



Prepared in cooperation with the Vermont Center for Geographic Information

# **A Logistic Regression Equation for Estimating the Probability of a Stream in Vermont Having Intermittent Flow**

Scientific Investigations Report 2006–5217

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



# **A Logistic Regression Equation for Estimating the Probability of a Stream in Vermont Having Intermittent Flow**

By Scott A. Olson and Michael C. Brouillette

Prepared in cooperation with the  
Vermont Center for Geographic Information

Scientific Investigations Report 2006–5217

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

**U.S. Department of the Interior**  
DIRK KEMPTHORNE, Secretary

**U.S. Geological Survey**  
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia: 2006

For product and ordering information:

World Wide Web: <http://www.usgs.gov/pubprod>

Telephone: 1-888-ASK-USGS

For more information on the USGS--the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:

World Wide Web: <http://www.usgs.gov>

Telephone: 1-888-ASK-USGS

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

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Suggested citation:

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.

## Contents

Abstract.....	1
Introduction.....	1
Description of Study Area .....	2
Site Selection.....	2
Data Collection .....	2
Basin Characteristics Database Development.....	4
Logistic Regression Equation.....	10
Limitations of the Logistic Regression Equation.....	13
Application of the Logistic Regression Equation.....	14
Summary and Conclusions.....	14
Selected References.....	15

## Figures

1. Map showing location of sites observed for perennial and intermittent streamflow in Vermont.....	3
2–4. Graphs showing—	
2. Distribution of <i>A</i> , drainage areas. <i>B</i> , elevations. <i>C</i> , channel slopes. <i>D</i> , waterbody areas. <i>E</i> , ratios of basin relief to basin perimeter. <i>F</i> , areas of well- or moderately well-drained soils for intermittent and perennial streams observed in Vermont .....	9
3. Observed frequency of perennial and intermittent streams by drainage area in Vermont.....	10
4. Receiver operating characteristic curve for the logistic regression equation for estimating the probability of a stream having intermittent flow in Vermont.....	13

## Tables

1. Streamgages monitored during the data collection and their median and 25th and 75th percentiles of August daily flows for ensuring field observations occurred during normal streamflow magnitudes .....	5
2. Basin characteristics developed for testing in the logistic regression analysis.....	6
3. Diagnostics for explanatory variables used in the logistic regression equation for estimating the probability of a stream in Vermont having intermittent flow.....	11
4. Classification table depicting accuracy of the logistic regression equation predicting the perennial or intermittent status of the stream sites held out of the logistic regression analysis for equation verification .....	12
5. Classification table for the logistic regression equation for estimating the probability of a stream in Vermont having intermittent flow, with a probability cutpoint ranging from 0 to 1 .....	12
6. Estimated covariance matrix for the logistic regression equation .....	14
7. Ranges of the basin characteristics used in the logistic regression equation.....	14

## Conversion Factors and Datum

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

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

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

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).

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

# A Logistic Regression Equation for Estimating the Probability of a Stream in Vermont Having Intermittent Flow

By Scott A. Olson<sup>1</sup> and Michael C. Brouillette<sup>2</sup>

## Abstract

A logistic regression equation was developed for estimating the probability of a stream flowing intermittently at unregulated, rural stream sites in Vermont. These determinations can be used for a wide variety of regulatory and planning efforts at the Federal, State, regional, county and town levels, including such applications as assessing fish and wildlife habitats, wetlands classifications, recreational opportunities, water-supply potential, waste-assimilation capacities, and sediment transport. The equation will be used to create a derived product for the Vermont Hydrography Dataset having the streamflow characteristic of “intermittent” or “perennial.” The Vermont Hydrography Dataset is Vermont’s implementation of the National Hydrography Dataset and was created at a scale of 1:5,000 based on statewide digital orthophotos.

The equation was developed by relating field-verified perennial or intermittent status of a stream site during normal summer low-streamflow conditions in the summer of 2005 to selected basin characteristics of naturally flowing streams in Vermont. The database used to develop the equation included 682 stream sites with drainage areas ranging from 0.05 to 5.0 square miles. When the 682 sites were observed, 126 were intermittent (had no flow at the time of the observation) and 556 were perennial (had flowing water at the time of the observation).

The results of the logistic regression analysis indicate that the probability of a stream having intermittent flow in Vermont is a function of drainage area, elevation of the site, the ratio of basin relief to basin perimeter, and the areal percentage of well- and moderately well-drained soils in the basin. Using a probability cutpoint (a lower probability indicates the site has perennial flow and a higher probability indicates the site has intermittent flow) of 0.5, the logistic regression equation correctly predicted the perennial or intermittent status of 116 test sites 85 percent of the time.

## Introduction

The Vermont Hydrography Dataset (VHD) provides a consistent, statewide, local-resolution<sup>3</sup> hydrography data layer. This data layer has significantly improved spatial resolution compared to other currently available datasets, as well as data characterization based on the National Hydrography Dataset (NHD) (U.S. Geological Survey and U.S. Environmental Protection Agency, 2004) framework. However, the classification of the intermittent or perennial character of the streamflow, an important NHD attribute, was not conflated from the NHD data because it was based on cartographic rather than hydrologic methods of determination.

Headwater streams with small drainage areas commonly have annually intermittent flows. In the northeastern United States, however, the probability of zero-flow occurrences generally decreases as drainage area and basin storage increase in a downstream direction. This report, prepared by the U.S. Geological Survey (USGS) in cooperation with the Vermont Center for Geographic Information (VCGI), presents a logistic regression equation for estimating the probability that an unregulated, rural stream in Vermont flows intermittently during normal summer low-streamflow conditions.

The equation can be applied to a wide variety of regulatory and planning efforts at the Federal, State, regional, county and town levels, including such applications as assessing fish and wildlife habitats, wetlands classifications, recreational opportunities, water-supply potential, waste-assimilation capacities, and sediment transport. The equation will be used to create a derived product for the Vermont Hydrography Dataset having the streamflow characteristic of “intermittent” or “perennial” and will be disseminated on the VCGI Web site (<http://www.vcgi.org>).

<sup>1</sup> U.S. Geological Survey.

<sup>2</sup> Vermont Center for Geographic Information.

<sup>3</sup> Local resolution VHD created at a nominal scale of 1:5,000, based on statewide digital orthophotos created by the Vermont Mapping Program.

## Description of Study Area

Vermont comprises 9,615 mi<sup>2</sup> in the northeastern United States (fig. 1). Land altitudes range from about 100 ft along Lake Champlain in the northwestern part of the State to 4,393 ft in the north-central part of the State. The higher elevations dominate the central part of the State from north to south. The climate of Vermont is humid, with average annual precipitation ranging from about 35 to 40 in. in the Lake Champlain and Connecticut River Valleys to about 50 to 70 in. in the higher elevations. Average annual temperature is about 43°F.

## Site Selection

The primary goal of the site-selection process for this investigation was to obtain a random sampling of all naturally flowing rural stream sites draining less than 5 mi<sup>2</sup> that covered the ranges of basin characteristics that could be observed in Vermont. To maximize the number of sites where observations could be made statewide within a limited timeframe, another goal of the site-selection process was to choose sites that were easily accessible. To satisfy the accessibility criteria, potential sites were determined by intersecting the Vermont Enhanced 9-1-1 Board's roads coverage (State of Vermont Enhanced 9-1-1 Board, 2005a) with the Vermont Hydrography Dataset one-dimensional drain features. This was an attempt to ensure that all sites would be accessible by car. The drainage basin upstream of each potential site was delineated, and all sites having a drainage area smaller than 0.25 mi<sup>2</sup> or larger than 5 mi<sup>2</sup> were eliminated. For safety reasons, sites on interstate highways and other high traffic volume roads were eliminated. This resulted in 8,900 potential sites for field observation.

