science for a changing world

In cooperation with the
VERMONT AGENCY OF TRANSPORTATION

A Geographic Information System Tool to Solve Regression Equations and Estimate Flow-Frequency Characteristics of Vermont Streams

Open-File Report 02-494

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

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2003	**N/A**	**-**

4. TITLE AND SUBTITLE
A Geographic Information System Tool to Solve Regression Equations and Estimate Flow-Frequency Characteristics of Vermont Streams

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Department of the Interior 1849 C Street, NW Washington, DC 20240

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**30**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

A Geographic Information System Tool to Solve Regression Equations and Estimate Flow-Frequency Characteristics of Vermont Streams

By Scott A. Olson, Gary D. Tasker, and Craig M. Johnston

U.S. GEOLOGICAL SURVEY

Open-File Report 02-494

Prepared in cooperation with the

VERMONT AGENCY OF TRANSPORTATION

Pembroke, New Hampshire
2003

U.S. DEPARTMENT OF THE INTERIOR

GALE A. NORTON, *Secretary*

U.S. GEOLOGICAL SURVEY

Charles G. Groat, *Director*

For additional information write to:

District Chief
U.S. Geological Survey
New Hampshire/Vermont District
361 Commerce Way
Pembroke, NH 03275-3718
http://nh.water.usgs.gov

CONTENTS

FIGURES

TABLES

CONVERSION FACTORS AND WEB SITES

CONVERSION FACTORS

Multiply	By	To obtain
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
square mile (mi^2)	2.589	square kilometer (km^2)
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

WEB SITES USEFUL FOR OBTAINING DATA AND INFORMATION PERTINENT TO THE VERMONT FLOW-FREQUENCY TOOL

Web Site	Owner	Source of
http://www.cits.rncan.gc.ca	Natural Resources Canada	Canadian elevation datasets
http://www.vcgi.org	Vermont Center for Geographic Information	Topographic images of Vermont
http://nhd.usgs.gov	U.S. Geological Survey	National Hydrography Dataset
http://edcnts12.cr.usgs.gov/ned	U.S. Geological Survey	National Elevation Dataset
http://vt.water.usgs.gov/CurrentProjects/vtfloodfreq/index.htm	U.S. Geological Survey, Vermont District	Vermont Flow-Frequency Tool
http://water.usgs.gov	U.S. Geological Survey, Water Resources Division	National Flood Frequency Program
http://www.climatesource.com	Climatesource	PRISM Precipitation Datasets

A Geographic Information System Tool to Solve Regression Equations and Estimate Flow-Frequency Characteristics of Vermont Streams

By Scott A. Olson, Gary D. Tasker, *and* Craig M. Johnston

ABSTRACT

Estimates of the magnitude and frequency of streamflow are needed to safely and economically design bridges, culverts, and other structures in or near streams. These estimates also are used for managing floodplains, identifying flood-hazard areas, and establishing flood-insurance rates, but may be required at ungaged sites where no observed flood data are available for streamflow-frequency analysis. This report describes equations for estimating flow-frequency characteristics at ungaged, unregulated streams in Vermont.

In the past, regression equations developed to estimate streamflow statistics required users to spend hours manually measuring basin characteristics for the stream site of interest. This report also describes the accompanying customized geographic information system (GIS) tool that automates the measurement of basin characteristics and calculation of corresponding flow statistics. The tool includes software that computes the accuracy of the results and adjustments for expected probability and for streamflow data of a nearby stream-gaging station that is either upstream or downstream and within 50 percent of the drainage area of the site where the flow-frequency characteristics are being estimated. The custom GIS can be linked to the National Flood Frequency program, adding the ability to plot peak-flow-frequency curves and synthetic hydrographs and to compute adjustments for urbanization.

INTRODUCTION

Estimates of the magnitude and frequency of streamflow are needed to safely and economically design bridges, culverts, and other structures in or near streams. These estimates also are used for managing floodplains, identifying flood-hazard areas, and establishing flood-insurance rates, but may be required at ungaged sites where no observed flood data are available for streamflow-frequency analysis. The U.S. Geological Survey (USGS), in cooperation with the Vermont Agency of Transportation (VTrans), developed regression equations for estimating peak-flow frequency for recurrence intervals of 2, 5, 10, 25, 50, 100, and 500 years and estimates of the 25, 50, and 75 percent flow duration (daily discharges exceeded 25, 50, and 75 percent of the time) for ungaged, unregulated streams in Vermont. Generalized least squares regression (Tasker, 1987) was used to relate basin characteristics to flow-frequency characteristics for the peak-flow equations. Ordinary least squares regression was used to develop the flow-duration equations. A report by Olson (2002) documents the development of those equations.

To solve the flow-frequency equations accurately and quickly, a computer application was developed to automate the delineation of drainage-basin characteristics required by the equations. A geographical information system (GIS) application, named the Vermont Flow-Frequency Tool, was created to automate basin-characteristic delineation. The National Hydrography Dataset (NHD) (U.S. Geological Survey and U.S. Environmental Protection Agency, 2002) tools are the foundation for the GIS tool customized for Vermont. The tool includes software that was written in the computer language, FORTRAN, to solve the regression equations, compute the accuracy of the results, and determine adjustments for expected probability and for data from a nearby stream-gaging station. A nearby station is defined as a station that is either upstream or downstream and within 50 percent of the drainage area of the site where the flow-frequency characteristics are being estimated. The Vermont Flow-Frequency Tool can be linked to the National Flood Frequency (Ries and Crouse, 2002) program, providing the abilities to plot peak-flow-frequency curves and synthetic hydrographs and compute adjustments for urbanization.

The purpose of this report is to describe (1) the equations for estimating flow-frequency characteristics, (2) an error analysis of the equations and (3) the Vermont Flow-Frequency Tool.

The authors wish to thank the following for their help or guidance in developing the GIS datasets and(or) programming of the computer application: Jennifer R. Hill of Horizon Systems Corporation, Lynn C. Bjorklund, Laura Hayes, David S. Nail, Kernell G. Ries, Keven S. Roth, and Peter A. Steeves of the U.S. Geological Survey.

ESTIMATING FLOW-FREQUENCY CHARACTERISTICS OF UNREGULATED, UNGAGED, RURAL VERMONT STREAMS

Generally, the most reliable estimates of flow-frequency characteristics are obtained by frequency analysis of flow data from gaged streams. Estimates, however, often are required at stream locations where no data are available. To solve this problem, regression equations were developed for estimating flood discharges at ungaged, unregulated streams in rural drainage basins in Vermont by relating flow-frequency characteristics to basin characteristics at gaged locations. Details of the development of the equations can be found in the report, “Flow-Frequency Characteristics of Vermont Streams” by Olson (2002).

Regression Equations

The equations for estimating flood-flow frequencies were developed using generalized least squares (GLS) regression methods (Tasker, 1987) and the GLSNET software (G.D. Tasker, K.M. Flynn, A.M. Lumb, W.O. Thomas, Jr., U.S. Geological Survey, written commun., 1995). Tasker and Stedinger (1989) demonstrated that GLS regression provides the most appropriate results for hydrologic purposes because the principles of GLS allow weight to be given to each station in the analysis to compensate for differences in length of record and for streamflow data that are correlated spatially and in time. The equations for estimating flow duration (daily-mean discharges exceeded a given percentage of time) were developed using ordinary least squares (OLS) regression procedures (SAS Institute Inc., 1990) because period-of-record flow-duration statistics lack the annual time series required by GLSNET software.

The equations used to estimate peak flow at recurrence intervals of 2, 5, 10, 25, 50, 100, and 500 years and the magnitude of daily-mean discharges exceeded 25, 50, and 75 percent of the time are

$$Q_2 = 38.1A^{0.914}L^{-0.294}E^{0.0776}Y^{-0.180}, \qquad (1)$$

$$Q_5 = 61.8A^{0.902}L^{-0.295}E^{0.0835}Y^{-0.253}, \qquad (2)$$

$$Q_{10} = 79.7A^{0.897}L^{-0.302}E^{0.0890}Y^{-0.298}, \qquad (3)$$

$$Q_{25} = 106A^{0.883}L^{-0.316}E^{0.104}Y^{-0.349}, \qquad (4)$$

$$Q_{50} = 129A^{0.874}L^{-0.327}E^{0.115}Y^{-0.385}, \qquad (5)$$

$$Q_{100} = 153A^{0.865}L^{-0.336}E^{0.125}Y^{-0.420}, \qquad (6)$$

$$Q_{500} = 217A^{0.846}L^{-0.355}E^{0.148}Y^{-0.497}, \qquad (7)$$

$$D_{75} = 0.000627A^{1.08}P^{1.55}E^{0.101}, \qquad (8)$$

$$D_{50} = 0.00152A^{1.04}P^{1.58}E^{0.0603}, \text{ and} \qquad (9)$$

$$D_{25} = 0.00431A^{1.01}P^{1.55}E^{0.0438}. \qquad (10)$$

where

Q_n is the calculated peak flow for recurrence interval *n*, in cubic feet per second;

D_n is the estimated daily discharge exceeded *n* percent of the time, in cubic feet per second;

A is the drainage area of the watershed delineated using a GIS (boundaries for the drainage areas were from a 12-digit hydrologic unit coverage (Natural Resources Conservation Service, 1996) or digitized when a watershed's boundary was not defined by this GIS coverage), in square miles;

L is the area of lakes and ponds in a watershed as a percent of drainage area, plus 1 percent, determined using a GIS from the 1:24,000 scale National Hydrography Dataset (U.S. Geological Survey and U.S. Environmental Protection Agency, 2002) using all features in the dataset that were lake or pond boundaries including ponds as small as several hundreths of an acre and pooled areas of streams;

E is the percent of the watershed that is at or greater than 1,200 ft altitude, plus 1 percent, computed using a GIS from the National Elevation Dataset (U.S. Geological Survey, 2001b);

Y is the northing of the centroid of the drainage basin determined with a GIS, in the Vermont State Plane Coordinate System, divided by 100,000, then increased by 1.0; and

P is the basinwide mean of the mean annual precipitation, in inches, determined using a GIS and the Parameter-elevation Regressions on Independent Slopes Model (PRISM) dataset (Daly, 2000) resampled with bilinear interpolation to a 30-m cell resolution using the GIS, ARC/INFO RESAMPLE command (Environmental Systems Research Institute, Inc., 1994). Canadian precipitation data (Ghislain Jacques, Environment Quebec, written commun., January 17, 2002) were merged with the PRISM precipitation data where needed along the northern Vermont border.

