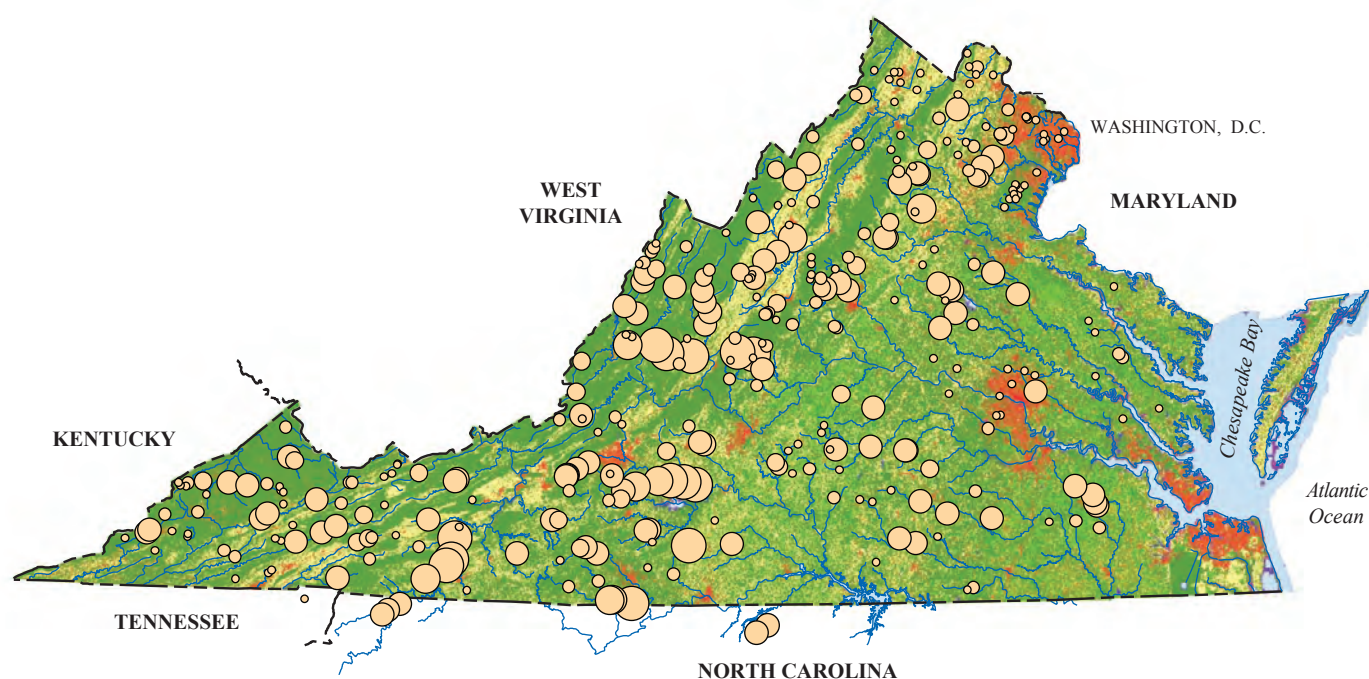


Prepared in cooperation with the Virginia Department of Environmental Quality

# Methods for Estimating Drought Streamflow Probabilities for Virginia Streams



Scientific Investigations Report 2014–5145

**Cover.** Map showing location of 339 selected streamflow study sites used to develop maximum likelihood logistic regression models for estimating drought streamflow probabilities for 259 hydrologic basins in Virginia. Taken from figure 1 (p. 3).

# **Methods for Estimating Drought Streamflow Probabilities for Virginia Streams**

By Samuel H. Austin

Prepared in cooperation with the Virginia Department of Environmental Quality

Scientific Investigations Report 2014–5145

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

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

SALLY JEWELL, Secretary

**U.S. Geological Survey**

Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2014

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit <http://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod>

To order this and other USGS information products, visit <http://store.usgs.gov>

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

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Austin, S.H., 2014, Methods for estimating drought streamflow probabilities for Virginia streams: U.S. Geological Survey Scientific Investigations Report 2014–5145, 20 p., <http://dx.doi.org/10.3133/sir20145145>.

ISSN 2328-0328 (online)

ISSN 2328-031X (print)

ISBN 978-1-4113-3855-5

## Acknowledgments

The author thanks Rob Burgholzer and Scott Kudlas of the Virginia Department of Environmental Quality, whose continued collaboration and assistance made this report possible.

George E. Harlow, Jr., of the U.S. Geological Survey is thanked for his advice and review of this report, and the author appreciates Jennifer Krstolic and Brian A. Hasty of the U.S. Geological Survey for developing a map for this report.

# Contents

Acknowledgments .....	iii
Abstract .....	1
Introduction .....	1
Purpose and Scope .....	1
Streamflow Requirements During Dry Times .....	2
Why Predict Drought Streamflow Probabilities? .....	2
Previous Studies .....	2
Methods of Study .....	2
Site Selection.....	2
Data Evaluation .....	2
Maximum Likelihood Logistic Regression .....	3
Reading Maximum Likelihood Logistic Regression Plots .....	6
Model Development .....	7
Model Verification and Diagnostics .....	9
Drought Streamflow Probability Estimates .....	10
Selection of an Equation.....	10
Application of the Equations with Examples .....	11
Example 1. Predicting the probability of exceeding a 5-percent July drought streamflow quantile threshold as a function of mean monthly streamflow from the previous December.....	11
Example 2. Predicting the probability of exceeding a 10-percent August drought streamflow quantile threshold as a function of mean monthly streamflow from the previous January.....	12
Example 3. Predicting the probability of exceeding a 25-percent September drought streamflow quantile threshold as a function of mean monthly streamflow from the previous February.....	14
Example 4. Predicting the probability of exceeding a 50-percent July drought streamflow quantile threshold as a function of averaged mean monthly streamflow from the previous November, December, January, and February .....	16
Summary.....	18
Selected References.....	19

## Figures

1. Map showing location of 339 selected streamflow study sites used to develop maximum likelihood logistic regression models for estimating drought streamflow probabilities for 259 hydrologic basins in Virginia .....	3
2. Three categories of maximum likelihood logistic regression curve shape illustrating the degree of relation strength.....	5
3. Three “zones of predictive power” for each maximum likelihood logistic regression plot, illustrating areas of strong predictive power <i>A</i> and <i>C</i> , and relatively weak predictive power <i>B</i> .....	6
4. Example MLLR drought threshold probability plot showing the chances of exceeding, and not exceeding, an historical July 25-percent daily streamflow threshold as a function of the average of mean daily streamflow values from November, December, January, and February .....	7
5. Diagram identifying MLLR probability models developed for evaluation and testing .....	8
6. MLLR drought threshold probability plot showing the chances of exceeding, and not exceeding, an historical July 5-percent daily streamflow threshold as a function of mean daily streamflow values from December .....	13
7. MLLR drought threshold probability plot showing the chances of exceeding, and not exceeding, an historical August 10-percent daily streamflow threshold as a function of mean daily streamflow values from January .....	14
8. MLLR drought threshold probability plot showing the chances of exceeding, and not exceeding, an historical September 25-percent daily streamflow threshold as a function of mean daily streamflow values from February .....	16
9. MLLR drought threshold probability plot showing the chances of exceeding, and not exceeding, an historical July 50-percent daily streamflow threshold as a function of the average of mean daily streamflow values from November, December, January, and February.....	18

## Tables

[Tables downloadable from <http://dx.doi.org/10.3133/sir20145145>, or from CD-ROM in back of printed report]

1. Maximum likelihood logistic regression equations for estimating drought streamflow threshold probabilities for Virginia streams
2. Range of maximum likelihood logistic regression explanatory variable values (x-values) and daily value quantiles for the period of record used for estimating drought streamflow threshold probabilities for Virginia streams

## Conversion Factors

Inch/Pound to SI

<b>Multiply</b>	<b>By</b>	<b>To obtain</b>
<b>Length</b>		
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
<b>Area</b>		
acre	0.4047	hectare (ha)
square foot (ft <sup>2</sup> )	0.09290	square meter (m <sup>2</sup> )
square mile (mi <sup>2</sup> )	259.0	hectare (ha)
<b>Volume</b>		
gallon (gal)	0.003785	cubic meter (m <sup>3</sup> )
cubic foot (ft <sup>3</sup> )	0.02832	cubic meter (m <sup>3</sup> )
acre-foot (acre-ft)	1,233	cubic meter (m <sup>3</sup> )
<b>Flow rate</b>		
acre-foot per day (acre-ft/d)	0.01427	cubic meter per second (m <sup>3</sup> /s)
foot per second (ft/s)	0.3048	meter per second (m/s)
cubic foot per second (ft <sup>3</sup> /s)	0.02832	cubic meter per second (m <sup>3</sup> /s)
<b>Hydraulic gradient</b>		
foot per mile (ft/mi)	0.1894	meter per kilometer (m/km)

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.



# Methods for Estimating Drought Streamflow Probabilities for Virginia Streams

By Samuel H. Austin

## Abstract

Maximum likelihood logistic regression model equations used to estimate drought flow probabilities for Virginia streams are presented for 259 hydrologic basins in Virginia. Winter streamflows were used to estimate the likelihood of streamflows during the subsequent drought-prone summer months. The maximum likelihood logistic regression models identify probable streamflows from 5 to 8 months in advance. More than 5 million streamflow daily values collected over the period of record (January 1, 1900 through May 16, 2012) were compiled and analyzed over a minimum 10-year (maximum 112-year) period of record. The analysis yielded the 46,704 equations with statistically significant fit statistics and parameter ranges published in two tables in this report. These model equations produce summer month (July, August, and September) drought flow threshold probabilities as a function of streamflows during the previous winter months (November, December, January, and February). Example calculations are provided, demonstrating how to use the equations to estimate probable streamflows as much as 8 months in advance.

## Introduction

Planning for drought conditions in Virginia streams is essential to the sound management of water resources and associated riparian and watershed ecosystems. Reliable estimations of the likelihood that streamflows during drought-prone months will exceed specific low-flow thresholds can provide advance warning of drought conditions, allowing extended lead times for improved drought awareness and effective management response. Improved knowledge and estimation of low flows in drought-prone months provide extended lead time for drought response even as precipitation, water withdrawals, land uses, and climate variables change over time. Drought streamflow probability estimates can be used for the management of riparian systems and the ecology associated with forested, agricultural, and urban landscapes. Drought streamflow probability estimates provide a basis for analysis of future streamflow response to changes in ecosystem and climate.

This study was conducted through a cooperative partnership between the Virginia Department of Environmental Quality (DEQ) and the U.S. Geological Survey (USGS). The DEQ and the USGS maintain a statewide network of more than 290 continuous streamgaging stations with many additional partial-record streamgaging stations. Together USGS and DEQ develop computations, analyses, and tools to support the understanding and effective management of streamflows. For this study the probabilities that streamflows in July, August, and September would exceed specific drought low-flow thresholds, as a function of streamflows during the previous November, December, January, and February, were estimated. The low-flow characteristics for each station were described previously by Austin and others (2011a). Maximum likelihood logistic regression (MLLR) was used to determine the probability of exceeding a set of flow thresholds.

Empirically determined MLLR equations were used to estimate drought streamflow probabilities for Virginia streams. Winter streamflows which correspond to the critical time for the majority of effective recharge to the groundwater system (Nelms and Moberg, 2010a, b) were used to estimate the likelihood of streamflows, composed mostly of base flow from groundwater discharge (Harlow and others, 2005; Nelms and Moberg, 2010b), during drought-prone summer months. This information aids in the understanding of water availability and the environmental health of watersheds and their associated ecosystems, and extends the lead-time of drought response by as much as 8 months.

## Purpose and Scope

This report presents a set of 46,704 predictive equations that describe the likelihood of streamflow during summer months exceeding each low-flow threshold, based on previous streamflows at 259 streamgaging stations for a minimum of 10 years and a maximum of 112 years. The development of the equations is described, and examples are provided to demonstrate how to use the equations to estimate probable streamflow as much as 8 months in advance. The predictive equations are listed in a table. MLLR drought threshold probabilities are shown in illustrations.

## Streamflow Requirements During Dry Times

Streamflows support fluvial processes and ecosystem functions that are particularly important during dry times and include maintaining channel shape (morphology). The dimension, pattern, and profile of a stream contribute to the physical properties associated with channel function, such as the movement of sediments and maintenance of streamwater width and depth. Ecosystems and habitat for aquatic organisms require these and other channel functions to maintain transportation corridors and sources of food and water for aquatic and terrestrial animals and plants.

Humans withdraw streamwater for industrial, agricultural, and domestic uses. Streamflows are an integrated response to the conditions upstream created by natural and anthropogenic sources, including precipitation received in previous months. When streamflows are low and available water is scarce, concern for maintaining water uses is heightened. Estimates of water availability allow water withdrawals to be adjusted, helping to ensure that fluvial processes, ecosystem functions, and human uses are maintained.