The target number of stream sites to be observed was 600. Simple random selection of 600 sites from the 8,900 potential sites was considered. Because the potential sites were found by intersecting streams and roads, and roads are commonly built on valley floors, there was a possibility that the mean elevation of 600 randomly selected sites would be biased low compared to typical Vermont streams. There was also a potential for a bias that most sites would be located in densely populated areas. To randomly select 600 sites from the 8,900 potential sites and reduce the biases in representation of basin characteristics that may be present in the potential sites, software for stratified random site selection (Scott, 1990) was utilized. The software, written in Arc Macro Language (AML) (Environmental Systems Research Institute, Inc., 1994), provided an approach for random selection of sites that represented combinations of magnitudes of basin characteristics.

Drainage area, site elevation, mean channel slope, mean basinwide annual precipitation, and areal location were the basin characteristics used in the stratified random selection process. In addition to the selected 600 sites, an alternate for

each site was also selected by the software should the primary site prove inaccessible in the field.

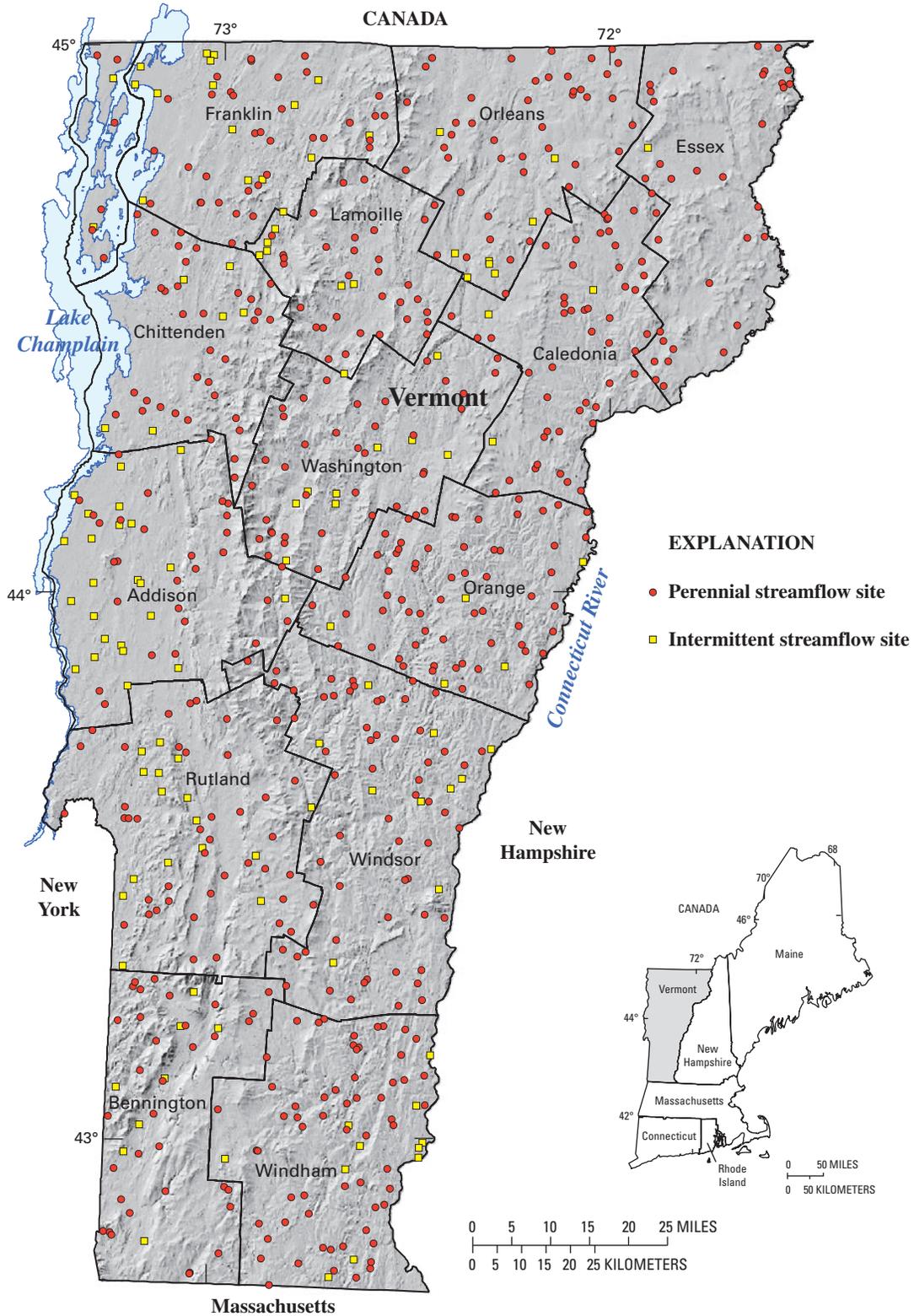
The selected 600 sites were examined for basin features that could prevent the stream from flowing naturally. If such a feature was found, the alternate site replaced the initially selected site. This review resulted in two sites being replaced because a dam was identified within its watershed, and five sites being replaced because of the urbanization or nonrural character of the basin. The decision to replace a site because of urbanization was subjective. Basins were identified as potentially urbanized on the basis of the number and density of emergency site locations identified within its watershed using the Emergency 9-1-1 Site Location database (State of Vermont Enhanced 9-1-1 Board, 2005b). Sites with more than 40 emergency site locations within its watershed were considered potentially urbanized. Potentially urbanized basins were then reviewed using all available mapping; if any urbanization or development that could have an impact on the natural flow character of the stream was apparent within a site's basin, the site was replaced by its alternate.

Following data collection in July and August 2005 at the 600 selected sites, about 90 percent of the sites had flow and were considered perennial for the purposes of this project. To better represent intermittent streams in the database, additional sites were selected that had drainage areas between 0.05 and 0.25 mi<sup>2</sup>. Because of the time constraints to complete the field observations, simple random selection was used to select 100 sites from a total of 9,101 available sites found intersecting GIS databases of roads and streams that had drainage areas between 0.05 and 0.25 mi<sup>2</sup>.

## Data Collection

An important aspect of the investigation was having the appropriate hydrologic and climatic conditions under which to observe sites for streamflow conditions. The intent of the investigation was to provide a methodology for estimating the probability that a stream flows intermittently during normal summer low streamflows—a condition that typically occurs in August or early September. Therefore, the selected sites needed to be observed for streamflow during normal summer low-streamflow conditions occurring from mid- to late summer. If hydrologic or climatic conditions were excessively wet, a site that typically goes dry and would be classified as intermittent could have streamflow at the time of observation and be classified as perennial. Conversely, if conditions were drier than normal, an observation site could be dry and would be classified as intermittent when, under normal hydrologic conditions, it is actually a site with perennial streamflow.

Sites were observed on selected dates from July 26 to August 26, 2005. Hydrologic and climatologic conditions during the data-collection period and the months leading up to the observation period were at or near normal. According to the Northeast Regional Climate Center (2006), precipitation



Base from Vermont Mapping Program digital elevation map, 1:5,000, 1994-2000  
 Universal Transverse Mercator projection, zone 18  
 Vermont State Plane coordinate system

**Figure 1.** Location of sites observed for perennial and intermittent streamflow in Vermont.

#### 4 A Logistic Regression Equation for Estimating the Probability of a Stream in Vermont Having Intermittent Flow

during the winter, spring, and summer in Vermont in 2005 was 111, 99, and 103 percent of normal, respectively, based on the period of record 1971 to 2000. During July 2005, precipitation was 0.18 in. above the normal July precipitation of 4.21 in. for the period of record. During August 2005, precipitation was 0.72 in. below the August normal of 4.49 in. for the period of record. The weekly Palmer Drought Severity Indices published by the National Weather Service (2006) were between -1.9 and 1.9 statewide for the entire period of data collection, an index range considered normal.