The equations with metric units are described in Appendix 1.

Allowable Range of Basin Characteristics for Use in the Regression Equations

Regression equations 1–10, developed to estimate flow-frequency characteristics, are limited to sites on ungaged, unregulated streams in rural Vermont drainage basins and to sites with basin characteristics that are within the range of basin characteristics used in the development of the equations. The ranges of basin characteristics used in the regression analysis are shown in table 1. If basin characteristics from an ungaged site are outside of the ranges, the accuracies of the predictions of flow frequency are unknown. The Vermont Flow-Frequency Tool indicates if basin characteristics are outside the allowable ranges.

Basin characteristics used for developing equations 1–10 were determined with a GIS using datasets and methods previously explained in the description of each of the equation input variables. Determining the basin characteristics for use in the regression equations without a GIS, or with data sources or methods that deviate from those described in this report, may introduce bias and produce streamflow-frequency estimates with unknown error.

Standard Error of Residuals and Prediction of the Regression Equations

Standard error of residuals and prediction are measures of the accuracy of a regression equation. Residuals are the difference between observations and predictions; therefore, the standard error of residuals is a measure of how well the regression estimates agree with the observed data. The standard error of residuals for each equation is shown in table 2.

The standard error of residuals is an indication of how well the observed data are being modeled, thus, it is only an approximation of model error or how well the equations will predict streamflow at ungaged locations. The average standard error of prediction has a model error component and a sampling error component. The sampling error is a measure of the error in the regression prediction due to the regression equation being developed with observed (sampled) streamflow characteristics instead of true streamflow characteristics. Thus, the average standard error of prediction is a measure of the expected accuracy to which a regression equation can estimate a

Table 1. Range of basin characteristics used in development of the regression equations for estimating flow-frequency characteristics of ungaged, unregulated streams in Vermont

Regression equation	Basin characteristic	Minimum	Maximum	Mean
Peak flow	Drainage area (square miles)	0.211	850	96.8
Peak flow	Percent lake and pond area	0	6.86	0.641
Peak flow	Percent of basin greater than 1,200 feet altitude	0	100	58.9
Peak flow	Centroid northing (Vermont State Plane)	-87	296,194	151,420
Flow duration	Drainage area (square miles)	2.09	850	148
Flow duration	Mean annual precipitation (inches)	32.8	63.8	44.5
Flow duration	Percent of basin above 1,200 feet altitude	0	100	63.8

Table 2. Standard error of residuals of the regression equations for estimating flow frequencies on ungaged, unregulated streams in rural drainage basins in Vermont

Flow-frequency characteristic	Standard error of residuals	
	(in $\log_{10}$ units)	(in percent)
Peak flow with 2-year recurrence interval	0.171	48.4 to –32.6
Peak flow with 5-year recurrence interval	.170	48.0 to –32.4
Peak flow with 10-year recurrence interval	.175	49.7 to –33.2
Peak flow with 25-year recurrence interval	.182	52.0 to –34.2
Peak flow with 50-year recurrence interval	.190	54.9 to –35.4
Peak flow with 100-year recurrence interval	.200	58.4 to –36.9
Peak flow with 500-year recurrence interval	.227	69.0 to –40.8
Daily discharge exceeded 75 percent of the time	.0540	13.2 to –11.7
Daily discharge exceeded 50 percent of the time	.0832	21.1 to –17.4
Daily discharge exceeded 25 percent of the time	.124	33.2 to –24.9

flow-frequency characteristic at an ungaged location with basin characteristics similar to the ones used to develop the regression equation.

Hodge and Tasker (1995) describe the mathematical equation for computing the standard error of prediction, SE_{pred}, of a flow-frequency estimate as

$$SE_{pred} = [\gamma^2 + x_i(X^T\Lambda^{-1}X)^{-1}x_i^T]^{1/2}, \qquad (11)$$

where

- γ^2 is the model-error variance (table 3);
- x_i is a row vector of the logarithms of the basin characteristics for the study site i, augmented by a 1 as the first element;
- T is the matrix algebra symbol for transposing (Grossman, 1986) a matrix; and
- $(X^T\Lambda^{-1}X)^{-1}$ is the (p x p) matrix for the regression equations (table 3) with X being a (n x p) matrix that has a row of logarithmically transformed basin characteristics augmented by a leading 1 and Λ being the (n x n) covariance matrix used for weighting sample data in the GLS regression; n is the number of stream-gaging stations used in the regression analysis and p is the number of basin characteristics, plus 1.

The standard error of prediction of a streamflow estimate at an ungaged location can be computed, because x_i is not required to be a row of X, The Vermont Flow-Frequency Tool accompanying this report computes and displays the standard error of prediction for each flow-frequency estimate. To find the average standard error of prediction of the equations that are shown in table 4, the standard error of prediction was computed for each of the stations used in the development of the flood-flow-frequency regression equation then averaged.

The standard error of prediction of estimates of flow duration at an ungaged site also is computed with equation 11, but with 2 substitutions for variables. This substitution is necessary because OLS regression was used instead of GLS regression to develop the equations for estimating flow duration. The first substitution is for model error, γ, which is approximated with the standard error of residuals (table 2). The second substitution is for the covariance matrix, Λ, used for weighting the station data. This matrix is simplified to a matrix with diagonal elements equal to γ^2 and off-diagonal elements equal to zero.

The standard error of prediction is converted to positive and negative percent errors with the following formulas:

$$S_{pos} = 100(10^{SE_{pred}} - 1)\text{, and} \qquad (12)$$

$$S_{neg} = 100(10^{-SE_{pred}} - 1) \quad . \qquad (13)$$

where

- S_{pos} is the positive percent error of prediction,
- S_{neg} is the negative percent error of prediction, and
- SE_{pred} is the standard error of prediction in logarithmic units.

Equations 12 and 13 apply not only to the standard error of prediction, but also to the average standard error of prediction and standard error of residuals by substituting the appropriate error term in logarithmic units. The probability that the true value of streamflow at a given frequency is between the positive-percent and negative-percent standard error of prediction is approximately 68 percent. The standard error of prediction is computed and included in the results each time a flow-frequency estimate is made with the Vermont Flow-Frequency Tool.

Equivalent Years of Record of Results of the Regression Equations

The equivalent years of record (Hardison, 1971) shown in table 4 and computed by the provided Vermont Flow-Frequency Tool is another measure of accuracy of the regression equations. The equivalent years of record are the number of years of data collection at an ungaged site that would be required to achieve frequency results with accuracy equal to the regression equations. It is computed by use of the equation

$$E = s^2[1 + k_n g + 0.5k_n^2(1 + 0.75g^2)]/SE_{pred}^2, \qquad (14)$$

Table 3. Model error variance and the $(X^{T} \Lambda^{-1} X)^{-1}$ matrices for the regression equations

[A $(X^{T}\Lambda^{-1}X)^{-1}$ matrix is the (p x p) matrix for the regression equations with X being a (n x p) matrix that has a row of logarithmically transformed basin characteristics augmented by a leading 1 and Λ being the (n x n) covariance matrix used for weighting sample data in the GLS regression; n is the number of stream-gaging stations used in the regression analysis and p is the number of basin characteristics, plus 1; Numbers in the $(X^{T}\Lambda^{-1}X)^{-1}$ matrices are in scientific notation; --, no data]