## Why Predict Drought Streamflow Probabilities?

Estimating drought streamflow probabilities allows for the anticipation of water availability during critical periods. Estimates facilitate improved planning for, and management of, water resources, particularly during intervals when streamflows are low. The benefits of accurate drought streamflow probabilities include (1) extended lead-times for drought response, as much as 8 months, (2) improved understanding of low flows in drought-prone months, and (3) accurate anticipation of future low flows as functions of change in precipitation, water withdrawals, land-use management, and climate. These benefits enhance management of water withdrawals and land uses, so that water availability may be sustained during dry months.

## Previous Studies

Previous studies have reported low-flow characteristics or analyzed low-flow data for Virginia streams. A study by Nuckels (1970) analyzed data from long-term, continuous-record streamgaging stations throughout Virginia. Hayes (1991) provides an analysis of low-flow characteristics in Virginia streams. Austin and others (2011a) provide a comprehensive analysis of low-flow characteristics in Virginia streams. Other studies, including those by Cushing and others (1973), Trainer and Watkins (1975), Mohler and Hagan (1981), Smith (1981), Wetzel and Bettendorff (1986), Lynch (1987), and Lynch and others (1987), focused on analysis of streamflow characteristics in portions of the State.

## Methods of Study

Analysis was conducted for five areas of focus: site selection, data evaluation, MLLR, model development, and model verification and diagnostics. Each of these focus areas is described below.

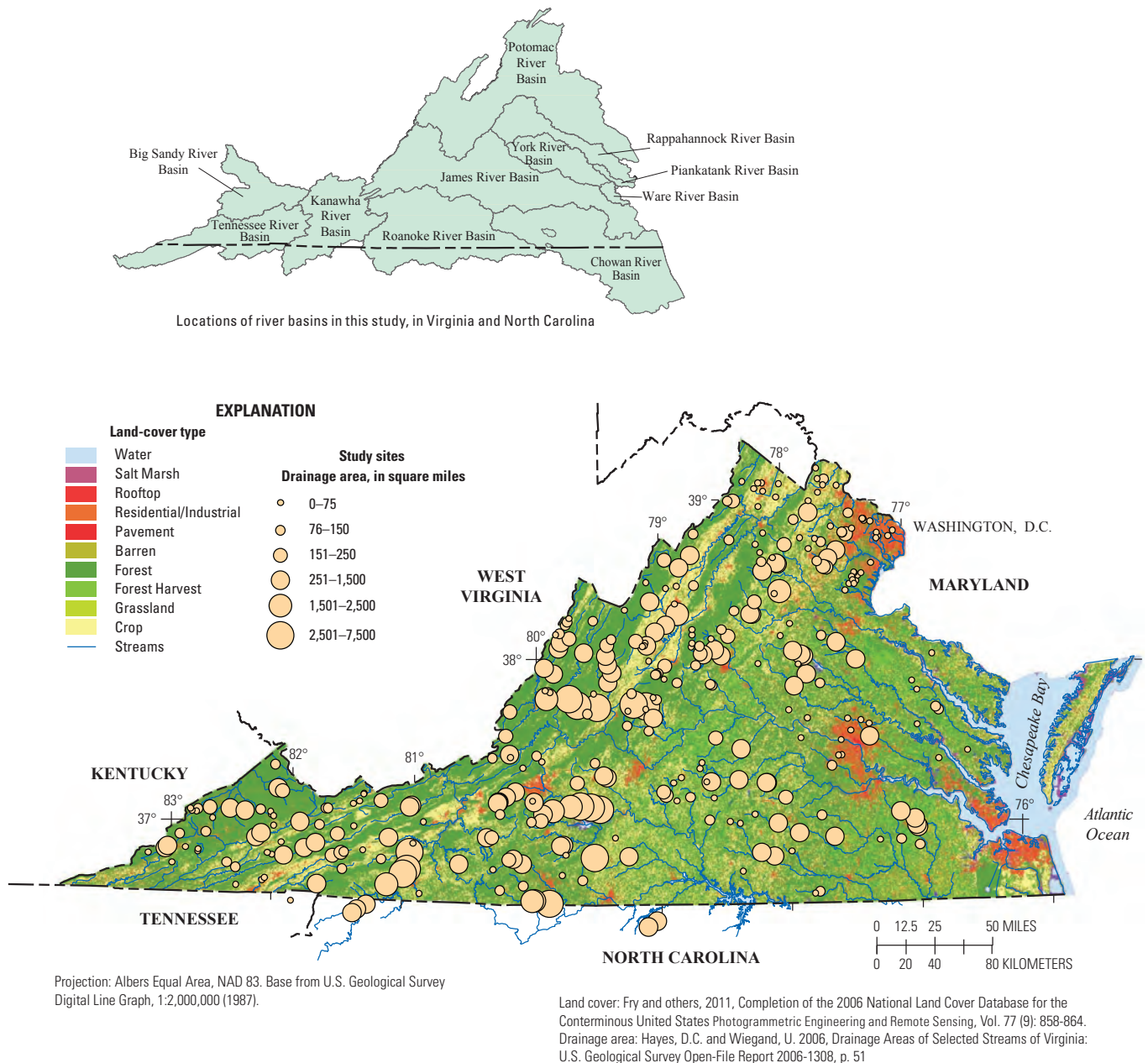
### Site Selection

A total of 361 Virginia streamgaging stations registered in the USGS National Water Information System (NWIS) were considered for analysis. Of these, 22 streamgaging stations were identified as not useful for estimating drought streamflow probabilities, including 20 streamgaging stations for which no daily value (DV) data are available in NWIS and 2 streamgaging stations with insufficient DV data in NWIS (less than 1 year of recorded DV data), when streamgaging station data were accessed on May 17, 2012. More than 5.12 million streamflow daily values (DV) collected over the period of record and compiled for the remaining 339 streamgaging stations were obtained from NWIS. These data and their supporting information were used to develop MLLR models to predict drought streamflow probabilities. Of these 339 streamgaging stations, 259 produced models with accompanying p-value statistics that are less than or equal to 0.05 based on streamflow data spanning a minimum 10-year (maximum 112-year) period of record. Models predicting drought streamflow probabilities for these 259 basins are presented in this report.

### Data Evaluation

Daily values of streamflow were grouped by streamgaging station for analysis, then organized into categories corresponding to the 11 major river basins in Virginia—Big Sandy, Chowan, James, Kanawha, Piankatank, Potomac, Rappahannock, Roanoke, Tennessee, Ware, and York (fig. 1). Within each river basin category, data for each streamgaging station were placed in one of two groups—(1) streamgaging stations for which flow duration statistics had been previously developed and (2) streamgaging stations without previously developed flow duration statistics. This helped to speed data evaluation by facilitating analysis of data for streamgaging stations with previously developed flow duration statistics.

Data were inspected and evaluated for each streamgaging station in an iterative fashion. Each list of daily values was visually scanned for errors, then after tabulation, daily values were scrutinized a second time. Each dataset was reviewed again as models were developed, and for selected models, plots of data in probability space produced as part of modeling output were evaluated for consistency with a previous tabulation of the same data. Finally, datapoints were plotted on each modeled MLLR plot to facilitate continued evaluation.



**Figure 1.** Location of 339 selected streamflow study sites used to develop maximum likelihood logistic regression models for estimating drought streamflow probabilities for 259 hydrologic basins in Virginia.

## Maximum Likelihood Logistic Regression

MLLR is used to predict the chance of exceeding, or not exceeding, a particular streamflow threshold. MLLR describes the odds of a particular outcome on the basis of the value of a predictive variable and provides a method to estimate the probability of a response level as a smooth function of a factor. For this study, the probability of streamflow exceeding a predefined drought flow threshold during dry months of the year was estimated as a function of mean streamflow during the previous wet months of the year. The regression is called a logistic regression because an S-shaped logistic curve is

used to fit the probabilities. This type of fitted curve is called a logistic function (Sall and others, 2007).

In its simplest form, the response (y-value) of a MLLR has two levels, and is, therefore, a binary response. In this study, the binary response is an indication of whether a drought streamflow threshold will, or will not, be exceeded. A “yes” response indicates that a particular drought streamflow threshold will be exceeded, and a “no” response indicates that a particular drought streamflow threshold will not be exceeded. The logistic curve is fitted on the basis of the difference in the logs of the two probabilities of the yes and no binary response, expressed as a linear function of the factor



#### 4 Methods for Estimating Drought Streamflow Probabilities for Virginia Streams

variable. In this case, the factor variable is the mean streamflow during the wet months of the year.

If  $p$  denotes the probability of the first response level, “yes,” then  $1-p$  denotes the probability of the second response level, “no.” The linear model may then be written as shown in equation 1, an example of a logistic model in linear form.

Equation 1: Example of a logistic model in linear form.

$$\begin{aligned} \log(p) - \log(1-p) &= \beta_0 + \beta_1 \cdot X \quad \text{or} \\ \log\left(\frac{p}{1-p}\right) &= \beta_0 + \beta_1 \cdot X, \end{aligned} \quad (1)$$

where

$\log\left(\frac{p}{1-p}\right)$  is called the logit of  $p$ , also known as the log odds-ratio,

- $p$  is the probability of exceeding a particular streamflow threshold,
- $\beta_0$  is an intercept parameter,
- $\beta_1$  is a slope parameter, and
- $X$  is an explanatory variable (x-axis value).

There is no error term associated with equation 1 because the predicted value (y-value) is not a response level as usually is expected in a linear regression model but rather a probability distribution of a response level. Because a probability distribution describes a chance, or likelihood, of either of two outcomes (yes or no) relative to an explanatory variable (x-value), an error cannot be associated with the fitted curve in the manner commonly identified with, and expressed in, regression models. Just as a weather forecast predicting a 90-percent chance of rain is not incorrect if rain does not occur, so too the MLLR models are correct even when the less likely of this study’s two possible outcomes occurs (Sall and others, 2007).

There is, however, a probability value (p-value) associated with each MLLR model, indicating the strength of each probability distribution relative to the likelihood that the distribution could simply be a product of chance. Each p-value is identified with the help of a Likelihood-ratio Chi-Square (Chi-Square) test for the hypothesis that all regression parameters are zero. The Chi-Square is computed by taking twice the difference in negative log-likelihoods ( $-2 \log L$ ) between the results of the fitted MLLR model and the results of the reduced version of the model that has no explanatory (x) variable, only intercepts. This score statistic is a function of the first and second derivatives of the log-likelihood function (Allison, 2012). The p-value is determined as the probability of obtaining a greater Chi-Square value by chance alone, if the fit of the specified model is no better than the fit of the reduced model that includes only intercepts. We can think of this as asking, “Is the MLLR probability model better than nothing?” Low p-values indicate that, yes, the MLLR probability model is highly significant and better than nothing (Allison, 2012).

The accounting to determine each modeled probability is done by summing the negative logarithms of the probabilities attributed by the model to the events that actually did occur. For example, if  $p$  represents a drought streamflow threshold probability, then the MLLR model assigns a value of  $-\log(p)$  to represent a chance of “yes”—the drought streamflow threshold will be exceeded—and a value of  $-\log(1-p)$  to represent a chance of “no”—the drought streamflow threshold will not be exceeded. A model that predicts perfectly yields  $p = 1$  when the drought streamflow threshold is exceeded since  $-\log(p) = 0$  when  $p = 1$ . This same perfectly predicting model yields  $p = 0$  when the drought streamflow threshold is not exceeded because  $-\log(1-p) = 0$  when  $p = 0$ . A value for the negative logarithm of the probability that is equal to zero (0) indicates that a perfect prediction has been made. That is to say, if  $-\log(p) = 0$ , then a perfect prediction has been made. Why does  $-\log(p)$  equal to 0 mean a perfect prediction? We can think about it this way: If we tried reversing the assignments of these negative logarithms of the probabilities such that a probability of zero (0) is attributed to an event that had actually occurred, rather than attributing a probability of one (1) to that event, then the  $-\log$ -likelihood of the event would be infinity because when  $p = 0$  then  $-\log(p) = \text{infinity } (\infty)$ . Because the accounting mechanism to determine modeled probability minimizes the sum of the negative logarithms of each modeled probability, that is, minimizes the sum of each  $-\log(p)$ , the sum of these negative logarithms ( $-\log$ -likelihood) would be infinity in this reversed assignment case, yielding a very high minimum value and therefore a very bad forecasting score. However, because the assignments of probability attributes are actually not reversed, attributing a probability of one (1) to an event that has actually occurred yields a very low  $-\log$ -likelihood value because when  $p = 1$ ,  $-\log(p) = \text{zero } (0)$ . This yields a low minimum value, which is a good forecasting score (Sall and others, 2007).