Although Vermont precipitation and drought indices were within normal ranges, streamflow conditions for June and July 2005 were reported by the USGS (2006a) as being predominantly above normal. Streamflows receded, however, and in August 2005, streamflow conditions were described as normal. Ground-water levels in June and July 2005 were predominantly in the normal range with some ground-water levels above and some below normal. By August 2005, water levels at 10 of the 13 ground-water stations in Vermont monitored by the USGS were within normal range, with the remaining 3 reported as being below normal. Normal streamflow or ground-water-levels are defined by the USGS (2006b) as being between the 25<sup>th</sup> and 75<sup>th</sup> percentile of streamflows or ground-water levels for the given month.

Overall, the statewide hydrologic and climatic conditions for the period of data collection—July 26 to August 26, 2005—can be considered at or near normal. Regardless, the climate and hydrologic conditions were monitored on a daily basis while making streamflow observations. To ensure that conditions were normal in the regions where observations were being made, additional criteria were followed when collecting data. These criteria included less than 0.4 in. of rain within 72 hours of a streamflow observation. Upon further review of precipitation data and the data-collection locations, no rain greater than 0.1 in. occurred within 24 hours of making an observation and no rain greater than 0.2 in. occurred within 48 hours of making an observation. Another criterion required that streamflows at the nearest or most hydrologically similar USGS streamgaging station to the observation site fall within the 25<sup>th</sup> and 75<sup>th</sup> percentile of August daily flows at the time of streamflow observation. Because streamflow data graphically reporting the percentiles in real time are available (U.S. Geological Survey, 2006c), this streamflow criterion was closely monitored. The median and the 25<sup>th</sup> and 75<sup>th</sup> percentile of August daily flows for gages on streams with little to no low-streamflow regulation monitored during the data collection are listed in table 1.

As stated above in this report, all observation sites were located at road crossings to expedite data collection. A global positioning system (GPS) was used to pinpoint the site. Streamflow conditions were observed at the upstream side and downstream side of the road crossing. A site was considered perennial if it had any flowing water at the time of observation. Sites that had dry channel beds or disconnected pools of water were considered intermittent. If a site had pooled water throughout the reach, it was considered to be a

perennial stream site unless a positive determination could be made that water was not flowing.

Several sites could not be reached because a roadway was either gated or impassable. Other sites were deemed inappropriate as an observation site because of the altered condition of the stream, such as a stream site within a golf course or a site where an artificial pond had been created upstream to store water. When a site was visited that was not appropriate or could not be reached, it was replaced by its alternate site. In some cases, the alternate sites had similar problems and could not be used. When this occurred, both the site and its alternate were removed from the database. At the end of the data collection effort, streamflow observations at 682 sites (fig. 1) were used to develop the logistic regression equation for estimating the probability that a stream has intermittent flow. Of the 682 sites, 126 were classified as intermittent and 556 were classified as perennial. The distribution of the observation sites is shown in figure 1.

### Basin Characteristics Database Development

Regression analysis involves developing a relation between a response variable and one or more explanatory variables. In this investigation, the basin characteristics of a stream site were used as the explanatory variables. Fifty basin characteristics for the 682 stream sites were determined for use in the logistic regression analysis. The basin characteristics that were computed and the data sources used to compute the basin characteristics are listed in table 2.

Basin delineation and the subsequent characterization effort required a Digital Elevation Model (DEM) that was spatially compliant with the local resolution VHD dataset. Barring the existence of such a dataset at any scale, a statewide “hydrologically enforced” DEM tiled by subbasin and other supporting derivative raster datasets, such as flow accumulation and flow direction, were created. The “hydrologically enforced” DEM was created with the ArcInfo TOPOGRID command and the following inputs: point elevation data from the Vermont Mapping Program digital orthophoto program; and one- and two-dimensional VHD features, such as stream centerlines and lakes and ponds.

The basin characteristics used in the logistic regression analysis were generated using ArcInfo® (v.8.0.2) on a UNIX platform and AML programming (Environmental Systems Research Institute, Inc., 1994). The basic approach to computing each characteristic revolved around the use of the site as a watershed outlet point, and its basin extent to control the extent of processing for computing basin characteristics.

Selected basin characteristics of the 682 streamflow observation sites were grouped by their intermittent or perennial classification and plotted. The plots in figures 2 and 3 display the distribution of selected basin characteristics by

**Table 1.** Streamgages monitored during the data collection and their median and 25th and 75th percentiles of August daily flows for ensuring field observations occurred during normal streamflow magnitudes.[Streamgages located in Vermont unless otherwise noted; mi<sup>2</sup>, square miles; ft<sup>3</sup>/s, cubic feet per second]

Station number	Streamgage	Drainage area, mi <sup>2</sup>	Median of August daily flows, ft <sup>3</sup> /s	25th percentile of August daily flows, ft <sup>3</sup> /s	75th percentile of August daily flows, ft <sup>3</sup> /s
01133000	East Branch Passumpsic River near East Haven	53.8	52.8	33.9	71.9
01134500	Moose River at Victory	75.2	56.3	27	87.1
01135150	Pope Brook near North Danville	3.25	2.6	1.7	3.8
01135300	Sleepers River near St. Johnsbury	42.9	30.4	16.9	56.5
01135500	Passumpsic River at Passumpsic	436	281	211	419
01139000	Wells River at Wells River	98.4	44.6	34	78.7
01139800	East Orange Branch at East Orange	8.95	4.3	2.7	6.6
01142500	Ayers Brook at Randolph	30.5	10	6.5	18.3
01144000	White River at West Hartford	690	264	194	477
01150900	Ottawaquechee River near West Bridgewater	23.4	17.8	9.3	37.5
01153550	Williams River near Rockingham	112	38.1	28	63.2
01154000	Saxtons River near Saxtons River	72.2	19.1	11	29.2
01169000	North River at Shattuckville, Massachusetts	89.0	35.1	22.6	63.3
01170100	Green River near Colrain, Massachusetts	41.4	17.6	11.6	37.6
01334000	Walloomsac River near North Bennington	111	79.7	61.9	129
04280350	Mettawee River near Pawlet	70.2	33	21.1	71.7
04282000	Otter Creek at Center Rutland	307	187	138	283
04282525	New Haven River at Brooksville	115	91.8	50.8	140
04282650	Little Otter Creek at Ferrisburg	57.1	11	6.3	25.1
04282795	Lewis Creek near North Ferrisburg	77.2	31.2	15.9	39
04282795	LaPlatte River at Shelburne Falls	44.6	7.5	2.9	16.9
04287000	Dog River at Northfield Falls	76.1	25.9	17.5	40.9
04288000	Mad River near Moretown	139	67.1	45.8	112.5
04292000	Lamoille River at Johnson	310	203	154	292
04293000	Missisquoi River near North Troy	131	82.7	54.1	162.2
04293500	Missisquoi River near East Berkshire	479	268	171	440
04294300	Pike River at East Franklin	34.5	11.5	6.2	81.4
04296000	Black River at Coventry	122	70.2	45.5	112

stream classification. Some interesting information can be determined from looking at the plots. For instance, figure 2 shows that the largest drainage area of an intermittent stream in this investigation is about 3.4 mi<sup>2</sup>. It also shows the median drainage area, site elevation, and area of well- or moderately well-drained soils are different between perennial and intermittent stream sites. In figure 3, the frequency of intermittent

sites consistently decreases from a drainage area of 0.05 mi<sup>2</sup> (the minimum drainage area used in this investigation) to a point where there are very few intermittent sites at 1.0 mi<sup>2</sup> and greater. Also, the number of intermittent and perennial sites with drainage areas between 0.05 and 0.25 mi<sup>2</sup> are nearly equal—above 0.25 mi<sup>2</sup>, the number of perennial stream sites greatly outnumbers the intermittent stream sites (fig. 3).