Flow-frequency characteristic	Model-error variance, γ^2	$(X^{T} \Lambda^{-1} X)^{-1}$ matrix				
Peak flow with a 2-year recurrence interval	0.0284	0.39106E-02	-0.22912E-03	-0.97874E-03	-0.10037E-02	-0.36445E-02
		-0.22912E-03	0.38268E-03	-0.63663E-03	-0.11393E-03	-0.63470E-04
		-0.97874E-03	-0.63663E-03	0.83023E-02	-0.33102E-04	0.10605E-02
		-0.10037E-02	-0.11393E-03	-0.33102E-04	0.74670E-03	-0.15507E-03
		-0.36445E-02	-0.63470E-04	0.10605E-02	-0.15507E-03	0.10302E-01
Peak flow with a 5-year recurrence interval	.0271	0.41068E-02	-0.26008E-03	-0.10128E-02	-0.97846E-03	-0.39752E-02
		-0.26008E-03	0.40128E-03	-0.61274E-03	-0.12005E-03	-0.86724E-04
		-0.10128E-02	-0.61274E-03	0.82284E-02	-0.45938E-04	0.10812E-02
		-0.97846E-03	-0.12005E-03	-0.45938E-04	0.74762E-03	-0.15726E-03
		-0.39752E-02	-0.86724E-04	0.10812E-02	-0.15726E-03	0.11254E-01
Peak flow with a 10-year recurrence interval	.0277	0.45209E-02	-0.29160E-03	-0.10952E-02	-0.10313E-02	-0.45038E-02
		-0.29160E-03	0.43815E-03	-0.63798E-03	-0.13321E-03	-0.10916E-03
		-0.10952E-02	-0.63798E-03	0.87145E-02	-0.60592E-04	0.11726E-02
		-0.10313E-02	-0.13321E-03	-0.60592E-04	0.80298E-03	-0.17208E-03
		-0.45038E-02	-0.10916E-03	0.11726E-02	-0.17208E-03	0.12762E-01
Peak flow with a 25-year recurrence interval	.0282	0.50699E-02	-0.32950E-03	-0.11943E-02	-0.10999E-02	-0.52262E-02
		-0.32950E-03	0.48450E-03	-0.67035E-03	-0.15032E-03	-0.14061E-03
		-0.11943E-02	-0.67035E-03	0.93168E-02	-0.80941E-04	0.12900E-02
		-0.10999E-02	-0.15032E-03	-0.80941E-04	0.87551E-03	-0.19373E-03
		-0.52262E-02	-0.14061E-03	0.12900E-02	-0.19373E-03	0.14825E-01
Peak flow with a 50-year recurrence interval	.0294	0.56002E-02	-0.36166E-03	-0.12972E-02	-0.11833E-02	-0.58946E-02
		-0.36166E-03	0.53107E-03	-0.71790E-03	-0.16740E-03	-0.16736E-03
		-0.12972E-02	-0.71790E-03	0.10048E-01	-0.97417E-04	0.14115E-02
		-0.11833E-02	-0.16740E-03	-0.97417E-04	0.95469E-03	-0.21490E-03
		-0.58946E-02	-0.16736E-03	0.14115E-02	-0.21490E-03	0.16725E-01
Peak flow with a 100-year recurrence interval	.0313	0.62124E-02	-0.39699E-03	-0.14203E-02	-0.12880E-02	-0.66509E-02
		-0.39699E-03	0.58615E-03	-0.78119E-03	-0.18749E-03	-0.19653E-03
		-0.14203E-02	-0.78119E-03	0.10972E-01	-0.11479E-03	0.15561E-02
		-0.12880E-02	-0.18749E-03	-0.11479E-03	0.10506E-02	-0.23911E-03
		-0.66509E-02	-0.19653E-03	0.15561E-02	-0.23911E-03	0.18869E-01
Peak flow with a 500-year recurrence interval	.0378	0.79249E-02	-0.49238E-03	-0.17810E-02	-0.16035E-02	-0.87217E-02
		-0.49238E-03	0.74532E-03	-0.98171E-03	-0.24512E-03	-0.27330E-03
		-0.17810E-02	-0.98171E-03	0.13796E-01	-0.15917E-03	0.19741E-02
		-0.16035E-02	-0.24512E-03	-0.15917E-03	0.13314E-02	-0.30477E-03
		-0.87217E-02	-0.27330E-03	0.19741E-02	-0.30477E-03	0.24721E-01

Table 3. Model error variance and the $(X^T \Lambda^{-1} X)^{-1}$ matrices for the regression equations--Continued

Flow-frequency characteristic	Model-error variance, γ^2	$(X^T \Lambda^{-1} X)^{-1}$ matrix				
Daily flow exceeded 25 percent of the time	.00292	0.54316E-01	-0.24661E-03	-0.34488E-01	0.17354E-02	--
		-0.24661E-03	0.83665E-04	0.76129E-04	-0.15047E-04	--
		-0.34488E-01	0.76129E-04	0.22208E-01	-0.13052E-02	--
		0.17354E-02	-0.15047E-04	-0.13052E-02	0.25861E-03	--
Daily flow exceeded 50 percent of the time	.00692	0.12877E-00	-0.58470E-03	-0.81768E-01	0.41146E-02	--
		-0.58470E-03	0.19836E-03	0.18049E-03	-0.35675E-04	--
		-0.81768E-01	0.18049E-03	0.52653E-01	-0.30946E-02	--
		0.41146E-02	-0.35675E-04	-0.30946E-02	0.61314E-03	--
Daily flow exceeded 75 percent of the time	.0153	0.28853E-00	-0.13100E-02	-0.18320E-00	0.92191E-02	--
		-0.13100E-02	0.44444E-03	0.40441E-03	-0.79931E-04	--
		-0.18325E-00	0.40441E-03	0.11797E-00	-0.69336E-02	--
		0.92191E-02	-0.79931E-04	-0.69336E-02	0.13737E-02	--

Table 4. Average standard error of prediction and average equivalent years of record of the regression equations for estimating peak flow-frequency characteristics on ungaged, unregulated streams in rural drainage basins in Vermont

Flow-frequency characteristic	Average standard error of prediction		Average equivalent years of record
	(in $\log_{10}$ units)	(in percent)	
Peak flow with 2-year recurrence interval	0.173	48.8 to –32.8	1.4
Peak flow with 5-year recurrence interval	.169	47.6 to –32.2	2.3
Peak flow with 10-year recurrence interval	.172	48.4 to –32.6	3.2
Peak flow with 25-year recurrence interval	.174	49.1 to –32.9	4.6
Peak flow with 50-year recurrence interval	.178	50.5 to –33.5	5.5
Peak flow with 100-year recurrence interval	.184	52.6 to –34.5	6.3
Peak flow with 500-year recurrence interval	.202	59.2 to –37.2	7.6

where

E is the equivalent years of record;

s is the standard deviation of annual events estimated from a regression between the standard deviation and drainage area, in square miles, A, at stations used in the regression analysis ($s = 0.0339A^{-0.102}$);

k_n is the log-Pearson type III frequency factor for the n-year event;

g is the skew coefficient used in the computation of the frequency curve (assumed to be zero when computing equivalent years of record for an estimate at an ungaged site); and

SE_{pred} is the standard error of prediction.

Because equation 14 requires variables that do not apply to flow duration estimates such as the frequency factor, k_n, equivalent years of record cannot be computed for the flow-duration estimates. Equivalent years of record are computed and output each time a flood-frequency estimate is made with the Vermont Flow-Frequency Tool.

Prediction Intervals of Results of the Regression Equations

Prediction intervals indicate the uncertainty in the use of the equations. For example, one can be 90-percent confident that the true value of a flow-frequency characteristic lies within the 90-percent prediction interval. The prediction interval can be computed as follows:

Let

$$V = 10^{(t_{\alpha/2,n-p} SE_{pred})}, \qquad (15)$$

then

$$(1/V)Q_{pred} < Q_{true} < (V)Q_{pred}, \qquad (16)$$

where

$t_{\alpha/2,n-p}$ is the critical value from a Students-t distribution (found in many introductory statistics textbooks) at alpha level α ($\alpha = 0.10$ for a 90-percent prediction interval) with $n - p$ degrees of freedom, n is the number of stations used in the regression analysis, and p is the number of basin characteristics in the regression equation, plus 1;

SE_{pred} is the standard error of prediction of a flow-frequency estimate;

Q_{pred} is the computed flow at a selected frequency from the regression equation; and

Q_{true} is the true value of flow at a selected frequency.

Prediction intervals at the 50-, 67-, 90-, and 95-percent confidence levels are computed and output each time a flow-frequency estimate is made with the Vermont Flow-Frequency Tool.

VERMONT FLOW-FREQUENCY TOOL

To determine the basin characteristics required by the flow-frequency estimating equations for streams in Vermont, a GIS application was developed. The application, called the Vermont Flow-Frequency Tool, allows the user to interactively select a point on a stream of interest. The tool automatically delineates the required basin characteristics and solves the regression equations. In addition, a link between the Vermont Flow-Frequency Tool and the National Flood Frequency (NFF) program (Ries and Crouse, 2002) provides the ability to plot frequency curves and synthetic hydrographs, and adjust flow-frequency curves for urbanization. The application customized for Vermont also calculates standard error of prediction, prediction intervals, and expected probability adjustments.

The Vermont Flow-Frequency Tool was developed in the Environmental Systems Research Institute, Inc. (ESRI) ArcView version 3.2 software (1996a) using ESRI Avenue programming language (Environmental Systems Research Institute, Inc., 1996b) and converted to an ESRI ArcView extension. Extensions are plug-in programs that can be easily loaded or unloaded into ArcView as needed. The drainage-basin delineation done by the Vermont Flow-Frequency Tool is based on the NHD and the NHD Toolkit (U.S. Geological Survey and U.S. Environmental Protection Agency, 2002) and can be downloaded from the USGS New Hampshire/Vermont District website at http://vt.water.usgs.gov/CurrentProjects/vtfloodfreq/index.htm.

System Requirements

The Vermont Flow-Frequency Tool requires the following software and toolkits:

1. ESRI ArcView version 3.1 or higher except versions 8.x, which do not support the Avenue programming language;
2. The ESRI ArcView Spatial Analyst extension; and
3. The NHD Toolkit (U.S. Geological Survey and U.S. Environmental Protection Agency, 2002).

The NHD Toolkit is available for download from the NHD website, http://nhd.usgs.gov/. The NHD Toolkit contains a suite of extensions some of which are used to process NHD data and some of which the Vermont Flow-Frequency Tool is dependent upon for execution.