As described in equation 1, the logit of the logistic model is  $\log\left(\frac{p}{1-p}\right)$ , which is the inverse logit of the model  $\beta_0 + \beta_1 \cdot X$ .

The logit expresses the probability of each response level as a function of an explanatory variable (x-value), and these estimates are in the MLLR so as to maximize the likelihood of their occurrence. This is equivalent to minimizing the negative log-likelihood ( $-\log$ -likelihood) of their occurrence, which is the negative sum of logs of the probabilities attributed to the response level (“yes” or “no”) for each streamflow observation. Therefore, plots of each MLLR trace a curve that solves for  $p$  in equation 1, which for a “yes” response is equivalent to

Equation 2: Logistic model of a “yes” response

$$p_1 = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot X)}} \quad \text{and} \quad (2)$$

and for a “no” response is equivalent to

Equation 3: Logistic model of a “no” response

$$p_2 = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot X)}} \quad (3)$$

where

- $e$  is the base of the natural logarithm,
- $p_1$  is the probability of exceeding a particular streamflow threshold (y-axis value is “yes”),
- $p_2$  is the probability of not exceeding a particular streamflow threshold (y-axis value is “no”),
- $\beta_0$  is an intercept parameter,
- $\beta_1$  is a slope parameter, and
- $X$  is an explanatory variable (x-axis value).

Equation 2 evaluates the probability of a “yes” response. Equation 3 evaluates the probability of a “no” response. The probability of a “yes” response is equivalent to  $1-p_2$ , and the probability of a “no” response is equivalent to  $1-p_1$  because the “yes” response and the “no” response must sum to 1 (Sall and others, 2007).

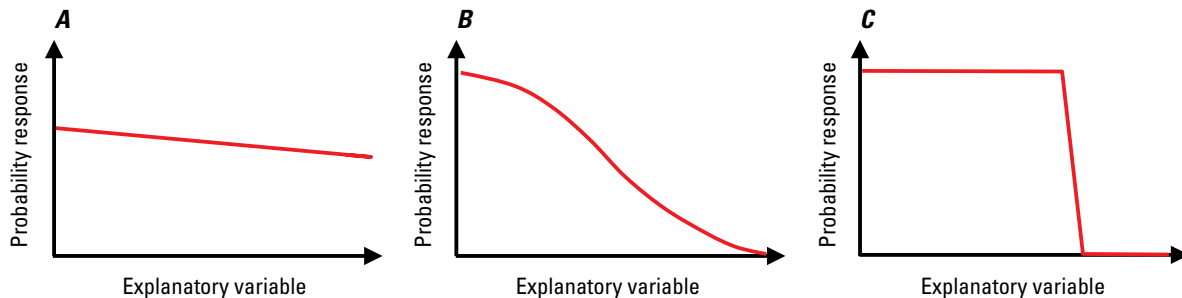
The strength of the x–y relation in a maximum likelihood logistic regression may be identified using three categories of curve shape, each illustrating a typical degree of relation strength. The phrase “strength of x–y relation” is deliberately used rather than “goodness of fit” because we are predicting probabilities, and all equation fits are equally “good.” However, predictions of the probability of a “yes” or a “no” response relative to an explanatory (x-axis) variable may be more distinct and, therefore, potentially more useful for decision making as a consequence of the shape of the MLLR response curve.

Three useful categories of MLLR response curve shape are illustrated in figure 2.

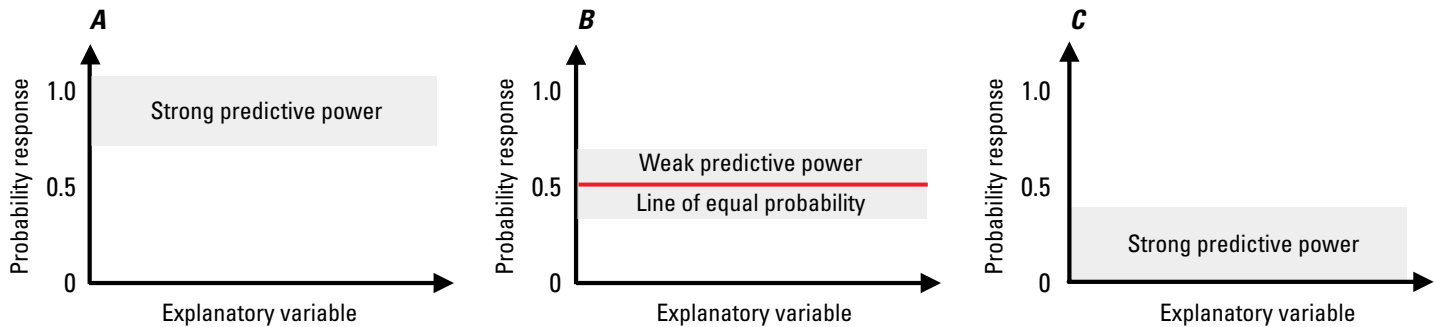
When the fit strength of an x–y relation is weak, the absolute value of the slope parameter in the MLLR model is

relatively small. This results in a gentle slope of the MLLR line over the range of the explanatory data, as illustrated by curve *A* in figure 2. When the fit strength of the x–y relation is strong, the absolute value of the slope parameter in the MLLR model is relatively large. This results in a steeper slope of the MLLR line over the range of the explanatory data, as illustrated by curve *B* in figure 2. In curve *B*, the logistic (s-shaped) nature of the regression line also becomes more pronounced. When the fit strength of the x–y relation is approaching a “perfectly definitive fit,” the absolute value of the slope parameter in the MLLR model becomes quite large, approaching infinity as the middle line segment of the curve approaches an orientation perpendicular to the x-axis of the graph. This results in an s-shaped MLLR line with sharper breaks in inclination and a section of near vertical slope, as illustrated by curve *C* in figure 2. With a highly definitive gradient such as this, distinctions between the y-axis categorical probabilities of “yes” and “no” become quite clear relative to the x-axis explanatory variable. With a perfectly definitive fit, the probability of “yes” changes from 0 to 1, and the probability of “no” changes from 1 to 0, at the same position on the x-axis. A perfectly definitive fit means that, before a certain value of  $X$ , all the responses are at one level (either a level of “yes” or “no”), and after that same value of  $X$ , all the responses are at another level (either a level of “yes” or “no”) (Sall and others, 2007).

In addition to the nuances of curve shape as a useful indicator of x–y relation fit strength, curve position within the probability space of the MLLR plot helps identify the predictive power of each relation. Three zones of predictive power may be associated with each MLLR curve—two zones of strong predictive power and one zone of weak predictive power. These are illustrated in figure 3.



**Figure 2.** Three categories of maximum likelihood logistic regression curve shape illustrating the degree of relation strength. Curve *A* illustrates a weak relation between the explanatory variable (x-value) and the probability response (y-value). Curve *B* illustrates a strong relation. Curve *C* illustrates an even stronger relation, a perfectly definitive fit (after Sall and others, 2007).



**Figure 3.** Three “zones of predictive power” for each maximum likelihood logistic regression plot, illustrating areas of strong predictive power *A* and *C*, and relatively weak predictive power *B*.

Thus, relation fit strength identified in figure 2 and predictive power identified in figure 3 may be used together to further refine the understanding of MLLR plots. A curve that moves through the three zones identified in figure 3 exhibits a strong *x*–*y* relation with strong predictive power. Figure 2*B* and figure 2*C* are examples of such a curve. A curve confined to the upper one-third or the lower one-third of an MLLR plot exhibits strong predictive power, as identified in figure 3*A* and figure 3*C*; a weak *x*–*y* relation is shown in figure 2*A*. Finally, a curve confined to the middle one-third of an MLLR plot exhibits weak predictive power, as identified in figure 3*B*, a line of equal probability defines a 0.5 probability of either response, and a weak *x*–*y* relation is shown in figure 2*A*.

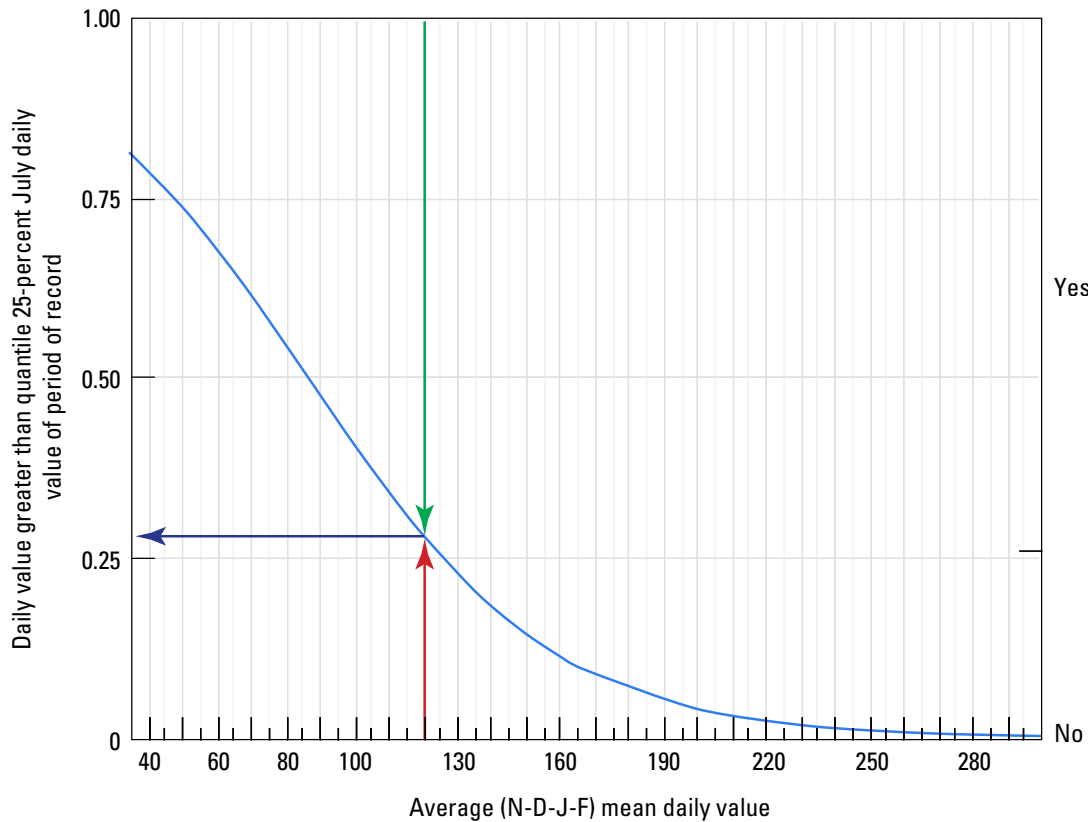
## Reading Maximum Likelihood Logistic Regression Plots

MLLR plots may be thought of as cumulative probability plots, where the line of fit partitions the entire probability space into response categories. The two response categories in the drought threshold probability plots—the probability of exceeding a particular drought threshold labeled as “yes” on the right vertical axis of the plot and the probability of not exceeding the same particular drought threshold labeled as “no” on the right vertical axis of the plot—may each be identified using the left vertical axis of the plot, as shown in figure 4. The probability of “no” is read directly from the left vertical axis of the plot. The probability of “yes” is equivalent to 1 minus the value read directly from the left vertical axis of the plot and represents the distance from the regression line to the top of the graph box. The probability of “yes,” therefore, equals 1 minus the probability of “no” or, using the variables identified in equation 2 and equation 3,  $p_1 = (1 - p_2)$ .

Figure 4 illustrates how to read MLLR plots. It is an example MLLR drought threshold probability plot showing the chance of exceeding and not exceeding a July 25-percent daily streamflow threshold as a function of the average of mean daily values from November, December, January, and February. On the graph, the line of fit partitions the whole

probability space into two response categories. A “no” response category occupies the area below the blue line, and a “yes” response category occupies the area above the blue line. The division into two areas is identified with the labels “Yes” and “No” on the right vertical axis. The probability of not exceeding a drought flow threshold (labeled “No”) may be read directly from the left vertical axis. For example, the probability of not exceeding a July daily streamflow threshold equal to 25-percent of historic July daily streamflow, when the average daily streamflow during the previous November, December, January, and February was 120 cubic feet per second (ft<sup>3</sup>/s), may be read from the graph as follows: (1) Enter the explanatory variable axis (*x*-axis) at a value of approximately 120 ft<sup>3</sup>/s, then (2) move vertically to the blue regression line, and (3) proceed horizontally to the left vertical axis (*y*-axis), then (4) read the probability value on the left vertical axis. This sequence of steps yields a probability value of approximately 0.26, as illustrated by the red and blue arrows in figure 4.