**Table 2.** Basin characteristics developed for testing in the logistic regression analysis.

[OFR, Open-File Report; PRISM, Parameter-elevation Regressions on Independent Slopes Model; SSURGO, Soil Survey Geographic Database; USGS, U.S. Geological Survey; UVM, University of Vermont; VCGI, Vermont Center for Geographic Information; VT, Vermont; --, not applicable]

Characteristic description	Units	Source	Minimum	Mean	Maximum
Compactness ratio <sup>1</sup>	dimensionless	Digital Elevation Model <sup>10</sup>	1.3	1.7	2.6
Relative relief <sup>2</sup>	feet/foot	Digital Elevation Model <sup>10</sup>	0.00150	0.0313	0.111
Basin azimuth <sup>3</sup>	degrees from north	Digital Elevation Model <sup>10</sup>	1.0	180.9	360.0
Basin length (basin centerline from divide to basin outlet)	feet	Digital Elevation Model <sup>10</sup>	1,235.0	7,122.9	22,226.0
Percent of basin above 1,200 feet	areal percentage	Digital Elevation Model <sup>10</sup>	0	56	100
Percent of basin above 1,500 feet	areal percentage	Digital Elevation Model <sup>10</sup>	0	35	100
Percent of basin above 1,800 feet	areal percentage	Digital Elevation Model <sup>10</sup>	0	16	100
Mean basin slope	feet/foot	Digital Elevation Model <sup>10</sup>	1.5	16.3	44.1
Percent of basin having greater than 10 degree slope	areal percentage	Digital Elevation Model <sup>10</sup>	.1	64.7	99.1
Percent of basin having greater than 15 degree slope	areal percentage	Digital Elevation Model <sup>10</sup>	.0	45.1	96.6
Stateplane x-coordinate of basin centroid	meters	Digital Elevation Model <sup>10</sup>	428,525	487,553	576,115
Stateplane y-coordinate of basin centroid	meters	Digital Elevation Model <sup>10</sup>	28,375	172,572	278,365
Land cover characteristics - Agricultural	areal percentage	2002 Land Use/Land Cover Dataset <sup>11</sup>	.0	15.3	94.1
Land cover characteristics - Barren land	areal percentage	2002 Land Use/Land Cover Dataset <sup>11</sup>	.0	.1	13.8
Land cover characteristics - Forested	areal percentage	2002 Land Use/Land Cover Dataset <sup>11</sup>	.0	71.2	99.5
Land use characteristics - Industrial	areal percentage	2002 Land Use/Land Cover Dataset <sup>11</sup>	.0	3.3	75.8
Land use characteristics - Residential	areal percentage	2002 Land Use/Land Cover Dataset <sup>11</sup>	.0	4.2	54.8
Mean annual precipitation	inches	PRISM Precipitation Data <sup>12</sup>	32	45	73
Mean August precipitation	inches	PRISM Precipitation Data <sup>12</sup>	3.6	4.6	7.5
Mean July precipitation	inches	PRISM Precipitation Data <sup>12</sup>	3.3	4.3	7.0
Mean September precipitation	inches	PRISM Precipitation Data <sup>12</sup>	3.29	4.06	7.02
Mean basin elevation	feet	Digital Elevation Model <sup>10</sup>	107	1,281	2,937
Main channel length	feet	Vermont Hydrography Dataset <sup>13</sup>	587	2,993	9,763
Main channel slope	percent rise	Digital Elevation Model <sup>10</sup>	.1	4.2	22.0
Elevation of observation point	feet	Digital Elevation Model <sup>10</sup>	91	934	2,340
Relief <sup>4</sup>	feet	Digital Elevation Model <sup>10</sup>	39	808	2,938
Basinwide mean annual runoff	inches	USGS OFR 96-395 <sup>14</sup>	9.80	22.44	39.00
Surficial geology: grouping A <sup>5</sup>	areal percentage	VCGI In-House Dataset <sup>15</sup>	.0	2.73	84.20
Surficial geology: grouping B <sup>6</sup>	areal percentage	VCGI In-House Dataset <sup>15</sup>	.0	3.46	99.60
Surficial geology: grouping A and B <sup>5,6</sup>	areal percentage	VCGI In-House Dataset <sup>15</sup>	.0	6.19	99.60

**Table 2.** Basin characteristics developed for testing in the logistic regression analysis.—Continued

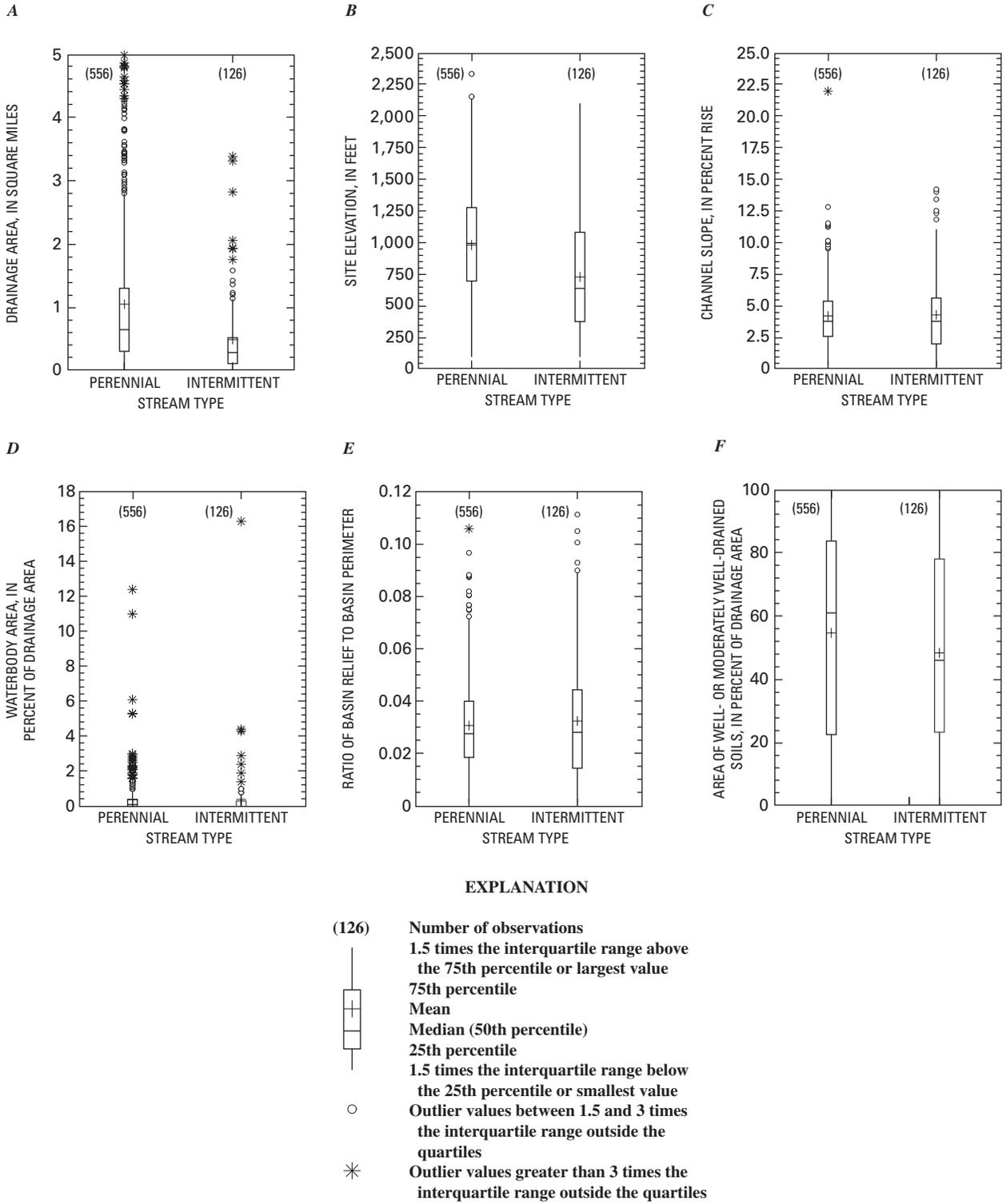
[OFR, Open-File Report; PRISM, Parameter-elevation Regressions on Independent Slopes Model; SSURGO, Soil Survey Geographic; USGS, United State Geological Survey; UVM, University of Vermont; VCGI, Vermont Center for Geographic Information; VT, Vermont; --, not applicable]