Although the computer and operating system requirements are defined by the ESRI ArcView software, the Vermont Flow-Frequency Tool does limit the operating system to the Microsoft Windows operating system. The Vermont Flow-Frequency Tool successfully runs in Microsoft Windows98, NT, and 2000. Other operating systems have not been tested. Notepad, a program for viewing text files commonly packaged with the Microsoft Windows operating system is required. Required disk space for the entire suite of topographic map images, NHD data, and other supporting datasets for computing basin characteristics throughout Vermont is estimated to be 5.3 gigabytes. Data such as the NHD and the topographic map images, however, can be installed in different locations on a local network. One gigabyte of virtual memory is required for delineating the full range of drainage basin sizes encountered in Vermont.

The Vermont Flow-Frequency Tool can be linked to the NFF Program (Ries and Crouse, 2002), which is discussed in greater detail in the section "Utilities of the National Flood Frequency Program." Although the Vermont Flow-Frequency Tool does not require NFF, users may find the utilities provided by NFF to be valuable. NFF is available on the Internet from the USGS (http://water.usgs.gov/software/nff.html). After NFF is installed, a copy of the NFF executable file needs to be placed in the Vermont Flow-Frequency Tool utilities folder.

Data Requirements

The Vermont Flow-Frequency Tool requires several gigabytes of data to work statewide. For the Tool to function properly, the data directory must be structured and use the naming conventions as shown in figure 1. For each hydrologic unit, there are three directories whose name begins with the hydrologic unit code and ends in either "_24shp," " _ff," or "_ws." The folder with the "_24shp" suffix contains the 1:24,000 scale NHD for Vermont projected into Vermont State Plane coordinates and converted to the ESRI ArcView shapefile format using the NHD Arc2shape extension. The naming convention of the subdirectories in the "_24shp" directory are generic to the NHD and are explained in the helpfiles of NHD extensions. The instructions for the NHD Watershed extension explain the processing steps in detail (Steeves, 2002). The NHD for Vermont and the processing instructions can be downloaded from the NHD Web site, http://nhd.usgs.gov/.

The folders with the "_ff" suffix (fig. 1) contain the data for computing basin characteristics in a hydrologic unit. The hydrologic unit is identified by the prefix of the folder name. Each of these "_ff" folders contain three grids projected into Vermont State Plane coordinates. The first grid dataset is the elevation grid clipped to the respective hydrologic unit and named "dem_grid." The source of the grid is the National Elevation Dataset (NED) (U.S. Geological Survey, 2001b). The second grid, "grid_1200ft," is for computing drainage basin areas greater than 1,200 ft altitude. This grid was computed from the NED with cells greater than 1,200 ft having a value of 2, and cells at or less than 1,200 ft having a value of 1. The final grid in the "_ff" folder, "precip_grd," contains the mean-annual-precipitation data. The Vermont Flow-Frequency Tool only requires the precipitation grid for computing daily discharges exceeded 25, 50, and 75 percent of the time. The precipitation data are not needed for estimating flood-flow frequency. The source of the precipitation grid is the PRISM dataset (Daly, 2000), which is available through Climate

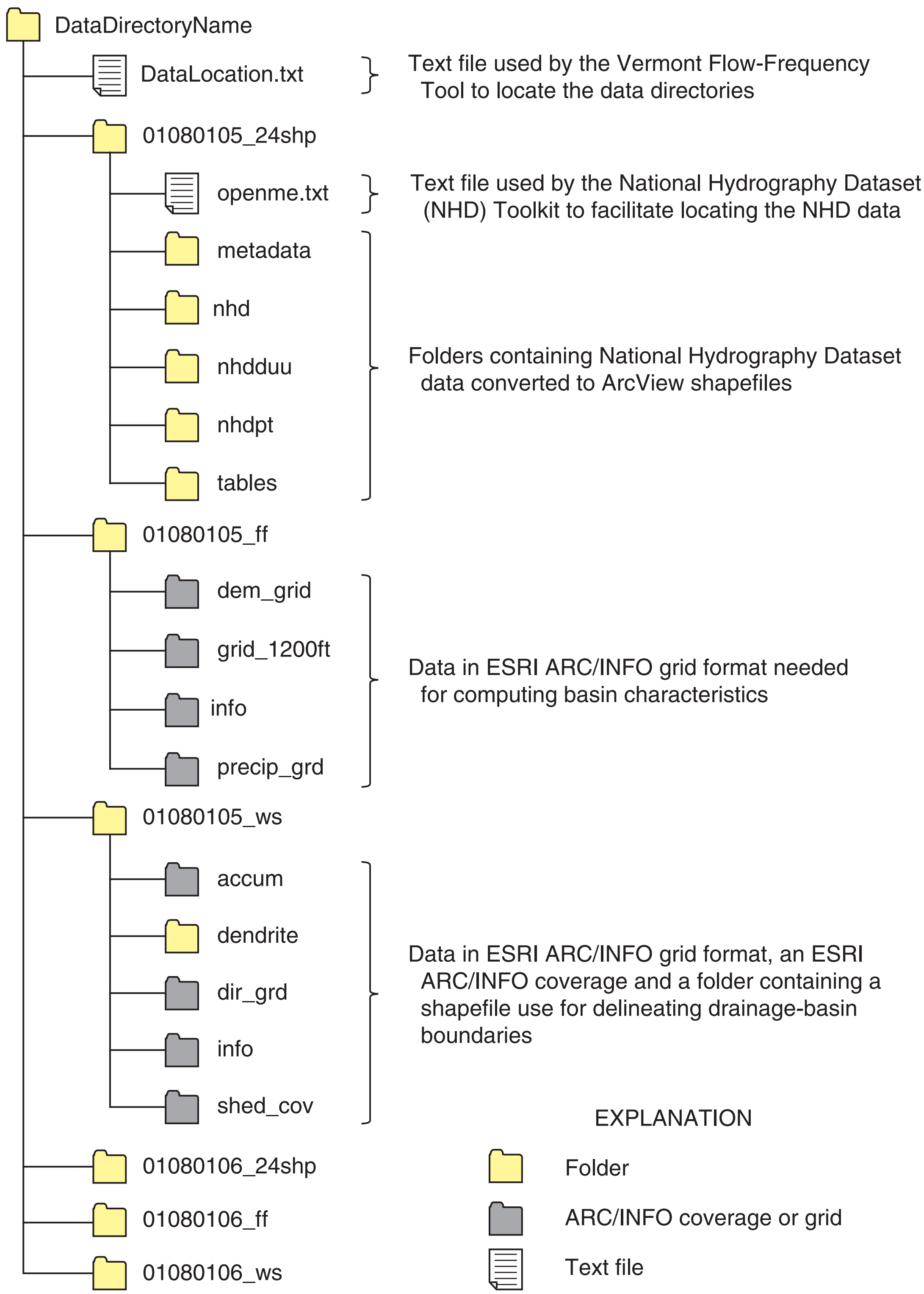

Figure 1. Directory structure of the data files required by the Vermont Flow-Frequency Tool. Hydrologic unit codes 01080105 and 01080106 are shown as examples.

Source at http://www.climatesource.com/. The PRISM data are projected into Vermont State Plane coordinates and resampled with bilinear interpolation to a 30-m cell resolution using the GIS ARC/INFO command RESAMPLE (Environmental Systems Research Institute, Inc., 1994). Obtaining, projecting, and processing the PRISM precipitation data covering Vermont is the responsibility of the user.

The PRISM precipitation data do not cover Canadian parts of basins shared between Vermont and Canada. Canadian mean-annual-precipitation data (Ghislain Jacques, Environment Quebec, written commun., January 17, 2002), therefore, were obtained for precipitation stations in areas adjacent to the northern Vermont border. A mean-annual-precipitation grid with cells at 30-m spacing covering the areas north of the Vermont border was interpolated from this point-precipitation data and merged with the PRISM precipitation data along the northern Vermont border.

The folders with the "_ws" suffix (fig. 1) contain data required for delineating a drainage basin. Detailed instruction for developing the data in the "_ws" folders can be downloaded from the Applications section of the NHD Web site, http://nhd.usgs.gov/. Within the "_ws" folder is a folder named "dendrite," which contains the dendritic stream network used in the selection of a point on a stream. This dendritic dataset is a copy of the NHD-reach route system downloaded from the NHD Web site, projected to Vermont State Plane coordinates and converted to ESRI ArcView shapefiles; however, this dendritic dataset has some stream reaches removed. The purpose of removing some of the reaches is to prevent users from selecting a stream reach in which the flow-frequency estimating equations do not apply. For example, artificial channels, stream braids, and the Connecticut River are all reaches where the equations do not apply and are removed from the dendrite dataset. The creation of the dendrite dataset is further explained in instructions for the NHD Watershed extension downloadable from the NHD Web site.