Similarly, the probability of exceeding a July daily streamflow threshold (labeled “Yes”) equal to 25-percent of historic July daily streamflow, when the average daily streamflow during the previous November, December, January, and February was 120 ft<sup>3</sup>/s, may be traced in like manner from the top of the graph-box downward and determined using simple subtraction, as follows: (1) Find the vertical above the blue regression line that corresponds to the vertical below the blue regression line associated with the explanatory variable (*x*) value of interest, in this case again, a value of approximately 120 ft<sup>3</sup>/s, (2) move vertically downward from the top of the graph box to the blue regression line, and (3) proceed horizontally to the left vertical axis, (4) read the probability value on the left vertical axis, and (5) subtract this value from the left vertical axis value of 1.00 at the top of the graph box. This sequence of steps yields a probability value of approximately 1.00 minus 0.26, which equals 0.74, as illustrated by the green and blue arrows in figure 4. For each explanatory variable value on the *x*-axis, a “yes” probability value is equal to 1 minus a “no” probability value, and a “no” probability value is equal to 1 minus a “yes” probability value (see fig. 4).



**Figure 4.** An example MLLR drought threshold probability plot showing the chances of exceeding, and not exceeding, an historical July 25-percent daily streamflow threshold as a function of the average of mean daily streamflow values from November, December, January, and February (N-D-J-F). Reading the 0.26 non-exceedance probability (“No”-value) of an explanatory variable value (x-axis value) of 120 ft<sup>3</sup>/s is demonstrated by red and blue arrows. Reading the 0.74 (1.00 minus 0.26) exceedance probability (“Yes”-value) of an explanatory variable value (x-axis value) of 120 ft<sup>3</sup>/s is demonstrated by green and blue arrows. (Site number 02030000, Hardware River at Route 637 below Briery Creek [Run], near Scottsville, Virginia).

## Model Development

MLLR models were developed to describe the chance of a particular average streamflow exceeding or not exceeding an identified drought streamflow threshold. A total of 70,120 MLLR models were developed describing a range of streamflow probabilities for each of 339 streamgaging stations. Of these, 55,428 models had very low p-values, less than or equal to a p-value of 0.05, and were identified as candidates for publication. Of these, 46,704 models were based on streamflow data spanning a period of record of 10 years or more. These 46,704 models for 259 basins were selected for publication in this report. Specifically, the models describe the chance of a particular average streamflow exceeding or not exceeding an identified streamflow threshold, as a function of streamflows from an earlier time period. The models answer the question: “What is the chance (likelihood) that a particular flow value will exceed a certain stream-flow threshold as a function of measured streamflows from an earlier time period?”

These logistic regression models fit nominal Y responses to a linear model of X terms. More specifically, the models fit probabilities for two response levels using a logistic function, as described in the previous section “Maximum Likelihood Logistic Regression.”

The fitting principal of maximum likelihood means that the variable coefficients ( $\beta$ s) are chosen to maximize the joint probability attributed by the model to the responses that did occur. This fitting principal essentially minimizes the negative log-likelihood (–LogLikelihood), as attributed by the model. Each model is fitted iteratively with negative log-likelihood values converging to the final estimates (SAS Institute, 2012). (See the section “Maximum Likelihood Logistic Regression.”)

The probability of exceeding nominal monthly and daily streamflow threshold response values for July, August, and September is predicted using November, December, January, and February mean daily streamflow values as explanatory variables. These explanatory variables are used as surrogates for precipitation.



## 8 Methods for Estimating Drought Streamflow Probabilities for Virginia Streams

The MLLR models describe the MLLR probability of “Yes” (P[Yes]), as shown in equation 2, and the maximum likelihood logistic regression probability of “No” (P[No]), as shown in equation 3. Drought streamflow threshold values for models of drought streamflow probabilities were identified using guidance from the Virginia DEQ and indicator streamflow values provided by the Virginia Drought Monitoring Task Force (<http://www.deq.virginia.gov/Programs/Water/WaterSupply/WaterQuantity/Drought/DroughtIndicators.aspx>, accessed May 18, 2012).

The following drought streamflow threshold categories were used to test streamflow probabilities in MLLR models:

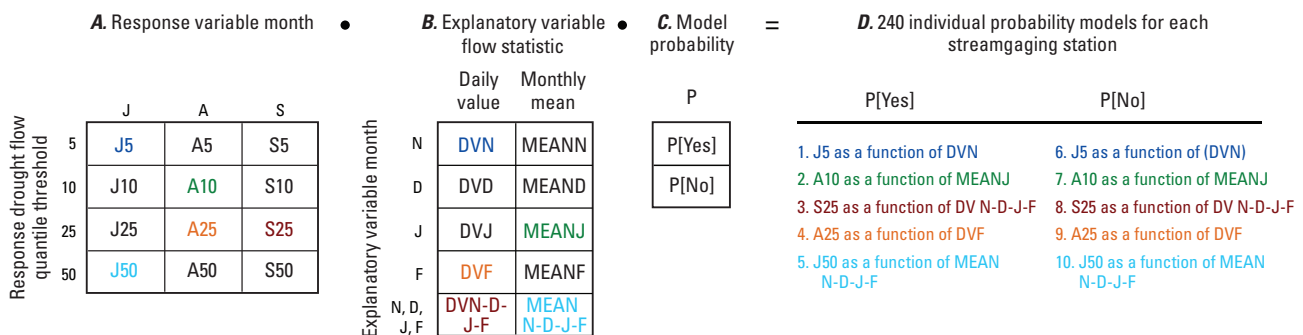
- the 5-percent streamflow quantile for the month, over the period of record;
- the 10-percent streamflow quantile for the month, over the period of record;
- the 25-percent streamflow quantile for the month, over the period of record; and
- the 50-percent streamflow quantile for the month, over the period of record.

The questions addressed using these drought flow threshold categories are

- What is the chance that streamflow exceeds the monthly 5-percent streamflow quantile?
- What is the chance that streamflow exceeds the monthly 10-percent streamflow quantile?
- What is the chance that streamflow exceeds the monthly 25-percent streamflow quantile?
- What is the chance that streamflow exceeds the monthly 50-percent streamflow quantile?

Data were prepared and models developed to perform analyses. As many as 240 individual MLLR probability models were developed for each streamgaging station, represented using 120 probability plots each illustrating 2 probability models, (P[Yes] and P[No]). The group of MLLR models developed and tested for each streamgaging station may be visualized using three matrices. For each station a 12-element matrix may be identified containing response (y) variable choices (fig. 5A), and a 10-element matrix may be identified containing explanatory (x) variable choices (fig. 5B). A third matrix may also be identified containing the two potential model probability outcomes (P[Yes] and P[No]) (fig. 5C). Multiplying the 12 elements in the first matrix, the 10 elements in the second matrix, and the 2 elements in the third matrix yields 240 individual MLLR probability models for testing and evaluation per streamgaging station, a potential total of 81,360 MLLR probability models for the 339 streamgaging stations. The 70,120 MLLR models were developed rather than 81,360 because data for some explanatory variables were unavailable in some instances.

For each station, the response (y) variable choices combine three response variable months (July [J], August [A], and September [S]) with four distinct streamflow thresholds (the 5-, 10-, 25-, and 50-percent streamflow quantiles). The explanatory (x) variable choices combine two explanatory variable flow statistics, mean monthly streamflow (Flow), and daily value flow (DV) with four distinct explanatory variable months plus the average of these four as a fifth explanatory variable (November [N] mean streamflow, December [D] mean streamflow, January [J] mean streamflow, February [F] mean streamflow, and the average N, D, J, F mean streamflow). Individual probability models combine selected response and explanatory variables (fig. 5). Ten examples of individual probability model types are listed in figure 5.



**Figure 5.** A diagram identifying MLLR probability models developed for evaluation and testing. For each streamgaging station, up to 240 individual MLLR probability models were developed for a combination of 3 potential response variable months (matrix A column labels), and 4 potential response drought streamflow quantile threshold categories (matrix A row labels), from 5 potential explanatory variable months (matrix B row labels), 2 potential explanatory variable streamflow statistics (matrix B column labels), and 2 potential model probability outcomes (matrix C). Colors show the progression of several example variable selections listed at the right of the figure.



Models were prepared using iterative methods in which the MLLR model fit converged to minimize the negative log-likelihood of each probable outcome, P[Yes] and P[No], over the range of explanatory variable (x) values. Minimizing negative log-likelihood is equivalent to maximizing log-likelihood, yielding MLLR probabilities. Model results were compiled in a table organized by streamgaging station number, listing model parameters and equations, and in graphs, showing a logistic regression response in probability space of probable outcomes over the range of a selected explanatory variable. These tabulated and plotted results were then evaluated.

## Model Verification and Diagnostics

Visual inspections of MLLR model response curves and evaluations of p-values associated with each MLLR model were used for model verification and diagnostics. Since a MLLR model describes a chance, or likelihood, of either of two possible outcomes, P[Yes] or P[No], relative to an explanatory variable (x-value), an error cannot be associated with the fitted model in the manner customarily used with regression models. An error term is not associated with a MLLR model because the predicted value (y-value) is not a response level as usually expected in a linear regression model, but rather a probability distribution of a response level. Thus, these MLLR models are correct even when the less likely of the two possible outcomes, P[Yes] or P[No], occurs (Sall and others, 2007).

A probability value (p-value) is associated with each MLLR model, indicating the strength of each probability distribution relative to the likelihood that the distribution could simply be a product of chance. Each p-value is determined by a Chi-Square test for the hypothesis that all regression parameters are zero. The Chi-Square is computed by calculating twice the difference in negative log-likelihoods ( $-2 \log L$ ) between the fitted MLLR model and a reduced version of the fitted MLLR model that has no explanatory (x) variable parameters, only intercepts, and then the resulting p-value is evaluated. The resulting p-value gives the probability of obtaining a greater Chi-Square value by chance alone, if the fit of the specified model is no better than the fit of the reduced model that includes only intercepts. Low p-values indicate that there is very little likelihood that the probability distribution associated with the MLLR model could be produced by chance alone. Thus, low p-values indicate it is highly likely that the fitted curve explaining the probabilities expressed in the MLLR model is an accurate response to the explanatory variable in the model. Model equations were ranked by the probability of obtaining a greater Chi-Square value by chance alone ( $\text{Prob} > \text{ChiSq}$ , or p-value). Smaller p-values were considered better than larger p-values. Only models with very low p-values, those less than or equal to a p-value of 0.05, were identified as candidates for publication. Of these 55,428 models, only those based on streamflow data spanning a period of record of 10 years or more (46,704 models), were selected for

publication in this report. This set of 46,704 models may be considered highly significant, with a high likelihood that the fitted model curve explains the probabilities expressed in the model and is an accurate response to the explanatory variable.

Visual inspection of each MLLR model response curve was used to identify the strength of the x–y relation and the predictive power of the x–y relation within each MLLR model. The strength of the x–y relation in each MLLR model was identified by comparing the shape of the response curve plotted for each model with the three categories of curve shape identified in figure 2. Each curve shape shown in figure 2 illustrates a distinct degree of relation strength. Note that relation strength is not goodness of fit because probabilities are predicted and all equation fits are considered “good” at predicting the likelihood of each response outcome within the given probability distribution. Predictions of P[Yes] or P[No] response relative to an explanatory (x-axis) variable may be more definitive, however, and therefore potentially more useful for decision making as a consequence of the shape of the MLLR response curve. When the fit strength of an x–y relation is weak, the absolute value of the slope parameter in the MLLR model is relatively small, producing a gentle slope of the MLLR line over the range of the explanatory data. When the fit strength of the x–y relation is strong, the absolute value of the slope parameter in the MLLR model is relatively large, producing a steeper slope of the MLLR line over the range of the explanatory data. When the fit strength of the x–y relation is approaching a “perfectly definitive fit,” the absolute value of the slope parameter in the MLLR model becomes quite large, approaching infinity as a segment of the curve approaches an orientation perpendicular to the x-axis of the graph, that is almost vertical. (See figure 2 and the section “Maximum Likelihood Logistic Regression” for additional illustration and discussion of the strength of the x–y relation.)

The predictive power of the x–y relation in each MLLR model was identified by comparing the position of the response curve plotted in probability space for each model with the three categories of curve position identified in figure 3. A curve that moves through all three zones of predictive power identified in figure 3 exhibits strong predictive power for probabilities P[Yes] and P[No] and strong explanatory power associated with the MLLR explanatory variable. A curve confined to the upper one-third or the lower one-third of an MLLR plot as identified in figure 3 also exhibits strong predictive power for probabilities P[Yes] and P[No] but weak explanatory power associated with the MLLR explanatory variable. A curve confined to the middle one-third of an MLLR plot exhibits weak predictive power for probabilities P[Yes] and P[No] and weak explanatory power associated with the MLLR explanatory variable. (See figure 3 and the section “Maximum Likelihood Logistic Regression” for additional illustration and discussion of the predictive power of the x–y relation.)