Characteristic description	Units	Source	Minimum	Mean	Maximum
Number of first order streams	count	Vermont Hydrography Dataset <sup>13</sup>	1	4.11	77
Stream order at observation point	stream order	Vermont Hydrography Dataset <sup>13</sup>	1	1.92	5
Total length of first order streams	miles	Vermont Hydrography Dataset <sup>13</sup>	0.0	1.7	14.3
Total length of second order streams	miles	Vermont Hydrography Dataset <sup>13</sup>	.0	0.7	8.1
Total length of third order streams	miles	Vermont Hydrography Dataset <sup>13</sup>	.0	.2	4.0
Soil characteristics - drainage; excessive <sup>7</sup>	areal percentage	SSURGO <sup>16</sup>	.0	17.3	99.3
Soil characteristics - drainage; poor <sup>8</sup>	areal percentage	SSURGO <sup>16</sup>	.0	14.9	95.5
Soil characteristics - drainage; well <sup>9</sup>	areal percentage	SSURGO <sup>16</sup>	.0	53.4	99.9
Soil characteristics - infiltration - hydrologic group A	areal percentage	SSURGO <sup>16</sup>	.0	4.7	95.7
Soil characteristics - Infiltration - hydrologic group B	areal percentage	SSURGO <sup>16</sup>	.0	12.9	85.0
Soil characteristics - Infiltration - hydrologic group C	areal percentage	SSURGO <sup>16</sup>	.0	43	100
Soil characteristics - Infiltration - hydrologic group D	areal percentage	SSURGO <sup>16</sup>	.0	28.2	99.3
Soil characteristics - hydraulic conductivity - permeability	inches/hour	SSURGO <sup>16</sup>	.0	121.8	999.0
Total length of streams	feet	Vermont Hydrography Dataset <sup>13</sup>	64	14,162	152,876
Drainage area	square miles	Digital Elevation Model <sup>10</sup>	.05	.959	4.99
Geologic-biophysical province	province	UVM-VT Biodiversity Project <sup>17</sup>	--	--	--
Waterbody area (lake and ponds)	areal percentage	Vermont Hydrography Dataset <sup>13</sup>	.0	.4	16.3
Wetland area - lacustrine	areal percentage	National Wetlands Inventory <sup>18</sup>	.0	.1	12.0
Wetland area - palustrine	areal percentage	National Wetlands Inventory <sup>18</sup>	.0	2.3	37.4
Wetland area - lacustrine and palustrine	areal percentage	National Wetlands Inventory <sup>18</sup>	.0	2.4	37.4

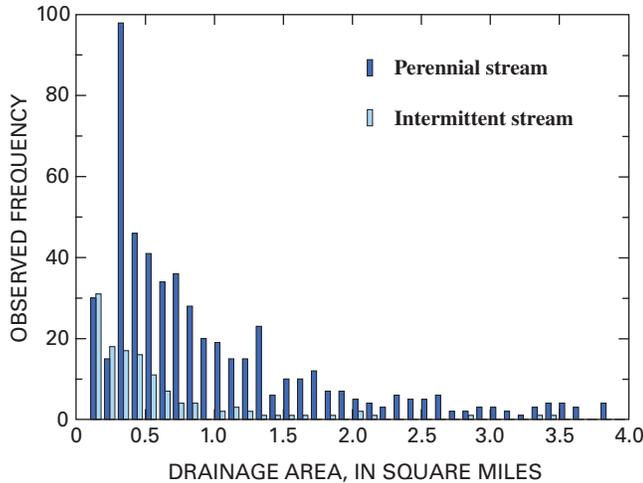
**Table 2.** Basin characteristics developed for testing in the logistic regression analysis.—Continued

[OFR, Open-File Report; PRISM, Parameter-elevation Regressions on Independent Slopes Model; SSURGO, Soil Survey Geographic; USGS, United States Geological Survey; UVM, University of Vermont; VCGI, Vermont Center for Geographic Information; VT, Vermont; --, not applicable]

- <sup>1</sup> Ratio of the perimeter of the basin to the perimeter of a circle having the area of the basin.
- <sup>2</sup> Ratio of basin relief to basin perimeter.
- <sup>3</sup> Line from point where main channel, when extended, intersects basin divide to the basin outlet.
- <sup>4</sup> Maximum basin elevation minus the minimum basin elevation.
- <sup>5</sup> Group A consists of beach gravel, delta gravel, delta sand, eolian sand, esker, fluvial, gravel, kame, kame terrace, kame moraine, lake gravel and outwash.
- <sup>6</sup> Group B consists of fluvial sand, lake sand, lake sand veneered with fluvial sand, moraine, pebbly sand, recent alluvium, and terminal moraine.
- <sup>7</sup> Excessive includes soils grouped in SSURGO as excessively and somewhat excessively well drained.
- <sup>8</sup> Poor includes soils grouped in SSURGO as somewhat poorly and poorly drained.
- <sup>9</sup> Well includes soils grouped in SSURGO as well and moderately well drained.
- <sup>10</sup> Vermont Department of Taxes, Vermont Mapping Program (2001).
- <sup>11</sup> University of Vermont Spatial Analysis Lab (2002).
- <sup>12</sup> Daly (2000).
- <sup>13</sup> Vermont Center for Geographic Information, Inc. (2004).
- <sup>14</sup> Randall (1996).
- <sup>15</sup> M.C. Brouillette, Vermont Center for Geographic Information, written commun., 2005.
- <sup>16</sup> U.S. Department of Agriculture, National Resource Conservation Service (2005).
- <sup>17</sup> Capen (1998).
- <sup>18</sup> U.S. Fish and Wildlife Service (2003).



**Figure 2.** Distribution of *A*, drainage areas. *B*, elevations. *C*, channel slopes. *D*, waterbody areas. *E*, ratios of basin relief to basin perimeter. *F*, areas of well- or moderately well-drained soils for intermittent and perennial streams observed in Vermont.



**Figure 3.** Observed frequency of perennial and intermittent streams by drainage area in Vermont.

## Logistic Regression Equation

Regression has been used in many water-resource investigations that examine the relation between a water-resource characteristic and one or more characteristics of its watershed or environment. The variables used to develop the regression equations are commonly continuous variables, and a least-squares or linear regression model is developed. In this investigation, the response variable being modeled is not continuous. The response variable is discrete—it is either a “1” for intermittent, or “0” for perennial. Logistic regression is the standard method of analysis when working with discrete response variables (Hosmer and Lemeshow, 2000). The logistic regression equation that is developed estimates the probability that the response variable is a “1” or that the stream is intermittent. Other investigations have used logistic regression to predict the intermittent or perennial classification of a stream site; most notable is a study by Bent and Archfield (2002) for the State of Massachusetts.

Before the logistic regression analysis began, 20 percent of the observations sites were randomly removed from the dataset and not used in the analysis for testing the predictive accuracy of the final logistic regression equation. This resulted in a database of 546 sites used to complete the regression analysis. All potential explanatory variables—including the basin characteristics (table 2), and the logarithms, squares, and square roots of the basin characteristics—for the stream observations sites were tested using SAS statistical software (SAS Institute, Inc., 1989). The explanatory variables were evaluated using forward selection, backward elimination, stepwise selection, and best subsets selection techniques to determine the most significant explanatory variables for intermittent flow.