The other datasets in the "_ws" folder include the flow accumulation grid, "accum;" the flow direction grid, "dir_grd;" and a coverage of drainage basin catchments for each stream reach in the NHD, "shed_cov." These datasets are processed from the NED (U.S. Geological Survey, 2001b). First, the ARC/INFO command TOPOGRID (Environmental Systems Research Institute, Inc., 1994) and the ARC/INFO script, "agree.aml," which is available at the Web site http://www.ce.utexas.edu/prof/maidment/gishydro/ferdi/, are applied to the NED to produce an elevation grid with a 10-m cell size that is hydrologically correct and horizontally matches the NHD network. Additional raster processing of the elevation grid is done by use of ARC/INFO to exaggerate basin boundaries taken from the Natural Resources Conservation Service (NRCS) 12-digit hydrologic unit coverage (Natural Resources Conservation Service, 1996) to enforce agreement of the elevation grid to these boundaries. Next, the ARC/INFO commands FILL, FLOWDIRECTION, and FLOWACCUMULATION are used to create the flow direction and accumulation grids. Last, the watershed catchment coverage, "shed_cov," is created using the ARC/INFO command WATERSHED and each catchment is attributed with the corresponding NHD reach code, matching each catchment to a stream segment in the NHD. The creation of these three datasets is explained further in the instructions for the NHD Watershed extension downloadable from the NHD Web site.

The utilities folder, named Utilities, also is important to the functionality of the Vermont Flow-Frequency Tool. The utilities folder contains some of the data that is displayed by the tool to make locating streams easier, such as roads and basin, town, and county boundaries. The utilities folder contains two shapefiles, "dams.shp" and "huc8.shp," which are used by the tool to check for upstream flood control structures and to load hydrologic unit data, respectively. The program (vtreg.exe) and datasets for solving the regression equations and analyzing the prediction intervals also are located in this directory. Each user of the Vermont Flow-Frequency Tool must have a copy of the utilities directory because this directory is used as a temporary workspace for saving files created during intermediate steps in computing flow frequency. A copy of the required utilities folder can be downloaded from the Web site http://vt.water.usgs.gov/CurrentProjects/vtfloodfreq/index.htm.

The last dataset used by the Vermont Flow-Frequency Tool is the topographic map images in electronic format at 1:24,000 scale. These maps are known also as Digital Raster Graphs. The images will make the process of accurately locating streams simple. The topographical images can be obtained from the Vermont Center for Geographic Information (http://www.vcgi.org/).

Installation of the Vermont Flow-Frequency Tool

Once the required ESRI software is installed and the datasets and utilities are loaded in the appropriate directory structure, and the Vermont Flow-Frequency Tool extension has been copied to the ESRI ArcView extensions directory, the Vermont Flow-Frequency Tool is ready for use. First, the ESRI ArcView software needs to be executed and the ESRI ArcView Spatial Analyst extension loaded and set as default. To load an extension, select the File menu in the ArcView window and select Extensions. A window of available extensions will open for the user to choose from. With the ESRI ArcView Spatial Analyst extension loaded, open the extension window again and load the extensions titled “NHD Load/Unload Workspace,” “NHD Navigate,” and “Vermont FF Tool.”

After the Vermont Flow-Frequency Tool extension is loaded, a menu named “VTSETUP” will appear in the project window. This set-up routine should be executed when the extension is first loaded into each new ESRI ArcView project file to identify the locations of required data and utilities folder and create the map of Vermont in the display window. Once the set-up routine is run, it will not be needed again in the ESRI ArcView project file and the “VTSETUP” menu will disappear. Saving and using only this project file will eliminate the need to run the set-up routine again.

The set-up routine first prompts the user for the location of the utilities folder. The user must browse for the “utilitiesLocation.txt” file located in the user’s utilities directory. The location of the data folders is requested in a similar manner, but the user must browse for the “dataLocation.txt” file (fig. 1) located in the directory that contains the data folders. The location of the topographic map database file (topo.dbf) also is requested; however, including the topographical maps is not required and search for this file can be canceled. The set up for the Vermont Flow-Frequency Tool is complete and the routine then opens a window displaying a map of Vermont. The Vermont Flow-Frequency Tool menu and buttons are now displayed in the ESRI ArcView menubar and toolbar, respectively. The ESRI ArcView window may need to be enlarged to display the entire menubar and toolbar. An example of an ESRI ArcView project window with the Vermont Flow-Frequency Tool loaded and set up is shown in figure 2.

Using the Vermont Flow-Frequency Tool

This section of the report includes instructions for using the Vermont Flow-Frequency Tool. To be successful, the user must be familiar with the basic functions of ESRI ArcView such as pan and zoom, making themes active, interactively identifying features, and turning on and off the display themes, such as the topographical maps.

With the set up complete, the Vermont Flow-Frequency Tool is ready for use. There are five buttons on the toolbar that run the major functions of the tool. These buttons are displayed in table 5. The first button, the load hydrologic unit button HUC, must be used to load the appropriate hydrography data and locate the appropriate support files for computing basin characteristics before a basin can be delineated. With the load hydrologic unit button depressed, the cursor is used to select the hydrologic unit where the stream of interest is located. The hydrologic unit boundaries are outlined in green by default (fig. 2).

With a hydrologic unit selected, the Vermont Flow-Frequency Tool button can be depressed and a point on a stream selected. For a point to be selected on a stream, the “HUC_24shp – Reach” theme, where HUC is the hydrologic unit code, must be active. The basin above this point is delineated by implementing the functionality of the NHD. The basin characteristics upstream of this point and the flow-frequency estimates for the selected point are automatically computed and displayed as shown in the example in figure 3. The tabular results are displayed and can be saved to disk using Notepad.

The Vermont Flow-Frequency Tool offers an option for computing channel characteristics, such as slope, for a drainage basin already delineated with the Flow-Frequency Tool. With the Channel Characteristics button depressed, the user must reselect the point from which the currently displayed drainage basin was delineated. The tool uses the NHD to determine the main channel. It then locates the upstream end of the main channel, delineates the catchment above this point, and extends the channel to the point that provides the longest flow length between the draingage divide and the upstream end of the channel identified on a USGS topographical map. Results are output in tabular form to a Notepad window. The tables include the elevation at the downstream end of the main channel, the elevation at the upstream end of the main channel, the elevation at

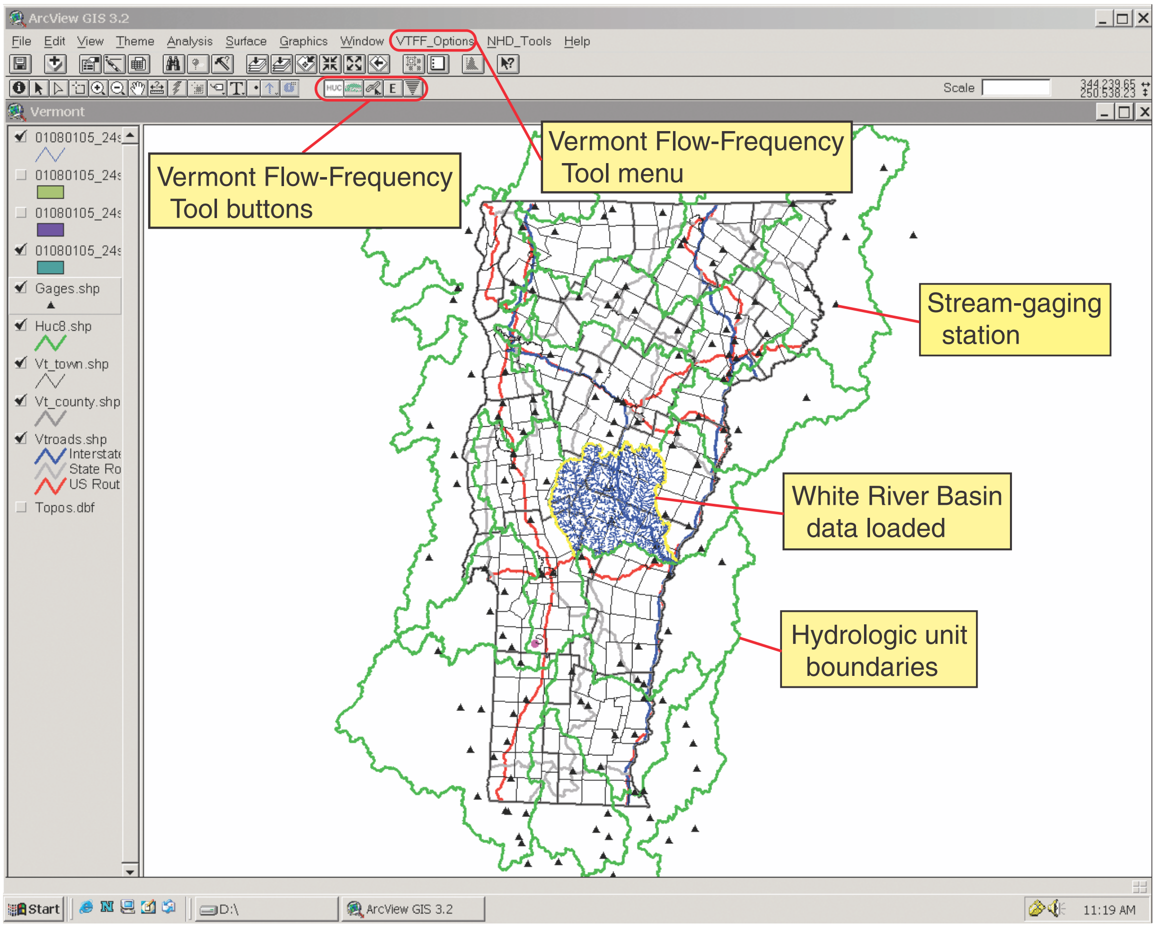

Figure 2. ArcView window with the Vermont Flow-Frequency Tool installed.