## Drought Streamflow Probability Estimates

A total of 70,120 MLLR models were developed, describing 70,120 drought streamflow probability estimates for 339 streamgaging stations. Of these, 55,428 models had an accompanying p-value statistic of 0.05 or less, indicating a highly statistically significant model of drought streamflow threshold probabilities. Of these, 46,704 models were based on streamflow data spanning a period of record of 10 years or more. These 46,704 highly significant models based on at least 10 years of streamflow data were selected for publication (table 1). Each of these models describes the chance of a particular daily-value streamflow or mean-monthly streamflow exceeding or not exceeding a 5-, 10-, 25-, or 50-percent streamflow quantile threshold for July, August, or September, as a function of mean streamflow from the previous November, December, January, or February, or the average of the previous November-December-January-February mean streamflows. The models were prepared using iterative methods in which the MLLR model fit converged to minimize the negative log-likelihood of each probable outcome,  $P[\text{Yes}]$  and  $P[\text{No}]$ , over the range of explanatory variable ( $x$ ) values. Because minimizing negative log-likelihood is equivalent to maximizing log-likelihood, this yields MLLR probabilities.

Drought streamflow probability estimates are compiled in a table organized by streamgaging station number, listing model parameters and equations, and in graphs, for each model showing a logistic regression response in probability space of probable outcomes over the range of an explanatory variable. These tabulated drought streamflow probability model parameters and equations are listed in table 1. A list of the minimum and maximum values of each explanatory ( $x$ ) variable, identifying the range of  $x$ -variable values over which drought streamflow threshold probabilities are estimated for Virginia streams, is presented in table 2.

## Selection of an Equation

As several models are available for each response variable ( $y$ ) at each streamgaging station for predicting the chance of July, August, or September streamflow exceeding or not exceeding a particular drought streamflow threshold as a function of November, December, January, or February mean daily-flow values, questions may arise as to which model is best or which equation to choose. An analyst potentially interested in using the models may wonder “How do I choose the best equation?” In this section, guidance is provided for selecting an appropriate equation.

The choice of most appropriate equation is often indicated by a mix of at least five factors: (1) the availability of an explanatory ( $x$ ) variable value to use in the equation; (2) the

time-frame within which the prediction will be useful; (3) the degree of accuracy and reliability associated with the selected explanatory variable measurement; (4) the strength of the  $x$ - $y$  relation in the MLLR accompanying each equation choice, as described in the section “Maximum Likelihood Logistic Regression;” and (5) the predictive power of the MLLR curve corresponding to each equation choice, also described in the section “Maximum Likelihood Logistic Regression.”

Reliable and accessible explanatory ( $x$ ) variable values are essential and may narrow the field of equation choices. If reliable explanatory variable values are available for only one particular month of the year, say for example February, then equations that use only February as their explanatory ( $x$ ) variable become the viable alternatives. Potential explanatory variable data need to be carefully evaluated, ensuring that measurements have been made accurately and using accepted methods.

The time frame within which the estimate will be useful may dictate viable equation options. For a given management decision, an analyst may need to predict drought streamflow probabilities within a particular interval of time. For example, if management policies direct adjustments to July streamflow withdrawals in order to avoid July drought streamflow conditions and at least 5 months are required to implement the adjustments, then equations that use explanatory variable values for March are best avoided because these values, although potentially accurate, provide predictions 4 months in advance of July streamflows when a minimum 5-month window is needed to successfully make the streamflow adjustments dictated by management policy. In contrast, if the time window for successful implementation of a particular drought management policy is relatively short, perhaps 2 months rather than 5 months, then the decision calculation may change. For example, if one set of equations using February values as the explanatory variable is available and a second set of equations using March values as the explanatory variable is available, and both equation sets have comparable  $x$ - $y$  relation strength and predictive power, then equations that use March values as the explanatory variable may be preferred as these provide predicted drought streamflow probabilities using the most recently acquired data as explanatory variable values, while still providing a time-window sufficient to successfully implement the streamflow adjustment management policy.

The degree of accuracy and reliability of an explanatory variable, relative to other explanatory variable choices may point to an appropriate equation choice. A certain explanatory variable streamflow statistic may be more accurate or reliable than another. For example, if a streamgaging station has few daily value data available for November and many data accurately measured and collected during January or February, then streamflow values from January or February may prove to be more accurate and reliable explanatory variables than those for November. Graphs of each MLLR model, plotting

the fitted MLLR line, can be compared to identify explanatory variables determined using more complete data coverage.

The strength of the x-y relation in the MLLR associated with a potential equation contributes to the usefulness of the MLLR equation as an estimator. As described in the section “Maximum Likelihood Logistic Regression,” the strength of the x-y relation in each MLLR model may be identified by comparing the shape of the response curve plotted for each model with the three categories of curve shape shown in figure 2. Each curve shape shown in figure 2 illustrates a distinct degree of x-y relation strength. Predictions of P[Yes] or P[No] response relative to an explanatory variable may be more definitive or less definitive and, therefore, potentially more or less useful for management decision making as a consequence of the shape of the MLLR response curve. When the fit strength of an x-y relation is weak, the MLLR line has a relatively gentle slope over the range of the explanatory data, as shown in figure 2A. When the fit strength of the x-y relation is strong, the MLLR line has a varying yet consistently relatively steep slope over the range of the explanatory data, expressing a more pronounced sigmoidal (s-shaped) curve, as shown in figure 2B. When the fit strength of the x-y relation is very strong, approaching a perfectly definitive fit, the absolute value of the slope of the MLLR line becomes quite large, often approaching infinity in a segment of the curve with a near-perpendicular orientation relative to the x-axis of the graph (almost vertical), as shown in figure 2C. Equations exhibiting highly definitive curve shapes are best chosen over equations exhibiting weak curve shapes.

The predictive power of the MLLR curve, corresponding to each equation choice can also help determine usefulness as a predictor. As described in the section “Maximum Likelihood Logistic Regression,” the predictive power of each MLLR model may be identified by comparing the position of the response curve plotted in probability space for each model with the three categories of curve position identified in figure 3. Equations with accompanying model graphs displaying MLLR curves that move through all three zones of predictive power (fig. 3) exhibit strong predictive power with strong and highly definitive curve shapes. Equations and curves of this type are preferred over equations and curves exhibiting weak predictive power and weak curve shape. In practice, a combination of the three zones of predictive power plus curve shape and the relative importance of each, as determined by the analyst, will identify the most appropriate equation for useful predictions of drought streamflow probability in each particular circumstance.

## Application of the Equations with Examples

MLLR equations are provided that predict the chance of exceeding or not exceeding one of a set of identified drought streamflow thresholds during the historically dry months of the year—July, August, and September—as a function of mean streamflow values from the previous historically wet months of the year—November, December, January, and February. Because equations are provided for each of the two probability categories associated with each response variable, these calculations are easy to make and easy to interpret. For a given streamgaging station, an equation may be chosen from table 1 that addresses the probability of exceeding a streamflow threshold value of interest P[Yes] or not exceeding a streamflow threshold value of interest P[No] on the basis of a mean monthly streamflow value for one of the preceding explanatory variable months of November, December, January, or February or the averaged mean monthly streamflow of these four months.

The following four example exercises provide step-by-step guidance for how to perform these calculations using the equations supplied in this report to predict the probability of exceeding or not exceeding a streamflow threshold. Each example exercise shows how to apply equations to calculate the probability of exceeding a drought streamflow threshold, P[Yes], and how to apply equations to calculate the probability of not exceeding a drought streamflow threshold, P[No]. The example exercises describe these calculations for each of the four response drought streamflow quantile thresholds identified in figure 5A and 4 of the 5 explanatory variable months identified in figure 5B.

### Example 1. Predicting the probability of exceeding a 5-percent July drought streamflow quantile threshold as a function of mean monthly streamflow from the previous December

- |           |  |
|-----------|--|
| Problem:  | At the start of the new year, you need to predict July drought streamflow probabilities for Stave Run, a stream in the Potomac River Basin near Reston, Virginia, in order to manage water withdrawals and avoid severe low flows. |
| Question: | What is the chance that July streamflow will exceed the monthly 5-percent streamflow quantile if the average December daily streamflow is 0.02 ft <sup>3</sup> /s?   |
| Given:    | Mean monthly streamflow from the previous December.  |



## 12 Methods for Estimating Drought Streamflow Probabilities for Virginia Streams

**Solution:** A streamgage identified as USGS 01644291 is on this stream below Parking Lot Road, near Reston, Virginia.

**Step 1:** In table 1, find station 01644291.

**Step 2:** Find the streamflow threshold criteria listed in table 1 that corresponds to this station number and the information given. In this case the flow threshold criteria in table 1 is listed as DV > Quantile 5% July DV POR By December Mean DV.

**Step 3:** Read Equation: Probability of “Yes” from table 1 corresponding to this station number and flow threshold criteria. In this instance, it is

$$P[\text{Yes}] = 1/(1 + e^{(0.46159633 + -11.06522 \cdot X)})$$

**Step 4:** Read Equation: Probability of “No” from table 1 corresponding to this station number and flow threshold criteria. In this instance, it is

$$P[\text{No}] = 1/(1 + e^{-(0.46159633 + -11.06522 \cdot X)})$$

**Step 5:** Calculate the probability P[Yes] that July streamflow will exceed the monthly 5-percent streamflow quantile if the average December daily streamflow is 0.02 ft<sup>3</sup>/s by substituting 0.02 for the X variable in the equation and then simplifying as follows:

$$\begin{aligned} P[\text{Yes}]_{0.02} &= 1/(1 + e^{(0.46159633 + -11.06522 \cdot X)}) \\ P[\text{Yes}]_{0.02} &= 1/(1 + e^{(0.46159633 + -11.06522 \cdot 0.02)}) \\ P[\text{Yes}]_{0.02} &= 1/(1 + e^{(0.46159633 + -0.22130)}) \\ P[\text{Yes}]_{0.02} &= 1/(1 + e^{0.24029}) \\ P[\text{Yes}]_{0.02} &= 1/(1 + 1.27162) \\ P[\text{Yes}]_{0.02} &= 1/2.27162 \\ P[\text{Yes}]_{0.02} &= 0.44 \end{aligned}$$

**Step 6:** Calculate the probability P[No] that July streamflow will not exceed the monthly 5-percent streamflow quantile if the average December daily streamflow is 0.02 ft<sup>3</sup>/s by substituting 0.02 for the X variable in the equation and then simplifying as follows:

$$\begin{aligned} P[\text{No}]_{0.02} &= 1/(1 + e^{-(0.46159633 + -11.06522 \cdot X)}) \\ P[\text{No}]_{0.02} &= 1/(1 + e^{-(0.46159633 + -11.06522 \cdot 0.02)}) \\ P[\text{No}]_{0.02} &= 1/(1 + e^{-(0.46159633 + -0.22130)}) \\ P[\text{No}]_{0.02} &= 1/(1 + e^{-0.24029}) \\ P[\text{No}]_{0.02} &= 1/(1 + 0.78640) \\ P[\text{No}]_{0.02} &= 1/1.78640 \\ P[\text{No}]_{0.02} &= 0.56 \end{aligned}$$

**Answer:** P[Yes]: There is a 0.44 (44-percent) chance that July streamflow will exceed the monthly 5-percent streamflow quantile if average

December daily streamflow is 0.02 ft<sup>3</sup>/s.

P[No]: There is a 0.56 (56-percent) chance that July streamflow will not exceed the monthly 5-percent streamflow quantile if average December daily streamflow is 0.02 ft<sup>3</sup>/s.

A plot of the MLLR line for the Potomac River Basin site number 01644291 confirms these values as shown in figure 6.

### Example 2. Predicting the probability of exceeding a 10-percent August drought streamflow quantile threshold as a function of mean monthly streamflow from the previous January

**Problem:** You need to predict August drought streamflow probabilities for the South Fork of the Roanoke River in the Roanoke River Basin near Shawsville, Virginia, in order to determine the feasibility of potential large-scale water withdrawals requested by a new business.

**Question:** What is the chance that August streamflow will exceed the monthly 10-percent streamflow quantile if the average January daily streamflow is 20 ft<sup>3</sup>/s?

**Given:** Mean monthly streamflow from the previous January.

**Solution:** A streamgage identified as USGS 02053800 is on the South Fork of the Roanoke River at Route 637 near Shawsville, Virginia.

**Step 1:** In table 1, find station number 02053800.

**Step 2:** Find the streamflow threshold criteria listed in table 1 that corresponds to this station number and the information given. In this case the flow threshold criteria in table 1 is listed as “DV > Quantile 10% August DV POR By January Mean DV.”