The following tools and techniques were also used for selecting the final variables to use in the logistics regression equation. Logit plots and partial residual plots (SAS Institute, Inc., 2001) were created for variables. Logit plots reveal whether or not there is a linear relation between the predictor variable and the response variable and will indicate if the predictor variable needs to be treated as a discrete variable. Partial residual plots indicate whether or not there is a linear relation between the predictor variable and the response variable and show what kind of transformation the variable may need, such as a logarithmic transformation, to get the predictor variable and response variable to have a linear relation. Index plots (SAS Institute, Inc., 2001) were created to identify influential observations and outliers in the datasets, and *p*-values were used to select the final variables for the equation. *P*-values indicate the significance of predictor variables in predicting the response variable. A *p*-value of about 0.01 indicates a positive significance; a *p*-value of about 0.001 indicates a strong significance; and a *p*-value of 0.0001 or less indicates a very strong significance (SAS Institute, Inc., 2001) when sample sizes are similar to the sample size found in this study.

During the final explanatory variable selection process, the soil characteristics data from the Soil Survey Geographic (SSURGO) database (U.S. Department of Agriculture, 2005) were determined to be significant in predicting the probability that a stream has intermittent flow. As of the fall of 2005, however, the SSURGO was not yet available in Caledonia or Essex Counties—an area covering about 14 percent of the State. Because the SSURGO data could improve the logistic regression equations, stream sites without SSURGO data available were removed from the regression analysis. This left 492 stream sites available to develop the final regression equation.

The Geologic-Biophysical Regions dataset, which divides Vermont into eight subregions on the basis of bedrock geology, gross physiography, climate, and broad-scale patterns of potential natural vegetation (Capen, 1998), was used to categorize sites into regions. Tests of regionalizing the sites based on this dataset indicated that a location variable such as this does have the potential to improve the predictive accuracy of the equation. But this dataset, and others that could be used for regionalization of sites, have not been developed at a scale and level of detail that was appropriate for this investigation.

The final four basin characteristics that were found to be significant in estimating the probability that a stream has intermittent flow were (1) drainage area; (2) elevation of the site; (3) percentage of the basin with soils classified as well- or moderately well-drained soils in the SSURGO database; and (4) the relative relief, which is a measure of basin slope. A logistic regression equation was created with the final four explanatory variables using SAS statistical software as follows:

$$p = 1 / \{1 + \exp[-5.30 + 1.98(\log_{10}(A)) + 2.38(\log_{10}(E)) + 0.525(\log_{10}(S + 1)) - 165(R^2)]\}, \quad (1)$$

where

- $p$  is the probability that a stream has intermittent flow;
  - $A$  is the drainage area, in square miles;
  - $E$  is the site elevation, in feet;
  - $S$  is the percentage of the basin with soils classified as well- or moderately well-drained soils in the SSURGO database;
- and
- $R$  is the relative relief, unitless (ratio of maximum relief to basin perimeter).

Diagnostics of the parameters used in this regression equation are presented in table 3. The  $p$ -values—ranging from less than 0.0001 to 0.018—indicate that all four of the explanatory variables are significant in predicting the response variable. The Variance Inflation Factors (VIF)—ranging from 1.00 to 1.18—indicate how many times greater the variance of the variable would be in the absence of collinearity. A VIF near 1.0 would indicate no collinearity (Allison, 2000).

The predictive accuracy of equation 1 was tested using 116 of the 136 sites randomly left out of the database for equation verification purposes. Not all 136 sites could be used for testing because 20 sites were in Essex or Caledonia Counties, which did not have the SSURGO soils data available. A probability cutpoint of 0.5 was used in this test. Therefore, if the probability,  $p$ , estimated by the logistic regression equation was greater than or equal to 0.5, the stream was predicted to have intermittent flow. If the probability,  $p$ , was less than 0.5, the stream was predicted to have perennial flow. The incorrect and correct responses from the regression equations are listed in the classification table (table 4). Using the results shown in this table, the sensitivity [ratio of the number of streams correctly predicted to be intermittent (8) to the total number of intermittent streams (23)] was 34.8 percent and the specificity [ratio of the number of streams correctly predicted to be perennial (91) to the total number of perennial streams (93)] was 97.8 percent. Overall, the equation correctly predicted the perennial or intermittent characteristic of the test sites 85.3 percent of the time.

A classification table (table 5) was also computed using SAS statistical software with probability cutpoints ranging from 0 to 1, using the sites that were used in developing the logistic regression equation. As shown in table 5, the sensitivity and the specificity of the logistic regression equation at a cutpoint of 0.50 is 28.3 and 95.7 percent, respectively. The table also shows that sensitivity and specificity can be optimized between a cutpoint of 0.15 and 0.20. In fact, sensitivity equals specificity at a cutpoint of approximately 0.19. At this cutpoint, there is a balance between the percentage of sites correctly predicted to be perennial or intermittent. There is a trade-off using 0.19 as the probability cutpoint; however, increasing the percentage of correct intermittent classification by lowering the cutpoint to 0.19 from 0.50 decreases the percentage of correct perennial stream predictions. The percentage of total correct responses, perennial and intermittent, at a cutpoint of 0.50 was more than 10 percent greater than at a cutpoint of 0.19 and, thus, 0.50 is the optimal cutpoint.

According to table 5, there is a slight (0.4 percent) improvement in the total number of correct predictions at a cutpoint of 0.55. However, using the sites held out of the logistic regression analysis for equation verification—the same sites represented in table 4—the percentage of correct predictions decreases from 85.3 to 83.6 percent when the cutpoint is increased from 0.50 to 0.55. Therefore, increasing the cutpoint to 0.55 is not recommended.

The data in the classification table shown in table 5 were used to develop a receiver operating characteristic (ROC) curve (fig. 4). This curve graphically depicts the predictive accuracy of the logistic regression equation. For a logistic model with high predictive accuracy, the ROC curve raises quickly and has a larger area under the ROC curve (SAS Institute, Inc. 2001). Hosmer and Lemeshow (2000) suggest that if the area under the ROC curve is between 0.8 and 0.9, the model has excellent discrimination; if the area under the curve is between 0.7 and 0.8, the model has acceptable discrimination; and if the area under the curve is 0.5, the model has no discrimination and is no better than flipping a coin. The area under the ROC curve computed for the logistic regression equation 1 is 0.79.

**Table 3.** Diagnostics for explanatory variables used in the logistic regression equation for estimating the probability of a stream in Vermont having intermittent flow.

[SAS Institute, Inc., 1989; --, not applicable; log<sub>10</sub>, base 10 logarithm]

Explanatory variable	Parameter estimate	95-percent confidence limits for parameter estimates		$p$ -value	Variance inflation factor
Intercept	5.30	2.89	7.71	<0.0001	--
Log <sub>10</sub> (drainage area)	-1.98	-2.61	-1.35	<.0001	1.03
Log <sub>10</sub> (site elevation)	-2.38	-3.22	-1.54	<.0001	1.18
Log <sub>10</sub> (areal percent well- or moderately well-drained soils+1)	-0.524	-.951	-0.0983	.016	1.00
Relative relief <sup>2</sup>	165	28.3	302	.018	1.18

<sup>2</sup>, Relative relief squared.

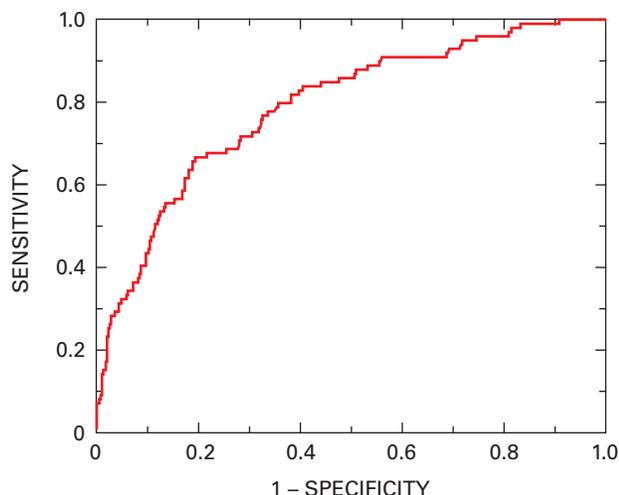
**Table 4.** Classification table depicting accuracy of the logistic regression equation predicting the perennial or intermittent status of the stream sites held out of the logistic regression analysis for equation verification.