Table 5. Buttons that run the functions of the Vermont Flow-Frequency Tool

[Buttons are shown in figures 2 and 3; NHD, National Hydrography Data Set]

Button	Name	Function
	Load hydrologic unit	Loads NHD data and locates support data sets for the hydrologic unit interactively selected.
	Flow-frequency tool	Delineates a drainage basin by use of NHD for a point on a stream interactively selected, computes the basin characteristics, and solves the regression equations.
	Channel characteristics	Estimates selected channel characteristics.
	Elevation query	Displays the elevation for a point interactively selected.
	Optional flow-frequency tool	Delineates a drainage basin by use of the flow direction and accumulation grids for a point interactively selected, computes the basin characteristics, and solves the regression equations.

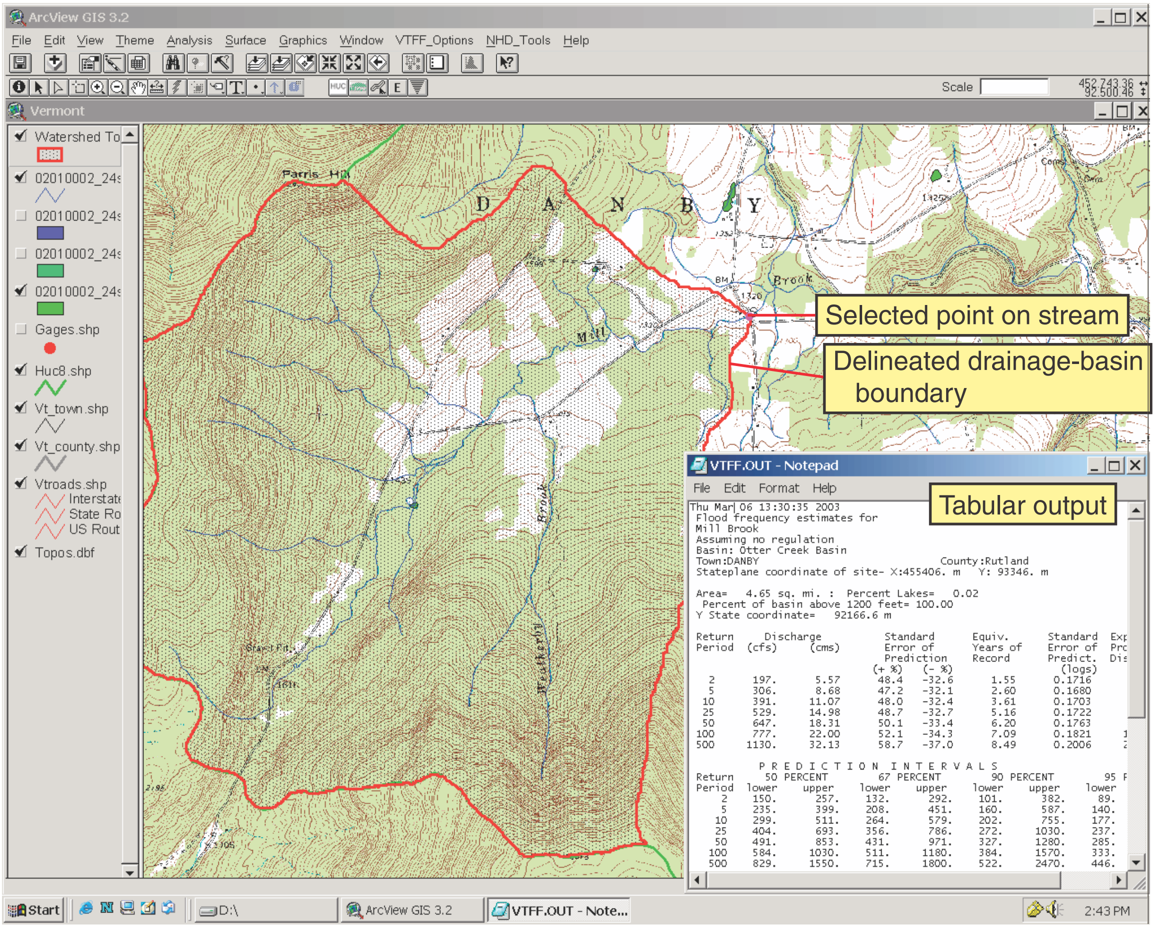

Figure 3. ArcView window displayed when a point on a stream is selected.

the upstream end of the main channel extended to the drainage divide, the elevation of points 10 and 85 percent up the main channel from the selected location to the drainage divide, the length of the main channel, the length of the main channel extended to the drainage divide, and the slope of the main channel between the 10 and 85 percent points described above.

If the user is not satisfied with the Vermont Flow-Frequency Tool attempt to extended the channel to the drainage divide, a manual option is available. By depressing and holding down the Channel Characteristics button, an optional button appears for selection. With the optional Channel Characteristics button selected, a point along the inside of the drainage divide where the main channel should be extended to can be selected. The Vermont Flow-Frequency Tool determines the new extended channel using the flow direction grid, computes the channel characteristics, and outputs the results to a new Notepad window.

The Elevation Query button is for identifying elevations on a map. With the Elevation Query button depressed, the cursor is used to select a point on a map for which the elevation will be displayed in feet.

The Optional Flow-Frequency Tool functions in the same manner as the Flow-Frequency Tool, except the basin delineation is done with a flow-direction and flow-accumulation grid. A flow-direction grid is a grid in which each cell contains information on the direction of flow at that cell. A flow-accumulation grid is a grid in which each cell contains the number of cells that flow into it. Basin delineation with these grids takes more time and computer resources than basin delineation using the NHD approach. Attempting to delineate large basins, such as those greater than a few hundred square miles with the optional Flow-Frequency Tool may exceed the capabilities of a personal computer.

Two technical difficulties are known to exist in the tool. The first is related to braided streams. User's should review the boundary across the braided reach and confirm that the Vermont Flow-Frequency Tool delineated this part of the boundary correctly. It is best to avoid a braided reach by selecting a point of interest either upstream or downstream of the braid. The second problem can occur if the point chosen for delineation is at a confluence. For example, if a drainage basin above a confluence is desired and the point selected is too close to the confluence, both catchments related to the reaches entering the confluence may be delineated; however, only the headwater catchments above the reach that the selected point is on will be included in the total drainage area. To avoid the problem, the Vermont Flow-Frequency Tool requires selected points to be farther than 100 ft or about 30 m in the display's scale upstream or downstream from the confluence.

ADDITIONAL UTILITIES OF VERMONT FLOW-FREQUENCY TOOL

Additional utilities are provided in the Vermont Flow-Frequency Tool. These include items such as adjusting the results with data from a nearby stream-gaging station, adjusting the results for expected probability, and executing the National Flood Frequency program.

Vermont Flow-Frequency Tool Options Menu

A menu on the ESRI ArcView menubar titled "VTFF_Options" is visible when the Vermont Flow-Frequency Tool extension is installed. This menu has 13 selections. Of the 13 selections, three—"Adjust Results for Nearby Gage," "Run NFF Program using English Units," and "Run NFF Program using Metric Units"—are discussed in more detail in the following sections. Ten of the selections need only the following brief descriptions. The first two selections in this menu are the "Select Utilities Location" and "Select Data Location" menu options. These can be used if datasets or the utilities folder are moved and the user does not want to run the set-up routines again. When one of these options is selected, the user must browse for the respective "utilitiesLocation.txt" or "dataLocation.txt file," as described in the previous section "Installation of the Vermont Flow-Frequency Tool."

Two options under the "VTFF_Options" menu are used to locate features and move around on the displayed map of Vermont. The "Zoom to Vermont Map" option zooms in or out so that the window displays the entire State of Vermont. The "Locate Stream by Name" option opens a window with all the names of the streams in the currently loaded hydrologic unit. Once the user selects a stream, the map display zooms to that stream. Another option, "Watershed display properties," allows the user to toggle the display of the watershed polygon from shaded to unshaded.

Options "Recompute Watershed Tool Results" and "Recompute Flow Accumulation Tool Results," compute the basin characteristics and estimated flow-frequency characteristics of a basin polygon that is already displayed. These options are helpful if the delineated basin boundary identified by the Watershed Tool or the Flow Accumulation Tool must be manually edited. The user must be proficient in hydrology and with the ESRI ArcView editing tools before using these options. The option "Display Gage Information" provides streamflow statistics for selected stream-gaging stations used for developing the flow-estimating equations. The station must first be selected with the ESRI ArcView Selection tool.

Finally, a menu item called "Clean Workspace" deletes many of the temporary work files that are left in the utilities directory. This option, however, does not delete files from the default ArcView Temp directory. A menu item called "About" provides a short explanation of the Vermont Flow-Frequency Tool.

Use of Regression Equations at or near Stream-Gaging Stations

A flow-frequency estimate from the regression equations can be improved with actual streamflow data if the estimate is made at or near a stream-gaging station. This is done by adjusting the regression equation results with the frequency curve computed from the streamflow data. A database of the streamflow data needed for this adjustment (through the 2000 water year, which ends on September 30) is included in the Vermont Flow-Frequency Tool. This method may be particularly useful when a limited number of years of record at a gaged site are available for frequency analysis. The procedure recommended in Bulletin 17B of the U.S. Interagency Advisory Committee on Water Data (1982) is to compute a weighted average of the regression estimate and station estimate using the following equation:

$$LogQ_{T,w} = \frac{\log Q_{T,s}(N) + \log Q_{T,r}(E)}{N+E}, \qquad (17)$$

where

- $Q_{T,w}$ is the weighted discharge for flow-frequency characteristic *T*,
- $Q_{T,s}$ is flow-frequency characteristic *T* computed from the streamflow data,
- $Q_{T,r}$ is the regression estimate for flow-frequency characteristic *T*,
- N is the number of years of streamflow data used to compute $Q_{T,s}$, and
- E is the equivalent years of record for $Q_{T,r}$.

