and surface water interactions in streams of the Palouse Basin, northwestern Idaho and southeastern Washington, Paper No. 36-15 *in* Annual Meeting of the North-Central Section of the Geological Society of America, 42d, Evansville, IN, April 24-25, 2008, Proceedings: accessed May 27, 2008 at http://gsa.confex.com/gsa/2008NC/finalprogram/abstract_137876.htm

Hellweger, Ferdi, and Maidment, David, 1998, AGREE-DEM Surface Reconditioning System, accessed March 3, 2007, at http://www.crwr.utexas.edu/gis/gishyd98/quality/agree/agree.htm

Hortness, J.E., 2006, Estimating low-flow frequency statistics for unregulated streams in Idaho: U.S. Geological Survey Scientific Investigations Report 2006-5035, 31 p.

Hortness, J.E., and Berenbrock, Charles, 2001, Estimating monthly and annual streamflow statistics at ungaged sites in Idaho: U.S. Geological Survey Water Resources-Investigations Report 2001-4093, 36 p.

Idaho Department of Environmental Quality, 2008, Surface water—monitoring and assessment program (BURP and WBAG): Idaho Department of Environmental Quality, accessed April 15, 2008, at http://www.deq.idaho.gov/water/data_reports/surface_water/monitoring/overview.cfm

Jaeger, K.L., Montgomery, D.R., and Bolton, S.M., 2007, Channel and perennial flow initiation in headwater streams—management implications of variability in source-area size: Journal of Environmental Management, v. 40, p. 775–786.

Jenson, S.K., and Domingue, J.O., 1988, Extracting topographic structure from digital elevation data for geographic information system analysis: Photogrammetric Engineering and Remote Sensing, v. 54, no. 11, p. 1593–1600.

Kahle, S.C., and Bartolino, J.R., 2007, Hydrogeologic framework and ground-water budget of the Spokane Valley-Rathdrum Prairie aquifer, Spokane County, Washington, and Bonner and Kootenai Counties, Idaho: U.S. Geological Survey Scientific Investigations Report 2007-5041, 48 p., 2 pls.

Kroll, C.N., Luz, Joana, Allen, Brad, and Vogel, R.M., 2004, Developing a watershed characteristics database to improve low streamflow prediction: Journal of Hydrologic Engineering, v. 9, no. 2, p. 116–125.

Link, Paul, and Welhan, John, 2002, Digital Atlas of Idaho—Geology: Idaho Museum of Natural History, accessed April 3, 2008, at http://imnh.isu.edu/digitalatlas/.

Molnau, Myron, 1995, Mean annual precipitation, 1961–1990, Idaho: Moscow, University of Idaho, Agricultural Engineering Department, State Climate Program, scale 1:1,000,000, accessed May 10, 2008, at http://snow.ag.uidaho.edu/Climate/reports.html

National Executive Committee for Space-Based Positioning, Navigation, and Timing, 2008, website v. April 2008, accessed May 15, 2008, at http://pnt.gov/.

National Oceanic and Atmospheric Administration, 1985, Climate narrative of the States, v. 1, Alabama – New Mexico (3d ed.): accessed May 1, 2008, at http://www.wrcc.dri.edu/CLIMATEDATA.html

Niswonger, R.G., Prudic, D.E., Pohll, Greg, and Constantz, Jim, 2005, Incorporation seepage losses into the unsteady streamflow equations for simulating intermittent flow along mountain front streams: Water Resources Research, v. 41, W06006.

North Carolina Division of Water Quality, 2005, Identification methods for the origins of intermittent and perennial streams, v. 3.1: Raleigh, North Carolina Department of Environment and Natural Resources, Division of Water Quality, 40 p.

Olson, S.A., and Brouillette, M.C., 2006, A logistic regression equation for estimating the probability of a stream in Vermont having intermittent flow: U.S. Geological Survey Scientific Investigations Report 2006-5217, 15 p.

Palmer, W.C., 1965, Meteorological drought: U.S. Department of Commerce Weather Bureau Research Paper No. 45, 58 p.

Pope, B.F., Tasker, G.D., and Robbins, J.C., 2001, Estimating the magnitude and frequency of floods in rural basins of North Carolina—revised: U.S. Geological Survey Water-Resources Investigations Report 01-4207, 44 p.

Rea, Alan, and Skinner, K.D., 2009, Estimated perennial streams of Idaho and related geospatial datasets, U.S. Geological Survey Data Series 412, 32 p., available at http://pubs.usgs.gov/ds/412/.

Riggs, H.C., 1968, Some statistical tools in hydrology: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A1, 39 p.

Ross, S.H., and Savage, C.N., 1967, Idaho earth science: Moscow, Idaho, Idaho Bureau of Mines and Geology, 285 p.

Saunders, W.K., 2000, Preparation of DEMs for use in environmental modeling analysis, *in* Maidment, David and Djokic, Dean, eds., Hydrologic and hydraulic modeling support, Redlands, Calif., Environmental Systems Research Institute, Inc., p. 29-51.

Skinner, K.D., Bartolino, J.R., and Tranmer, A.W., 2007, Water-resource trends and comparisons between partial development and October 2006 hydrologic conditions, Wood River valley, south-central, Idaho: U.S. Geological Survey Scientific Investigations Report 2007-5258, 30 p.

State of Idaho, 2006, IDAPA 58.01.02—Water Quality Standards: Idaho Administrative Code, Department of Environmental Quality, 173 p.

U.S. Department of Commerce Technology Administration, 2000, Civilian benefits of discontinuing selective availability, May 1, 2000: U.S. Department of Commerce Technology Administration Press Release.

U.S. Environmental Protection Agency and U.S. Geological Survey, 1999, Standards for National Hydrography Dataset—High Resolution (Draft) (7/1999): U.S. Geological Survey database, accessed September 24, 2007, at http://rockyweb.cr.usgs.gov/nmpstds/nhdstds.html

U.S. Environmental Protection Agency and U.S. Army Corps of Engineers, 2007, Joint legal memorandum: Clean Water Act jurisdiction following U.S. Supreme Court's decision in Rapanos v. United States and Carabell v. United States: issued June 5, 2007, accessed August 15, 2007, at http://www.epa.gov/owow/wetlands/pdf/RapanosGuidance6507.pdf

U.S. Environmental Protection Agency, 1994, History of the U.S. EPA's River Reach File—a national hydrographic database available for ARC/INFO applications: U.S. Environmental Protection Agency, accessed August 26, 2008, at http://www.epa.gov/waters/doc/historyrf.pdf

U.S. Environmental Protection Agency, 2008, Environmental monitoring and assessment program: U.S. Environmental Protection Agency, accessed January 20, 2008, at http://www.epa.gov/emap/.

Vogel, R.M., and Kroll, C.N., 1992, Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics: Water Resources Research, v. 28, no. 9, p. 2451-2458.

Yu, P.S., Yang, T.C., and Liu, C.W., 2002, A regional model of low flow for southern Taiwan: Hydrologic Processes, v. 16, p. 2017–2034.

This page intentionally left blank.

Publishing support provided by the U.S. Geological Survey
Publishing Network, Tacoma Publishing Service Center

For more information concerning the research in this report, contact the
Director, Idaho Water Science Center
U.S. Geological Survey
230 Collins Road
Boise, Idaho 83702
http://id.water.usgs.gov