	Predicted intermittent stream	Predicted perennial stream	Total
Observed intermittent stream	8	15	23
Observed perennial stream	2	91	93
Total	10	106	116

**Table 5.** Classification table for the logistic regression equation for estimating the probability of a stream in Vermont having intermittent flow, with a probability cutpoint ranging from 0 to 1.

[SAS Institute, Inc. (1989); Cutpoint is the probability level that could be used to classify the stream as intermittent or perennial. Sensitivity is the ratio of correctly classified intermittent streams to the observed number of intermittent streams. Specificity is the ratio of correctly classified perennial streams to the observed number of perennials streams. False positive is the ratio of sites incorrectly predicted to be intermittent to all sites predicted to be intermittent. False negative is the ratio of sites incorrectly predicted to be perennial to all sites predicted to be perennial]

Cutpoint	Total correct		Total incorrect		Percent				
	Intermittent	Perennial	Intermittent	Perennial	Correct	Sensitivity	Specificity	False positive	False negative
0.00	99	0	393	0	20.1	100	0.0	79.9	0.0
0.05	95	77	316	4	35.0	96.0	19.6	76.9	4.9
0.10	87	171	222	12	52.4	87.9	43.5	71.8	6.6
0.15	79	248	145	20	66.5	79.8	63.1	64.7	7.5
0.20	67	286	107	32	71.7	67.7	72.8	61.5	10.1
0.25	61	321	72	38	77.6	61.6	81.7	54.1	10.6
0.30	55	331	62	44	78.5	55.6	84.2	53.0	11.7
0.35	42	348	45	57	79.3	42.4	88.5	51.7	14.1
0.40	34	358	35	65	79.7	34.3	91.1	50.7	15.4
0.45	32	370	23	67	81.7	32.3	94.1	41.8	15.3
0.50	28	376	17	71	82.1	28.3	95.7	37.8	15.9
0.55	24	382	11	75	82.5	24.2	97.2	31.4	16.4
0.60	17	385	8	82	81.7	17.2	98.0	32.0	17.6
0.65	13	386	7	86	81.1	13.1	98.2	35.0	18.2
0.70	8	389	4	91	80.7	8.1	99.0	33.3	19.0
0.75	6	391	2	93	80.7	6.1	99.5	25.0	19.2
0.80	2	393	0	97	80.3	2.0	100	0.0	19.8
0.85	1	393	0	98	80.1	1.0	100	0.0	20
0.90	0	393	0	99	79.9	0.0	100	0.0	20.1
0.95	0	393	0	99	79.9	0.0	100	0.0	20.1
1.00	0	393	0	99	79.9	0.0	100	0.0	20.1



**Figure 4.** Receiver operating characteristic curve for the logistic regression equation for estimating the probability of a stream having intermittent flow in Vermont.

The lower and upper confidence limits of the predicted probability of intermittent flow are dependent on the basin characteristics input into the equation and can be computed by the following equations (SAS Institute, Inc., 1989):

$$P_l = 1 / \{1 + \exp[-5.30 + 1.98(\log_{10}(A)) + 2.38(\log_{10}(E)) + 0.525(\log_{10}(S+1)) - 165(R^2) + z_{\alpha/2}SE]\}, \quad (2)$$

and

$$P_u = 1 / \{1 + \exp[-5.30 + 1.98(\log_{10}(A)) + 2.38(\log_{10}(E)) + 0.525(\log_{10}(S+1)) - 165(R^2) - z_{\alpha/2}SE]\}, \quad (3)$$

where

- $P_l$  is the lower confidence limit of the predicted probability;
- $P_u$  is the upper confidence limit of the predicted probability;
- $A$  is the drainage area, in square miles;
- $E$  is the site elevation, in feet;
- $S$  is the percentage of the basin with soils classified as well- or moderately well-drained soils in the SSURGO database;
- $R$  is the relative relief, unitless (ratio of maximum relief to basin perimeter);
- $z_{\alpha/2}$  is the percentile point of a standard normal distribution at alpha level  $\alpha$  ( $\alpha=0.10$  for a 90-percent prediction interval);

and

$SE$  is the standard error of prediction and can be computed as

$$SE = [x_i V_b x_i']^{1/2}, \quad (4)$$

where

$x_i$  is a row vector containing a 1,  $\log_{10}(A)$ ,  $\log_{10}(E)$ ,  $\log_{10}(S+1)$ , and  $R^2$ ;

$V_b$  is the estimated covariance matrix of the parameter estimates, as shown in table 6;

and

$x_i'$  is the transpose of the  $x_i$  matrix.

## Limitations of the Logistic Regression Equation

The logistic regression equation 1 in this report is applicable only to sites on unregulated, naturally flowing streams in rural Vermont drainage basins. It cannot be applied to streams that have a drainage basin in which there are surface-water withdrawals, diversions, or flow augmentation. Users of the logistic regression equation also must be aware of substantial ground-water withdrawals in a drainage basin and evaluate the effect of such withdrawals on streamflow before applying the equation. Use of the equation should be limited to sites with basin characteristics that are within the range of explanatory variables used in the development of the equation. The ranges of the basin characteristics used in the analysis are shown in table 7. If explanatory variables used in the logistic regression equation are outside of the ranges shown in the table, the probability prediction is an extrapolation and the result could be unreliable. Furthermore, the confidence limits of the predicted probability cannot be defined if the probability prediction is an extrapolation.

Basin characteristics used to develop the logistic regression equation 1 were determined with a Geographic Information System (GIS) using specific databases as described earlier in this report. Determining basin characteristics for use in the regression equation with alternative data sources or methods could introduce bias and thus produce erroneous results. This limitation prevents the use of the logistic regression equation at sites in Essex and Caledonia Counties as the SSURGO database currently does not include these two counties. In the future, if the SSURGO data for these two counties become available, the applicability of the equation should be tested, because it was developed with sites that had SSURGO data outside of those counties.

The observations of streamflow were completed in a short time frame (20 non-consecutive days) during the summer of 2005. Although overall hydrologic conditions in Vermont were considered to be normal summer conditions, accurately quantifying normal streamflow at each stream observation site was outside of the scope of this study.

**Table 6.** Estimated covariance matrix for the logistic regression equation.

	Intercept	Drainage area	Site elevation	Areal percent well- or moderately well-drained soils +1	Relative relief
Intercept	1.513714	0.009184	-0.50145	-0.08015	23.59203
Drainage area	0.009184	.103271	.011219	-.00139	2.460786
Site elevation	-.50145	.011219	.185337	.002103	-10.9912
Areal percent well- or moderately well-drained soils +1	-.08015	-.00139	.002103	.047288	0.364138
Relative relief	23.59203	2.460786	-10.9912	.364138	4,861.633

**Table 7.** Ranges of the basin characteristics used in the logistic regression equation.

Basin characteristic	Minimum	Mean	Maximum
Drainage area, in square miles	0.05	0.959	4.99
Site elevation, in feet	91	934	2,340
Areal percent well- or moderately well-drained soils	.0	53.4	99.9
Relative relief, unitless	.00150	.0313	.111

### Application of the Logistic Regression Equation

An example application of the logistic regression equation for estimating the probability of a site having intermittent flow is provided using a site that has a drainage area of 0.25 mi<sup>2</sup> and a site elevation of 948 ft. The percentage of the basin with well- or moderately well-drained soils is 9.6 percent and the ratio of the basin relief (1,037 ft) to basin perimeter (15,880 ft) is 0.0653 ft/ft. Equation 1 is solved for this site as follows:

$$p = 1 / \{ 1 + \exp[-5.30 + 1.98(\log_{10}(0.25)) + 2.38(\log_{10}(948)) + 0.525(\log_{10}(9.6 + 1)) - 165(0.0653^2)] \}$$

The resulting probability, *p*, that this stream site has intermittent flow is 0.39. This probability is less than the 0.5 cutpoint probability; therefore, the site would be classified as a perennial streamflow site.