Flow frequency at sites that are not at but relatively near a stream-gaging station and on the same stream, can be calculated by combining the results of the regression equations and the nearby station data. This method is adapted from Sauer (1974) and is applicable to statistics of peak flow and flow duration. For sites with drainage area within 50 percent of the drainage area of the gaged site on the same stream, the following equations are used. For the gaged site, the ratio, *R*, is computed as

$$R = \log Q_{T,w} / \log Q_{T,r}. \qquad (18)$$

This ratio is linearly decayed to 1.0 at plus or minus 50 percent of the gaged drainage area. Thus, for the ungaged site a modified ratio, *R*', is computed as

$$R' = R - \frac{\Delta A}{0.5A_g}(R-1), \qquad (19)$$

where

- ΔA is the absolute value of the difference between the drainage area at the ungaged site and drainage area at the gaged site, and
- A_g is the drainage area of the gaged site.

To calculate the adjusted discharge estimate at the ungaged site, *R*' is applied to the logarithmically transformed regression estimate using the following equation:

$$Q_{T,a} = 10^{R'(\log Q_{T,u})}, \qquad (20)$$

where

- $Q_{T,a}$ is the adjusted discharge estimate of flow-frequency characteristic *T* at the ungaged site, and
- $Q_{T,u}$ is the regression estimate of flow-frequency characteristic *T* at the ungaged site.

After a flow-frequency estimate had been made for a point on a stream with the Vermont Flow-Frequency Tool, the above computations (equations 17–20) can be made to adjust for a nearby station by selecting “Adjust Results for Nearby Gage” under the “VTFF_Options” menu. When this menu option is selected, a stream-gaging station number can be entered or the application can search for potential stations. The adjusted flow-frequency estimates are output in table format to a new Notepad window.

Expected Probability Adjustment to Flood Discharges

There is error in all flood-frequency curves because of the limited years of record available for analysis. These errors are magnified when estimating low-frequency floods, such as a 100-year event, from a finite series of annual peaks as few as 10 years. Furthermore, the distribution of errors around the true flood-frequency characteristics is not symmetrical. For example, a 100-year flood is a flood that will be exceeded on average one time in a hundred years. This flood estimate could actually be exceeded 2, 3, 4, or more times in a hundred years, but it can never be exceeded fewer than zero times in a hundred years. Because the distribution is not symmetrical, the median and mean of the distribution are not the same.

Exceedence probabilities computed on the basis of guidelines in Bulletin 17B of the U.S. Interagency Advisory Committee on Water Data (1982) are median-probability estimates. Expected probability (Beard, 1960) conforms to the mean of the distribution and is defined as the average of all estimated true probabilities for a specified flow frequency that might be made from successive samples of a specified size (U.S. Interagency Advisory Committee on Water Data, 1982). Thomas (1976) documented that the mean-probability estimate exceeds the median-probability estimate and the median-probability estimate approaches the true probability of a time series as the length of the streamflow record increases.

There are limitations and advantages in the use of median- and mean-probability estimates, but the use of the flood-frequency estimate will define which is appropriate. For example, when a user wants the result to have an equal likelihood of being too high or too low, the median probability estimate is appropriate. If the average number of years likely to experience various flood discharges over a number of sites are desired, the mean probability (expected probability) estimate is appropriate (Thomas, 1976). According to the U.S. Interagency Advisory Committee on Water Data (1982), expected probability is most often used in estimates of annual flood damages and in establishing design flood criteria.

Hardison and Jennings (1972) provide a method for estimating expected probabilities at ungaged sites. This method involves the use of a table in their report that converts exceedence probability to expected probability for selected record lengths (Hardison and Jennings, 1972, table 3, p. 422). For a normal distribution, the expected probabilities in their table can be approximated by the use of the following equation:

$$Prob\{t_{n-1} > K_p[n/(n+0.25)]^{0.5}\}, \qquad (21)$$

where

n is the number of year of annual peaks,

K_p is the standard normal frequency factor for exceedence probability p, and

t_{n-1} is the Student's t-statistic with $n-1$ degrees of freedom.

Thus, expected probability discharges can be computed as follows:

Let

q_T represent the base 10 logarithm of the estimated T-year flood from the regression,

N_e represent the computed equivalent years for the estimate rounded up to nearest integer, and

$t_{Ne-1,p}$ the Student's t-statistics for $N_e - 1$ degrees of freedom and probability, $p = 1/T$.

Then, the expected probability frequency factor, K'_p, is estimated as

$$K'_p = t_{Ne-1,p}[(N_e + 0.25)/N_e], \qquad (22)$$

and the logarithm of the expected probability discharge, $q_{e,T}$, is estimated by

$$q_{e,T} = (q_T - q_2)(K'_p/K_p) + q_2. \qquad (23)$$

The expected probability discharge, $Q_{e,T}$, is computed by raising 10 to the power $q_{e,T}$. The software attached to this report automatically computes the expected-probability discharge for each estimated peak-flow frequency.

National Flood Frequency Program

The National Flood-Frequency (NFF) program (Ries and Crouse, 2002) was developed as a compilation of regression equations used across the United States for estimating flood frequency. In addition to solving regression equations, the program also contains options for plotting flood-flow-frequency curves and synthetic hydrographs, computing estimates of flood-flow frequency for urban basins, and displaying results in English or metric units. The NFF program and these options can be accessed through the menu options titled “Run NFF Program using English Units” and “Run NFF Program using Metric Units” under the “VTFF_Options” menu of the Vermont Flow-Frequency Tool. Some of the highlights of the

NFF program utilities are listed in the following paragraphs. Instructions and a complete description of the NFF program's functions can be found in the help files.

1. The NFF program contains equations for estimating the magnitude and recurrence intervals for floods in urbanized areas based on Sauer and others (1983). The equations, which are considered valid for use throughout the United States, have seven input parameters—drainage area, channel slope, the 2-hour 2-year rainfall, basin storage, percent impervious surface, a basin-development factor, and the discharge at the respective recurrence interval of an equivalent rural basin. The Vermont Flow-Frequency Tool does not determine the additional basin characteristics required for computing estimates of urban flood frequency.
2. The NFF program also contains routines for plotting the frequency curve and a synthetic hydrograph. For the synthetic hydrograph, the user is required to input the basin lag time—the elapsed time, in hours, from the center of the mass of rainfall excess to the peak of the hydrograph. The NFF program then computes the synthetic hydrograph using a dimensionless hydrograph method developed by Stricker and Sauer (1982).

The NFF program is not required by Vermont Flow-Frequency Tool. The Vermont Flow-Frequency Tool writes a file called "current.nff" to the Utilities folder (see section titled "Data Requirements" for description of the Utilities folder), which the NFF program will read when executed through the "VTFF_Options" menu. The NFF program can be downloaded from http://water.usgs.gov/software/nff.html

SUMMARY

This report describes a tool that has been developed by the U.S. Geological Survey, in cooperation with the Vermont Agency of Transportation, for applying regression equations that estimate the peak-flow frequency for recurrence intervals of 2, 5, 10, 25, 50, 100, and 500 years and flow duration for daily discharges exceeded 25, 50, and 75 percent of the time for ungaged, unregulated streams in Vermont. The tool automates the delineation of basin characteristics required by the equations and solves the equations, significantly reducing the processing time normally required when using regression equations to determine flow-frequency manually.

The customized geographical information system (GIS) application, Vermont Flow-Frequency Tool, is adapted from the National Hydrography Dataset toolkit. The functionality of the National Hydrography Dataset provides the basin boundary delineation. Additional programming capabilities included in the Vermont Flow-Frequency Tool evaluate basin characteristics, solve the regression equations, and compute the accuracy of the results and an expected probability adjustment to the streamflow estimates.

Several options are made available in the Vermont Flow-Frequency Tool. One option in the tool allows the recomputation of flow-frequency characteristics adjusted for data from a nearby stream-gaging station that may be on the same stream. Another option provides a link to the utilities of the National Flood-Frequency Program. The National Flood-Frequency program utilities include plotting of frequency curves, plotting of synthetic hydrographs, and computing peak-flow-frequency estimates for urban drainage basins.

SELECTED REFERENCES

Beard, L.R., 1960, Probability estimates based on small normal-distribution samples: Journal of Geophysical Research, v. 65, no. 7, p. 2, 143-2, 148.

Benson, M.A., 1962, Factors influencing the occurrence of floods in a humid region of diverse terrain: U.S. Geological Survey Water-Supply Paper 1580-B, 64 p.

Daly, Christopher, 2000, PRISM (Parameter-elevation Regressions on Independent Slopes Model), U.S. Department of Agriculture and Natural Resources Conservation Service Climate Mapping Project: National Water and Climate Center, Oregon State University Partnership.

Environmental Systems Research Institute, Inc. (ESRI), 1994, ARC/INFO users guides, version 7: Redlands, Calif.

———1996a, ArcView GIS: The geographic information system for everyone: Redlands, Calif., 350 p.

———1996b, Avenue: Customization and application development for ArcView: Redlands, Calif., 259 p.