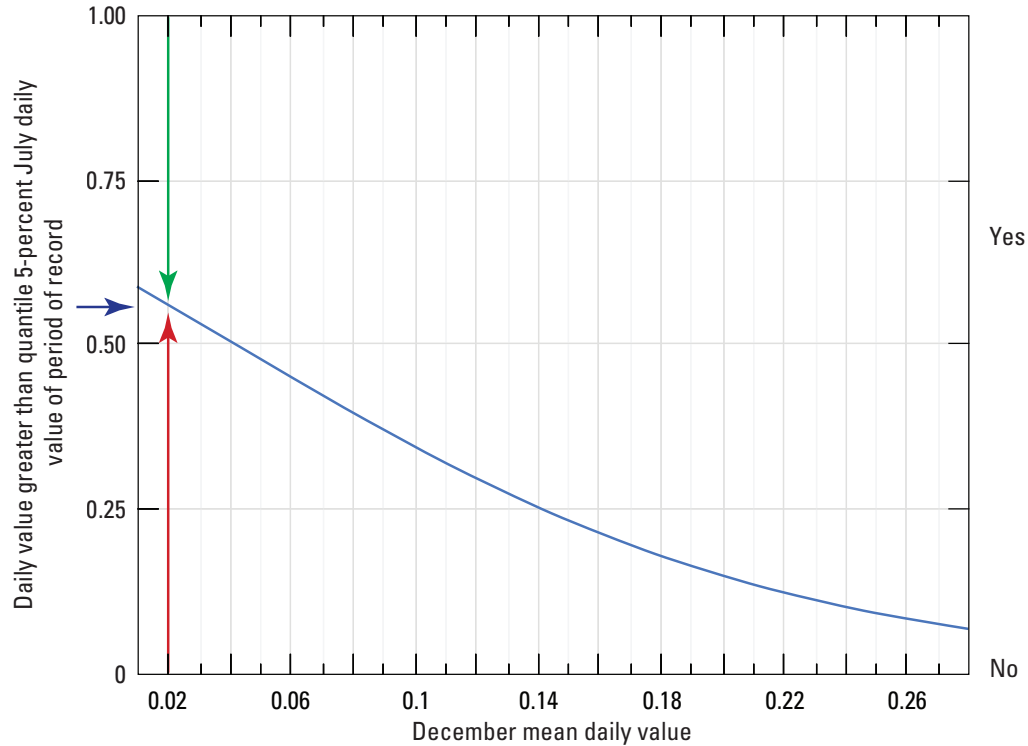
**Step 3:** Read Equation: Probability of “Yes” from table 1 corresponding to this station number and flow threshold criteria. In this instance, it is

$$P[\text{Yes}] = 1/(1 + e^{(0.09725373 + -0.0223555 \cdot X)})$$

**Step 4:** Read Equation: Probability of “No” from table 1 corresponding to this station number and flow threshold criteria. In this instance, it is

$$P[\text{No}] = 1/(1 + e^{-(0.09725373 + -0.0223555 \cdot X)})$$

**Step 5:** Calculate the probability P[Yes] that August streamflow will exceed the monthly 10-percent streamflow quantile if the average January daily streamflow is 20 ft<sup>3</sup>/s by substituting 20



**Figure 6.** A MLLR drought threshold probability plot showing the chances of exceeding, and not exceeding, an historical July 5-percent daily streamflow threshold as a function of mean daily streamflow values from December. Red and blue arrows illustrate reading a  $P[\text{No}] = 0.56$  response with an explanatory variable value of  $0.02 \text{ ft}^3/\text{s}$ . Green and blue arrows illustrate reading a  $P[\text{Yes}] = 0.44$  response with an explanatory variable value of  $0.02 \text{ ft}^3/\text{s}$ . (Site number 01644291, Stave Run Below Parking Lot Road, near Reston, Virginia.)

for the X variable in the equation and then simplifying as follows:

$$P[\text{Yes}]_{20} = 1/(1 + e^{(0.09725373 + -0.0223555 \cdot X)})$$

$$P[\text{Yes}]_{20} = 1/(1 + e^{(0.09725373 + -0.0223555 \cdot 20)})$$

$$P[\text{Yes}]_{20} = 1/(1 + e^{(0.09725373 + -0.44711)})$$

$$P[\text{Yes}]_{20} = 1/(1 + e^{-0.34986})$$

$$P[\text{Yes}]_{20} = 1/(1 + 0.70479)$$

$$P[\text{Yes}]_{20} = 1/1.70479$$

$$P[\text{Yes}]_{20} = 0.59$$

Step 6:

Calculate the probability  $P[\text{No}]$  that August streamflow will not exceed the monthly 10-percent streamflow quantile if the average January daily streamflow is  $20 \text{ ft}^3/\text{s}$  by substituting 20 for the X variable in the equation and then simplifying as follows:

$$P[\text{No}]_{20} = 1/(1 + e^{-(0.09725373 + -0.0223555 \cdot X)})$$

$$P[\text{No}]_{20} = 1/(1 + e^{-(0.09725373 + -0.0223555 \cdot 20)})$$

$$P[\text{No}]_{20} = 1/(1 + e^{-(0.09725373 + -0.44711)})$$

$$P[\text{No}]_{20} = 1/(1 + e^{0.34986}) \quad P[\text{No}]_{20} = 1/(1 + 1.41886)$$

$$P[\text{No}]_{20} = 1/2.41886$$

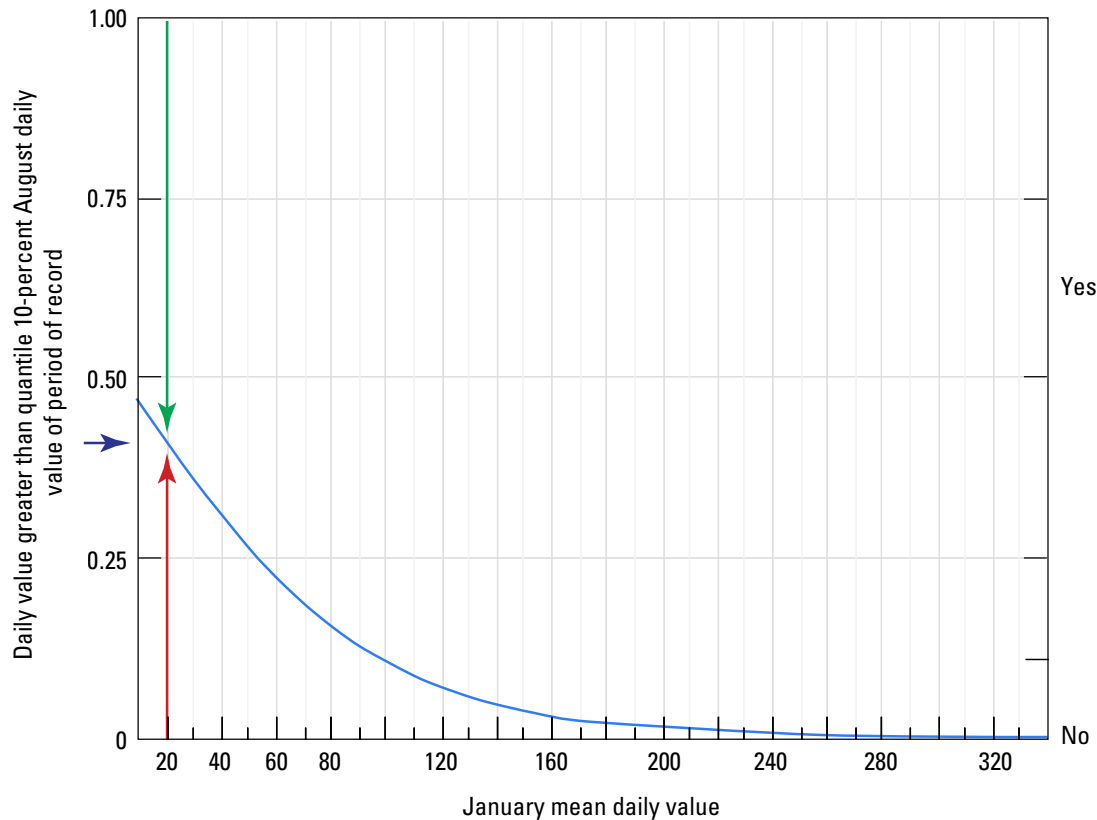
$$P[\text{No}]_{20} = 0.41$$

Answer:

$P[\text{Yes}]$ : There is a 0.59 (59-percent) chance that August streamflow will exceed the monthly 10-percent streamflow quantile if average January daily streamflow is  $20 \text{ ft}^3/\text{s}$ .

$P[\text{No}]$ : There is a 0.41 (41-percent) chance that August streamflow will not exceed the monthly 10-percent streamflow quantile if average January daily streamflow is  $20 \text{ ft}^3/\text{s}$ .

A plot of the MLLR line for the Roanoke River Basin site number 02053800 confirms these values, as shown in figure 7.



**Figure 7.** A MLLR drought threshold probability plot showing the chances of exceeding, and not exceeding, an historical August 10-percent daily streamflow threshold as a function of mean daily streamflow values from January. Red and blue arrows illustrate reading a  $P[\text{No}] = 0.41$  response with an explanatory variable value of 20  $\text{ft}^3/\text{s}$ . Green and blue arrows illustrate reading a  $P[\text{Yes}] = 0.59$  response with an explanatory variable value of 20  $\text{ft}^3/\text{s}$ . (Site number 02053800, South Fork Roanoke River at Route 637 near Shawsville, Virginia.)

### Example 3. Predicting the probability of exceeding a 25-percent September drought streamflow quantile threshold as a function of mean monthly streamflow from the previous February

**Problem:** You have noticed over the past 5 years that intervals of drought during the month of September have become more frequent and of longer duration than they seemed to be previously. You have good streamflow data from February and need to predict September drought streamflow probabilities for the North Mayo River in the Roanoke River Basin near Spencer, Virginia, in order to anticipate conditions that may warrant implementing voluntary water-use reduction programs or mandatory water-rationing policies during occasions of severe drought.

**Question:** What is the change in the probability that September streamflow will exceed the monthly 25-percent streamflow quantile if the average February daily streamflow drops from 200  $\text{ft}^3/\text{s}$  to 40  $\text{ft}^3/\text{s}$ ?

**Given:** Mean monthly streamflow from the previous February.

**Solution:** A streamgage identified as USGS 02070000 is on the North Mayo River below Route 629, near Spencer, Virginia.

**Step 1:** In table 1, find station number 02070000.

**Step 2:** Find the streamflow threshold criteria listed in table 1 that corresponds to this station number and the information given. In this case the flow threshold criteria in table 1 is listed as  $\text{DV} > \text{Quantile } 25\% \text{ September DV POR By February Mean DV}$ .

Step 3: Read Equation: Probability of “Yes” from table 1 corresponding to this station number and flow threshold criteria. In this instance, it is

$$P[\text{Yes}] = 1/(1 + e^{(1.4162127 + -0.0184251 \cdot X)}).$$

Step 4: Read Equation: Probability of “No” from table 1 corresponding to this station number and flow threshold criteria. In this instance, it is

$$P[\text{No}] = 1/(1 + e^{-(1.4162127 + -0.0184251 \cdot X)}).$$

Step 5: Calculate the change in probability P[Yes] that September streamflow will exceed the monthly 25-percent streamflow quantile if the average February daily streamflow drops from 200 ft<sup>3</sup>/s to 40 ft<sup>3</sup>/s by first substituting 200 for the X variable in the equation, then simplify the equation. Then, for comparison, solve the equation again this time substituting 40 for the X variable in the equation, and simplifying as follows:

Solve the first P[Yes] equation:

$$\begin{aligned} P[\text{Yes}]_{200} &= 1/(1 + e^{(1.4162127 + -0.0184251 \cdot X)}) \\ P[\text{Yes}]_{200} &= 1/(1 + e^{(1.4162127 + -0.0184251 \cdot 200)}) \\ P[\text{Yes}]_{200} &= 1/(1 + e^{(1.4162127 + -3.68502)}) \\ P[\text{Yes}]_{200} &= 1/(1 + e^{-2.26881}) \\ P[\text{Yes}]_{200} &= 1/(1 + 0.10344) \\ P[\text{Yes}]_{200} &= 1/1.10344 \\ P[\text{Yes}]_{200} &= 0.91 \end{aligned}$$

Solve the comparative P[Yes] equation:

$$\begin{aligned} P[\text{Yes}]_{40} &= 1/(1 + e^{(1.4162127 + -0.0184251 \cdot X)}) \\ P[\text{Yes}]_{40} &= 1/(1 + e^{(1.4162127 + -0.0184251 \cdot 40)}) \\ P[\text{Yes}]_{40} &= 1/(1 + e^{(1.4162127 + -0.73700)}) \\ P[\text{Yes}]_{40} &= 1/(1 + e^{0.67921}) \\ P[\text{Yes}]_{40} &= 1/(1 + 1.97232) \\ P[\text{Yes}]_{40} &= 1/2.97232 \\ P[\text{Yes}]_{40} &= 0.34 \end{aligned}$$