To determine the lower and upper confidence limits for the probability, the standard error, *SE*, must be computed using equation 4. To solve equation 4, the *x<sub>i</sub>* vector for this site,

$$x_i = [1, \log_{10}(0.25), \log_{10}(948), \log_{10}(9.6 + 1), 0.0653^2],$$

and the covariance matrix, *V<sub>b</sub>*, shown in table 4, is needed. Solving the matrix multiplication of equation 4 results in a value of 0.0578 so that the standard error can be computed

by *SE*=(0.0578)<sup>0.5</sup>. Thus, the standard error, *SE*, is 0.240 and equations 2 and 3 reduce to  $1 / \{ 1 + \exp[0.427 \pm z_{\alpha/2} 0.240] \}$ . If the 90-percent confidence interval is desired, then  $\alpha=1-0.9$  and *z<sub>α/2</sub>*, determined from a standard normal distribution, is 1.6449. Therefore, the lower and upper 90-percent confidence limits for this site are 0.305 and 0.492, respectively.

### Summary and Conclusions

A logistic regression equation for estimating the probability that an unregulated, rural stream in Vermont has intermittent flow was developed by the U.S. Geological Survey, in cooperation with the Vermont Center for Geographic Information. The database used to develop the logistic regression equation consisted of 492 streamflow observations completed at 682 sites in July and August of 2005 during normal summertime hydrologic conditions and the basin characteristics of each site. The final logistic regression equation incorporates drainage area, elevation of the observation site, percentage of the basin containing well- and moderately well-drained soils, and relative relief (the ratio of the maximum relief of the basin to its perimeter) as explanatory variables. The equation correctly predicted the perennial or intermittent classification of the stream site 85 percent of the time at a 0.50 cutpoint. The cutpoint is the probability level used to classify the stream as either intermittent or perennial.

The logistic regression equation is applicable to sites on unregulated, naturally flowing streams in rural Vermont drainage basins. It does not apply to streams that have a drainage basin in which there are surface-water withdrawals, diversions, or flow augmentation. Users of the logistic regression equation also must be aware of ground-water withdrawals in a drainage basin and determine if the withdrawals could affect streamflow before applying the equation. Because the SSURGO soils database did not include Essex and Caledonia Counties at the time of this investigation, the applicability of the logistic regression equation in those counties will need to be evaluated as the data become available.

## Selected References

- Allison, P.D., 2000, Logistic regression using the SAS system, theory and application: Cary, N.C., SAS Institute, Inc., 287 p.
- Bent, G.C., and Archfield, S.A., 2002, A logistic regression equation for estimating the probability of a stream flowing perennially in Massachusetts: U.S. Geological Survey Water-Resources Investigations Report 02–4043, 45 p.
- Capen, D.E., 1998, Biophysical regions of Vermont from the VT Biodiversity Project: Waterbury, Vt., Vermont Center for Geographic Information, Inc., accessed April 4, 2006, at <http://www.vcgi.org>
- Daly, Christopher, 2000, PRISM (Parameter-elevation regressions on independent slopes model), U.S. Department of Agriculture and Natural Resources Conservation Service Climate Mapping Project, United States Average Monthly or Annual Precipitation, 1971–2000: National Water and Climate Center, Oregon State University Partnership.
- Environmental Systems Research Institute, Inc., 1994, ARC/INFO users guide, version 7.0: Redlands, Calif., variously paged.
- Hosmer, D.W., and Lemeshow, Stanley, 2000, Applied logistic regression (2d ed.): New York, John Wiley, 375 p.
- National Weather Service, 2006, Palmer Drought Severity index, accessed July 14, 2006, at [http://www.cpc.ncep.noaa.gov/products/monitoring\\_and\\_data/drought.shtml](http://www.cpc.ncep.noaa.gov/products/monitoring_and_data/drought.shtml).
- Northeast Regional Climate Center, 2006, Monthly and seasonal climate summary tables, accessed June 29, 2006, at [http://met-www.cit.cornell.edu/climate/Climate\\_summary.html](http://met-www.cit.cornell.edu/climate/Climate_summary.html).
- Randall, A.D., 1996, Mean annual runoff, precipitation, and evapotranspiration in the glaciated northeastern United States, 1951–80: U.S. Geological Survey Open-File Report 96–395, 2 pls.
- SAS Institute, Inc., 1989, SAS/STAT user's guide, vers. 6, 4th ed., v. 2: Cary, N.C., SAS Institute, Inc., 846 p.
- SAS Institute, Inc., 2001, Categorical data analysis using logistic regression: Cary, N.C., SAS Institute, Inc., 364 p.
- Scott, J.C., 1990, Computerized stratified random site-selection approaches for design of a ground-water-quality sampling network: U.S. Geological Survey Water-Resources Investigations Report 90–4101, 17 p.
- State of Vermont Enhanced 9-1-1 Board, 2005a, Emergency E911\_RDS, accessed April 25, 2006, at <http://www.vcgi.org/>.
- State of Vermont Enhanced 9-1-1 Board, 2005b, Emergency E911\_ESITE, accessed April 25, 2006, at <http://www.vcgi.org/>.
- University of Vermont Spatial Analysis Lab, 2002, 2002 Land cover/land use dataset for Vermont, accessed April 26, 2006, at <http://www.vcgi.org/>.
- U.S. Department of Agriculture, Natural Resources Conservation Service, Soil Survey Division, 2005, Soil survey geographic database (SSURGO), National SSURGO database, accessed March 22, 2006, at <http://www.ncgc.nrcs.usda.gov/products/datasets/ssurgo/index.html>
- U.S. Fish and Wildlife Service, 2003, National wetlands inventory, accessed April 26, 2006, at <http://www.vcgi.org/>.
- U.S. Geological Survey, 2002, National elevation dataset, accessed April 26, 2006, at <http://ned.usgs.gov/>.
- U.S. Geological Survey and U.S. Environmental Protection Agency, 2004, National hydrography dataset, accessed May 5, 2004, at <http://nhd.usgs.gov/>.
- U.S. Geological Survey, 2006a, Precipitation and drought data, accessed July 14, 2006, at <http://water.usgs.gov/cgi-bin/waterwatch>.
- U.S. Geological Survey, 2006b, Streamflows and ground-water level data, accessed on July 14, 2006, at [http://vt.water.usgs.gov/WaterData/all\\_months.htm](http://vt.water.usgs.gov/WaterData/all_months.htm).
- U.S. Geological Survey, 2006c, Real-time streamflow data, accessed July 14, 2006, at <http://waterdata.usgs.gov/vt/nwis/rt>.
- Vermont Center for Geographic Information, Inc., 2004, The Vermont hydrography dataset, accessed April 4, 2006, at <http://www.vcgi.org/>.
- Vermont Department of Taxes, Vermont Mapping Program, 2001, Digital Elevation Models: Waterbury, Vt., Vermont Mapping Program.

Prepared by Publications Service Center 1

For more information concerning the research in this report, contact:

Keith W. Robinson, Director

U.S. Geological Survey

New Hampshire-Vermont Water Science Center

361 Commerce Way

Pembroke, NH 03275

or visit our Web site at:

<http://nh.water.usgs.gov>