Grossman, S.I., 1986, Multivariable calculus, linear algebra, and differential equations, Second Edition: New York, Academic Press, Inc., 876 p.

Harbeck, G.E., 1966, Reservoirs in the United States: U.S. Geological Survey Water-Supply Paper 1838, 115 p.

Hardison, C.H., 1971, Prediction error of regression estimates of streamflow characteristics at ungaged sites: U.S. Geological Survey Professional Paper 750-C, p. 228-236.

Hardison, C.H., and Jennings, M.E., 1972, Bias in computed flood risk, Proceedings: Journal of the Hydraulics Division, American Society of Civil Engineers, v. 98, no. HY3, p. 415-427.

Harvey, C.A., and Eash, D.A., 1995, Description, instructions, and verification for Basinsoft, a computer program to quantify drainage-basin characteristics: U.S. Geological Survey Water-Resources Investigations Report 95-4287, 25 p.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier Science Publishing Co., 522 p.

Hodge, S.A., and Tasker, G.D., 1995, Magnitude and frequency of floods in Arkansas: U.S. Geological Survey Water-Resources Investigations Report 95-4224, p. 37-39.

Jennings, M.E., Thomas, W.O., Jr., and Riggs, H.C., 1994, Regional regression equations for estimating magnitude and frequency of floods for ungaged sites, 1993: U.S. Geological Survey Water-Resources Investigations Report 94-4002, 189 p.

Johnson, C.G., and Tasker, G.D., 1974, Progress report on flood magnitude and frequency of Vermont streams: U.S. Geological Survey Open-File Report 74-130, 37 p.

Maidment, D.R., 1992, Handbook of hydrology: New York, McGraw-Hill, Inc., 29 chap.

Natural Resources Canada, 2002, National Topographic Data Base, 1:50:000-scale, accessed May 13, 2002, at URL<http://www.cits.rncan.gc.ca/.

Natural Resources Conservation Service, 1996, Watershed boundaries of Vermont based on U.S. hydrologic units, accessed December 21, 2001, at URL < http://www.vcgi.org/pub/dataware/gisdata/layers_vcgi/WaterHydro_WATSHED.exe

Olson, S.A., 2002, Flow-frequency characteristics of Vermont streams: U.S. Geological Survey Water-Resources Investigations Report 02-4238, 47 p.

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 plates, scale 1:1,000,000.

Ries, K.G., III, and Crouse, M.Y., 2002, The National Flood Frequency Program, version 3: A computer program for estimating magnitude and frequency of floods for ungaged sites: U.S. Geological Survey Water-Resources Investigations Report 02-4168, 41 p.

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

———1968b, Frequency curves: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A2, 15 p.

———1968c, Regional analyses of streamflow characteristics: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. B3, 15 p.

SAS Institute, Inc., 1990, SAS/STAT User's guide (vers. 6, 4th ed.): Cary, N.C., v. 1 and v. 2, 1,686 p.

Sauer, V.B., 1974, Flood characteristics of Oklahoma streams: U.S. Geological Survey Water-Resources Investigations Report 52-73, 301 p.

Sauer, V.B., Thomas, W.O., Jr., Stricker, V.A., and Wilson, K.V., 1983, Flood characteristics of urban watersheds in the United States: U.S. Geological Survey Water-Supply Paper 2207, 63 p.

Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis 1. Ordinary, weighted, and generalized least squares compared: Water Resources Research, v. 21, no. 9, p. 1,421-1,432.

Steeves, P.A., 2002, NHD Applications: NHD Watershed, accessed September 2002, at URL <http://nhd.usgs.gov/app_page.htm

Stricker, V.A., and Sauer, V.B., 1982, Techniques for estimating flood hydrographs for ungaged urban watersheds: U.S. Geological Survey Open-File Report 82-365, 24 p.

Tasker, G.D., 1986, Generating efficient gaging plans for regional information, *in* Moss, M.E., ed., Integrated design of hydrologic networks, Budapest Symposium, July 1986, Proceedings: International Association of Hydrologic Sciences, Publication 158, p. 269-281.

———1987, Theory and application of generalized least squares, *in* Medina, K.D., Analysis of surface-water data network in Kansas for effectiveness in providing regional streamflow information: U.S. Geological Survey Water-Supply Paper 2303, p. 24-26.

Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: Journal of Hydrology, v. 3, p. 361-375.

Thomas, D.M., 1976, Flood frequency: Expected and unexpected probabilities: U.S. Geological Survey Open-File Report 76-775, 7 p.

U.S. Geological Survey, 2001a, STATSGO soil characteristics for the conterminous United States: U.S. Geological Survey Open-File Report 656, accessed December 21, 2001, at URL <http://water.usgs.gov/GIS/metadata/usgswrd/muid.html

———2001b, National elevation dataset, accessed December 21, 2001, at URL <http://edcnts12.cr.usgs.gov/ned/.

U.S. Geological Survey and U.S. Environment Protection Agency, 2002, National hydrography dataset, accessed June 18, 2002, at URL <http://nhd.usgs.gov/.

U.S. Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood-flow frequency, Bulletin 17-B of the Hydrology Subcommittee: Reston, Va., U.S. Geological Survey, Office of Water Data Coordination, 183 p.

APPENDIX 1. USE OF REGRESSION EQUATIONS WITH METRIC UNITS

APPENDIX 1. USE OF REGRESSION EQUATIONS WITH METRIC UNITS

The equations used to estimate peak flow at recurrence intervals of 2, 5, 10, 25, 50, 100, and 500 years and the magnitude of daily-mean discharges exceeded 25, 50, and 75 percent of the time when using metric units for drainage area and precipitation are

$$Q_2 = 0.452A^{0.914}L^{-0.294}E^{0.0776}Y^{-0.180}, \qquad (1)$$

$$Q_5 = 0.742A^{0.902}L^{-0.295}E^{0.0835}Y^{-0.253}, \qquad (2)$$

$$Q_{10} = 0.962A^{0.897}L^{-0.302}E^{0.0890}Y^{-0.298}, \qquad (3)$$

$$Q_{25} = 1.30A^{0.883}L^{-0.316}E^{0.104}Y^{-0.349}, \qquad (4)$$

$$Q_{50} = 1.59A^{0.874}L^{-0.327}E^{0.115}Y^{-0.385}, \qquad (5)$$

$$Q_{100} = 1.90A^{0.865}L^{-0.336}E^{0.125}Y^{-0.420}, \qquad (6)$$

$$Q_{500} = 2.75A^{0.846}L^{-0.355}E^{0.148}Y^{-0.497}, \qquad (7)$$

$$D_{75} = (4.28\times10^{-8})A^{1.08}P^{1.55}E^{0.101}, \qquad (8)$$

$$D_{50} = 9.73\times10^{-8}A^{1.04}P^{1.58}E^{0.0603}, \text{ and} \qquad (9)$$

$$D_{25} = (3.08\times10^{-7})A^{1.01}P^{1.55}E^{0.0438}. \qquad (10)$$

where

Q_n is the calculated peak flow for recurrence interval *n*, in cubic meters per second;

D_n is the estimated daily discharge exceeded *n* percent of the time, in cubic meters per second;

A is the drainage area of the watershed delineated using a GIS (boundaries for the drainage areas from a 12-digit hydrologic unit coverage (Natural Resources Conservation Service, 1996) or digitized when a watershed’s boundary was not defined by this GIS coverage), in square kilometers;

L is the area of lakes and ponds in a watershed as a percent of drainage area, plus 1 percent, determined using a GIS from the 1:24,000 scale National Hydrography Dataset (U.S. Geological Survey and U.S. Environmental Protection Agency, 2002) using all features in the dataset that were lake or pond boundaries including ponds as small as several hundreths of an acre and pooled areas of streams;

E is the percent of the drainage basin that is at or greater than 365.7 meters (1,200 ft) altitude, plus 1 percent, computed using a GIS from the National Elevation Dataset (U.S. Geological Survey, 2001);

Y is the northing of the centroid of the drainage basin determined with a GIS, in the Vermont State Plane Coordinate System, divided by 100,000, then increased by 1.0; and

P is the basinwide mean of the mean-annual precipitation, in millimeters, determined using a GIS and the PRISM dataset (Daly, 2000) resampled with bilinear interpolation to a 30-meter-cell resolution using the ESRI ARC/INFO RESAMPLE command (Environmental Systems Research Institute, Inc., 1994). Canadian precipitation data (Ghislain Jacques, Environment Quebec, written commun., January 17, 2002) was merged with the PRISM precipitation data where needed along the northern Vermont border.

SELECTED REFERENCES

Daly, Christopher, 2000, PRISM (Parameter-elevation Regressions on Independent Slopes Model): U.S. Department of Agriculture and Natural Resources Conservation Service Climate Mapping Project, National Water and Climate Center, Oregon State University Partnership.

Environmental Systems Research Institute, Inc. (ESRI), 1994, ARC/INFO users guides, version 7: Redlands, Calif.

Natural Resources Conservation Service, 1996, Watershed boundaries of Vermont based on U.S. hydrologic units, accessed December 21, 2001, at URL http://www,vcgi.org/pub/dataware/gisdata/layers_vcgi/WaterHydro_WATSHED.exe.

U.S. Geological Survey, 2001, National elevation dataset, accessed December 21, 2001, at URL http://edcnts12.cr.usgs.gov/ned/.

U.S. Geological Survey and U.S. Environmental Protection Agency, 2002, National hydrography dataset, accessed June 18, 2002, at URL http://nhd.usgs.gov/.