Step 6: Calculate the change in probability P[No] that September streamflow will not exceed the monthly 25-percent streamflow quantile if the daily streamflow drops from 200 ft<sup>3</sup>/s to 40 ft<sup>3</sup>/s by substituting 200 for the X variable in the equation, then simplify the equation. For comparison, solve the equation again, this time substituting 40 for the X variable in the equation, and simplify as follows:

Solve the first P[No] equation:

$$\begin{aligned} P[\text{No}]_{200} &= 1/(1 + e^{-(1.4162127 + -0.0184251 \cdot X)}) \\ P[\text{No}]_{200} &= 1/(1 + e^{-(1.4162127 + -0.0184251 \cdot 200)}) \\ P[\text{No}]_{200} &= 1/(1 + e^{-(1.4162127 + -3.68502)}) \\ P[\text{No}]_{200} &= 1/(1 + e^{2.26881}) \\ P[\text{No}]_{200} &= 1/(1 + 9.66786) \\ P[\text{No}]_{200} &= 1/10.66786 \\ P[\text{No}]_{200} &= 0.09 \end{aligned}$$

Solve the comparative P[No] equation:

$$\begin{aligned} P[\text{No}]_{40} &= 1/(1 + e^{-(1.4162127 + -0.0184251 \cdot X)}) \\ P[\text{No}]_{40} &= 1/(1 + e^{-(1.4162127 + -0.0184251 \cdot 40)}) \\ P[\text{No}]_{40} &= 1/(1 + e^{-(1.4162127 + -0.73700)}) \\ P[\text{No}]_{40} &= 1/(1 + e^{-0.67921}) \\ P[\text{No}]_{40} &= 1/(1 + 0.50702) \\ P[\text{No}]_{40} &= 1/1.50702 \\ P[\text{No}]_{40} &= 0.66 \end{aligned}$$

Answer:

P[Yes]<sub>200</sub>: There is a 0.91 (91-percent) chance that September streamflow will exceed the monthly 25-percent streamflow quantile if average February daily streamflow is 200 ft<sup>3</sup>/s.

P[No]<sub>200</sub>: There is a 0.09 (9-percent) chance that September streamflow will not exceed the monthly 25-percent streamflow quantile if average February daily streamflow is 200 ft<sup>3</sup>/s.

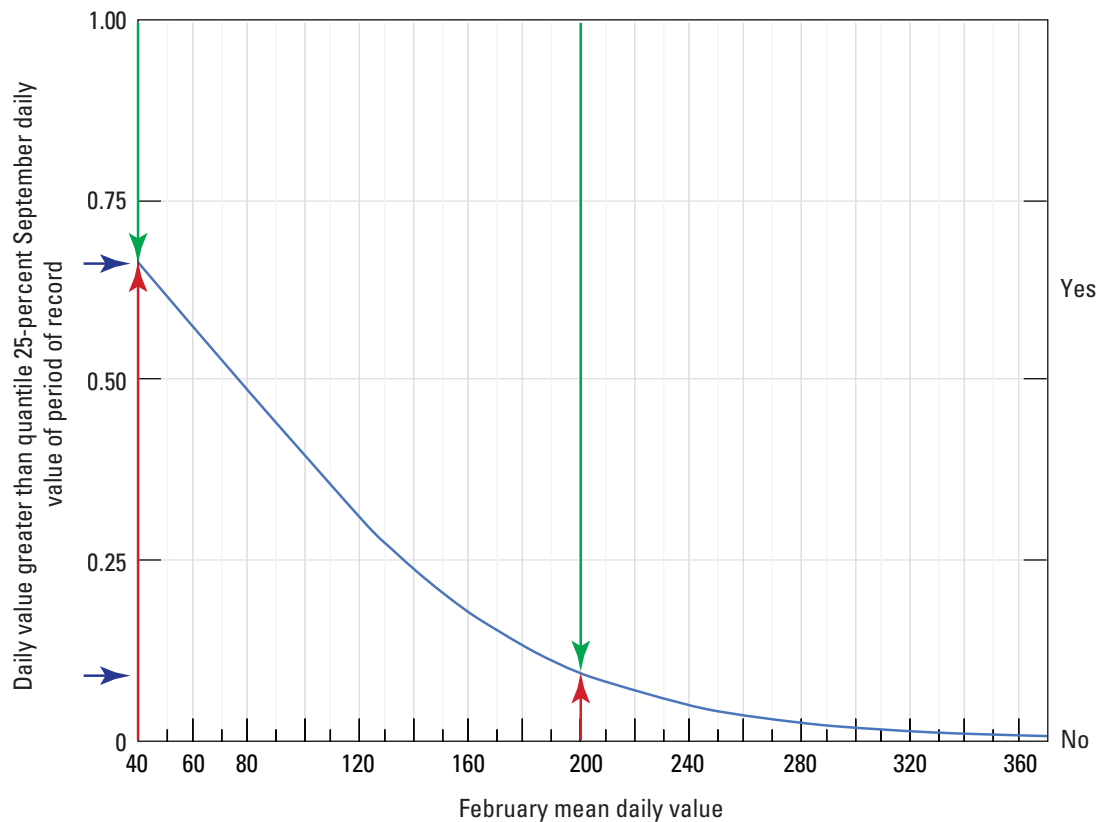
P[Yes]<sub>40</sub>: There is a 0.34 (34-percent) chance that September streamflow will exceed the monthly 25-percent streamflow quantile if average February daily streamflow is 40 ft<sup>3</sup>/s.

P[No]<sub>40</sub>: There is a 0.66 (66-percent) chance that September streamflow will not exceed the monthly 25-percent streamflow quantile if average February daily streamflow is 40 ft<sup>3</sup>/s.

The probability that September streamflow will exceed the monthly 25-percent streamflow quantile if the average February daily streamflow drops from 200 ft<sup>3</sup>/s to 40 ft<sup>3</sup>/s is reduced from P[Yes]<sub>200</sub> = 0.91 to P[Yes]<sub>40</sub> = 0.34, a decrease of 0.57 (57 percentage points).

The probability that September streamflow will not exceed the monthly 25-percent streamflow quantile if the average February daily streamflow drops from 200 ft<sup>3</sup>/s to 40 ft<sup>3</sup>/s is increased from P[No]<sub>200</sub> = 0.09 to P[No]<sub>40</sub> = 0.66, an increase of 0.57 (57 percentage points).

A plot of the MLLR line for the Roanoke River Basin site number 02070000 confirms these values, as shown in figure 8.



**Figure 8.** A MLLR drought threshold probability plot showing the chances of exceeding, and not exceeding, an historical September 25-percent daily streamflow threshold as a function of mean daily streamflow values from February. Red and blue arrows illustrate reading a  $P[No]_{200} = 0.09$  response with an explanatory variable value of 200  $ft^3/s$ , and a  $P[No]_{40} = 0.66$  response with an explanatory variable value of 40  $ft^3/s$ . Green and blue arrows illustrate reading a  $P[Yes]_{200} = 0.91$  response with an explanatory variable value of 200  $ft^3/s$ , and a  $P[Yes]_{40} = 0.34$  response with an explanatory variable value of 40  $ft^3/s$ . (Site number 02070000, North Mayo River below Route 629 near Spencer, Virginia.)

**Example 4. Predicting the probability of exceeding a 50-percent July drought streamflow quantile threshold as a function of averaged mean monthly streamflow from the previous November, December, January, and February**

**Problem:** A land-use re-classification (development) is proposed that may increase the demand for potable water. July river flow is frequently the lowest flow of the year in your district. County supervisors have scheduled a vote on the land-use re-classification in 2 weeks time. You need to predict July drought streamflow probabilities for the Robinson River in the Rappahannock River Basin in order to determine whether water flows are sufficient to sustain the water demands of the proposed development.

- Question 1:** What is the chance that July streamflow will exceed the monthly 50-percent streamflow quantile if the average of November, December, January, and February daily streamflow is maintained at 500  $ft^3/s$ ?

**Question 2:** What is the chance that July streamflow will exceed the monthly 50-percent streamflow quantile if the average of November, December, January, and February daily streamflow is reduced to 120  $ft^3/s$ ?
- Given:** The average of mean monthly streamflow from the previous November, December, January, and February.

**Solution:** A streamgage identified as USGS 01666500 is on the Robinson River at Route 614 near Locust Dale, Virginia.

**Step 1:** In table 1, find station number 01666500.



- Step 2: Find the streamflow threshold criteria listed in table 1 that corresponds to this station number and the information given. In this case, the flow threshold criteria in table 1 is listed as  $DV > \text{Quantile } 50\% \text{ July DV POR By Average N-D-J-F Mean DV}$ .
- Step 3: Read Equation: Probability of “Yes” from table 1 corresponding to this station number and flow threshold criteria. In this instance, it is  

$$P[\text{Yes}] = 1/(1 + e^{(2.12923861 + -0.0085836 \cdot X)})$$
- Step 4: Read Equation: Probability of “No” from table 1 corresponding to this station number and flow threshold criteria. In this instance, it is  

$$P[\text{No}] = 1/(1 + e^{-(2.12923861 + -0.0085836 \cdot X)})$$
- Step 5: Calculate the probability  $P[\text{Yes}]$  that July streamflow will exceed the monthly 50-percent streamflow quantile if the average November, December, January, and February daily streamflow is maintained at 500 ft<sup>3</sup>/s by substituting 500 for the X variable in the equation and then simplifying as follows:  

$$P[\text{Yes}]_{500} = 1/(1 + e^{(2.12923861 + -0.0085836 \cdot X)})$$

$$P[\text{Yes}]_{500} = 1/(1 + e^{(2.12923861 + -0.0085836 \cdot 500)})$$

$$P[\text{Yes}]_{500} = 1/(1 + e^{(2.12923861 + -4.2918)})$$

$$P[\text{Yes}]_{500} = 1/(1 + e^{2.16256})$$

$$P[\text{Yes}]_{500} = 1/(1 + 0.11503)$$

$$P[\text{Yes}]_{500} = 1/1.11503$$

$$P[\text{Yes}]_{500} = 0.90$$
- Step 6: Calculate the probability  $P[\text{No}]$  that July streamflow will not exceed the monthly 50-percent streamflow quantile if the average November, December, January, and February daily streamflow is maintained at 500 ft<sup>3</sup>/s by substituting 500 for the X variable in the equation and then simplifying as follows:  

$$P[\text{No}]_{500} = 1/(1 + e^{-(2.12923861 + -0.0085836 \cdot X)})$$

$$P[\text{No}]_{500} = 1/(1 + e^{-(2.12923861 + -0.0085836 \cdot 500)})$$

$$P[\text{No}]_{500} = 1/(1 + e^{-(2.12923861 + -4.2918)})$$

$$P[\text{No}]_{500} = 1/(1 + e^{2.16256})$$

$$P[\text{No}]_{500} = 1/(1 + 8.69338)$$

$$P[\text{No}]_{500} = 1/9.69338$$

$$P[\text{No}]_{500} = 0.10$$
- Answer:  $P[\text{Yes}]_{500}$ : There is a 0.90 (90-percent) chance that July streamflow will exceed the monthly 50-percent streamflow quantile if the average November, December, January, and February daily streamflow is maintained at 500 ft<sup>3</sup>/s.
- $P[\text{No}]_{500}$ : There is a 0.10 (10-percent) chance that July streamflow will not exceed the monthly 50-percent streamflow quantile if the average November, December, January, and February daily streamflow is maintained at 500 ft<sup>3</sup>/s.
- Step 7: Calculate the probability  $P[\text{Yes}]$  that July streamflow will exceed the monthly 50-percent streamflow quantile if the average November, December, January, and February daily streamflow is reduced to 120 ft<sup>3</sup>/s by substituting 120 for the X variable in the equation and then simplifying as follows:  

$$P[\text{Yes}]_{120} = 1/(1 + e^{(2.12923861 + -0.0085836 \cdot X)})$$

$$P[\text{Yes}]_{120} = 1/(1 + e^{(2.12923861 + -0.0085836 \cdot 120)})$$

$$P[\text{Yes}]_{120} = 1/(1 + e^{(2.12923861 + -1.030032)})$$

$$P[\text{Yes}]_{120} = 1/(1 + e^{1.09921})$$

$$P[\text{Yes}]_{120} = 1/(1 + 3.00178)$$

$$P[\text{Yes}]_{120} = 1/4.00178$$

$$P[\text{Yes}]_{120} = 0.25$$
- Step 8: Calculate the probability  $P[\text{No}]$  that July streamflow will not exceed the monthly 50-percent streamflow quantile if the average November, December, January, and February daily streamflow is reduced to 120 ft<sup>3</sup>/s by substituting 120 for the X variable in the equation and then simplifying as follows:  

$$P[\text{No}]_{120} = 1/(1 + e^{-(2.12923861 + -0.0085836 \cdot X)})$$

$$P[\text{No}]_{120} = 1/(1 + e^{-(2.12923861 + -0.0085836 \cdot 120)})$$

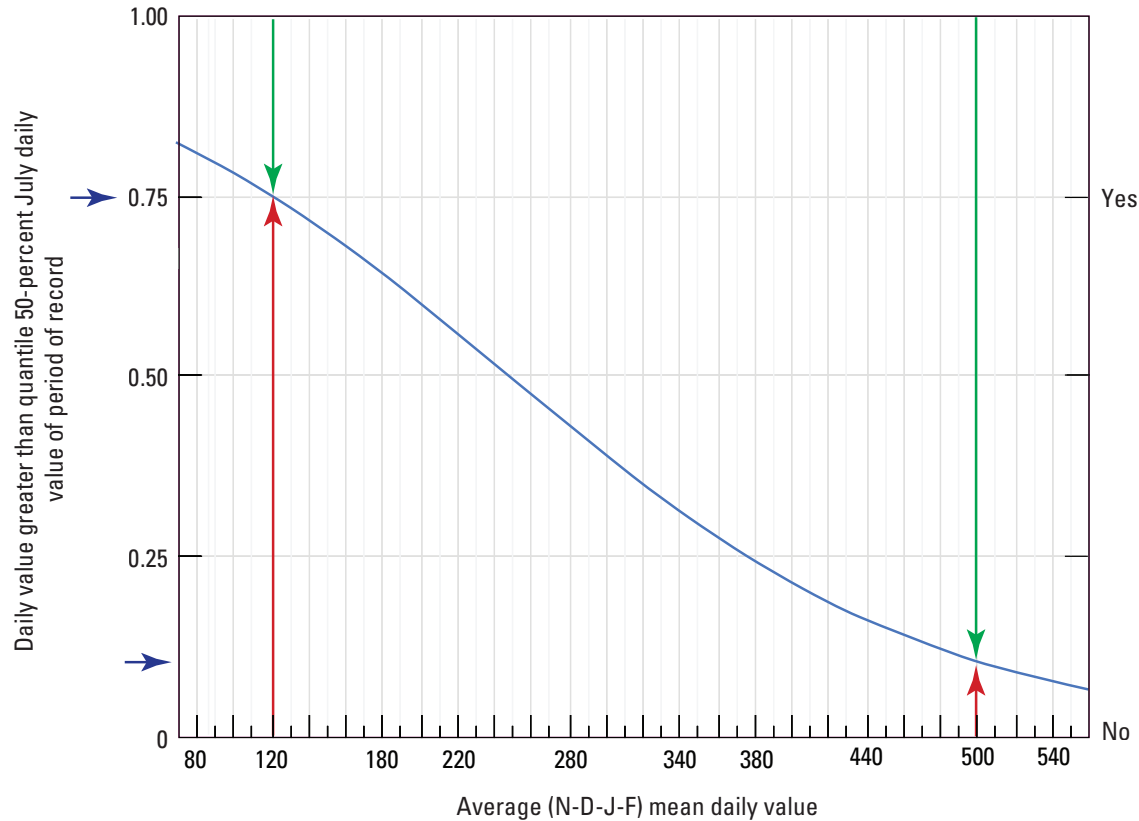
$$P[\text{No}]_{120} = 1/(1 + e^{-(2.12923861 + -1.030032)})$$

$$P[\text{No}]_{120} = 1/(1 + e^{-1.09921})$$

$$P[\text{No}]_{120} = 1/(1 + 0.33314)$$

$$P[\text{No}]_{120} = 1/1.33314$$

$$P[\text{No}]_{120} = 0.75$$
- Answer:  $P[\text{Yes}]_{120}$ : There is a 0.25 (25-percent) chance that July streamflow will exceed the monthly 50-percent streamflow quantile if the average November, December, January, and February daily streamflow is reduced to 120 ft<sup>3</sup>/s.
- $P[\text{No}]_{120}$ : There is a 0.75 (75-percent) chance that July streamflow will not exceed the monthly 50-percent streamflow quantile if the average November, December, January, and February daily streamflow is reduced to 120 ft<sup>3</sup>/s.
- A plot of the MLLR line for the Rappahannock River Basin site number 01666500 confirms these values, as shown in figure 9.



**Figure 9.** A MLLR drought threshold probability plot showing the chances of exceeding, and not exceeding, an historical July 50-percent daily streamflow threshold as a function of the average of mean daily streamflow values from November, December, January, and February (N-D-J-F). Red and blue arrows illustrate reading a  $P[\text{No}]_{500} = 0.10$  response with an explanatory variable value of 500  $\text{ft}^3/\text{s}$ , and a  $P[\text{No}]_{120} = 0.75$  response with an explanatory variable value of 120  $\text{ft}^3/\text{s}$ . Green and blue arrows illustrate reading a  $P[\text{Yes}]_{500} = 0.90$  response with an explanatory variable value of 500  $\text{ft}^3/\text{s}$ , and a  $P[\text{Yes}]_{120} = 0.25$  response with an explanatory variable value of 120  $\text{ft}^3/\text{s}$ . (Site number 01666500, Robinson (Robertson) River at Route 614 near Locust Dale, Virginia.)

## Summary

Anticipating drought conditions in Virginia streams is essential to sound management of the water resources and associated riparian and watershed ecosystems. Reliable estimates of the likelihood that stream low flows during drought-prone months will exceed specific low-flow thresholds can provide advance warning of drought conditions, allowing extended lead times for improved drought awareness and effective management response. This report, a product of a cooperative partnership between the Virginia Department of Environmental Quality (DEQ) and the U.S. Geological Survey, describes empirically determined maximum likelihood logistic regression (MLLR) equations that predict

drought streamflow probabilities for Virginia streams. Winter streamflows are used to estimate the likelihood of probable streamflows during drought-prone summer months. Probabilities of streamflows exceeding specific drought streamflow thresholds are characterized in a set of 70,120 MLLR equations for 339 basins in Virginia. A 46,704-member subset of these equations is published in this report for 259 Virginia basins, each with an accompanying p-value statistic that is less than or equal to 0.05 and is based on streamflow data spanning a minimum 10-year (maximum 112-year) period of record. More than 5.12 million streamflow daily values, collected over the period of record, were compiled and used in the analysis. Example calculations are provided demonstrating how to use the equations to estimate probable streamflows up to 8 months in advance.

## Selected References

- Aitchison, J., and Brown, J.A.C., 1957, *The lognormal distribution*: Cambridge, U.K., Cambridge University Press, 176 p.
- Allison, P.D., 2012, *Logistic regression using SAS®: Theory and Application* (2d ed.): Cary, NC, SAS Institute Inc., 339 p.
- Austin, S.H., Krstolic, J.L., and Wiegand, Ute, 2011a, Low-flow characteristics of Virginia streams: U.S. Geological Survey Scientific Investigations Report 2011–5143, 122 p. + 9 tables on CD. (Also available online at <http://pubs.usgs.gov/sir/2011/5143/>.)
- Austin, S.H., Krstolic, J.L., and Wiegand, Ute, 2011b, Peak-flow characteristics of Virginia streams: U.S. Geological Survey Scientific Investigations Report 2011–5144, 106 p. + 3 tables and 2 appendixes on CD. (Also available online at <http://pubs.usgs.gov/sir/2011/5144/>.)
- Austin, S.H., and Wiegand, U., 2009, Annual maximum stages and discharges of selected streams in Virginia through 2007: U.S. Geological Survey Open-File Report 2009–1007, 733 p. (Available only online at <http://pubs.water.usgs.gov/ofr2009-1007>.)
- Belsley, D.A., Kuh, E., and Welsch, R.E., 1980, *Regression diagnostics*: New York, John Wiley and Sons, 292 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.
- Bent, G.C., and Steeves, P.A., 2006, A revised logistic regression equation and an automated procedure for mapping the probability of a stream flowing perennially in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2006–5031, 107 p.
- Burnham, K.P., and Anderson, D.R., 2002, *Model selection and inference—A practical information-theoretic approach*: New York, Springer-Verlag, 488 p.
- Cook, R.D., and Weisberg, S., 1982, *Residuals and influence in regression*: New York, Chapman and Hall, 229 p.
- Cushing, E.M., Kantrowitz, I.H., and Taylor, K.R., 1973, *Water resources of the Delmarva Peninsula*: U.S. Geological Survey Professional Paper 822, 58 p.
- Dicken, C.L., Nicholson, S.W., Horton, J.D., Kinney, S.A., Gunther, G., Foote, M.P., and Mueller, J.A.L., 2005, Preliminary integrated geologic map databases for the United States—Delaware, Maryland, New York, Pennsylvania, and Virginia: U.S. Geological Survey Open-File Report 2005–1325, 45 p. (Also available at <http://pubs.usgs.gov/of/2005/1325/>.)
- Fenneman, N.M., 1938, *Physiography of Eastern United States*: New York, McGraw-Hill, 714 p.
- Fenneman, N.M., and Johnson, D.W., 1946, *Physiographic divisions of the conterminous U.S.*: U.S. Geological Survey map product, scale 1:7,000,000. (Also available at <http://water.usgs.gov/GIS/metadata/usgswrd/XML/physio.xml>.)
- Harlow, G.E., Jr., Orndorff, R.C., Nelms, D.L., Weary, D.J., and Moberg, R.M., 2005, Hydrogeology and ground-water availability in the carbonate aquifer system of Frederick County, Virginia: U.S. Geological Survey Scientific Investigations Report 2005–5161, 30 p.
- Hayes, D.C., 1991, Low-flow characteristics of streams in Virginia: U.S. Geological Survey Water-Supply Paper 2374, 69 p.
- Hayes, D.C., and Nelms, D.L., 2001, Base-flow characteristics of streams in the Valley and Ridge, Blue Ridge, and Piedmont Physiographic Provinces of Virginia and other mid-Atlantic States [abs.], in U.S. Geological Survey Appalachian Region integrated science workshop proceedings, Gatlinburg, Tennessee, October 22–26, 2001: U.S. Geological Survey Open-File Report 01–406, p. 87–88.
- Hayes, D.C., and Wiegand, U., 2006, Drainage areas of selected streams in Virginia: U.S. Geological Survey Open-File Report 2006–1308, 51 p.
- Helsel, D.R., and Hirsch, R.M., 2002, *Statistical methods in water resources*: U.S. Geological Survey Techniques of Water-Resources Investigations Report, book 4, chap. A3, 510 p.
- Hirsch, R.M., Slack, J.R., and Smith, R.A., 1982, Techniques of trend analysis for monthly water quality data: *Water Resources Research*, v. 18, no. 1, p. 107–121.
- Homer, C., Huang, C., Yang, L., Wylie, B., and Coan, M., 2004, Development of a 2001 National Landcover Database for the United States: *Photogrammetric Engineering and Remote Sensing*, v. 70, no. 7, July 2004, p. 829–840.
- Kirby, W., 1980, Computer routines for probability distributions, random numbers, and related functions: U.S. Geological Survey Open-File Report 80–448, 61 p.
- Lane, E.W., and Lei, K., 1950, Stream flow variability: *Transcript from American Society of Engineers*, v. 115, p. 1084–1134.
- Langbein, W.B., and Iseri, K.T., 1995, General introduction and hydrologic definitions, manual of hydrology—Part 1. General surface-water techniques: U.S. Geological Survey Water-Supply Paper 1541-A, 29 p. (Also available at <http://water.usgs.gov/wsc/glossary.html>.)

- Lynch, D.D., 1987, Hydrologic conditions and trends in Shenandoah National Park, Virginia: U.S. Geological Survey Water-Resources Investigations Report 87-4131, 115 p.
- Lynch, D.D., Nuckels, E. H., and Zenone, C., 1987, Low-flow characteristics and chemical quality of streams in the Culpeper geologic basin, Virginia and Maryland: U.S. Geological Survey Miscellaneous Investigations Map I-1313-H, scale 1:125,000.
- Mixon, R.B., Berquist, C.R., Jr., Newell, W.L., Johnson, G.H., Powars, D.S., Schindler, J.S., and Rader, E.K., 1989, Geologic map and generalized cross sections of the Coastal Plain and adjacent parts of the Piedmont, Virginia with generalized geologic cross sections: U.S. Geological Survey Miscellaneous Investigations Map I-2033, scale 1:250,000.
- Mohler, E.H., Jr., and Hagan, G.F., 1981, Low flow of streams in Fairfax County, Virginia: U.S. Geological Survey Open-File Report 81-63, 30 p.
- Mosteller, F., and Tukey, J. W., 1977, Data analysis and regression—A second course in statistics: Reading, MA, Addison-Wesley, 588 p.
- National Oceanic and Atmospheric Administration, 2007, National Weather Service Hydrometeorological Design Studies Center, precipitation frequency data server, accessed January 27, 2009, at <http://dipper.nws.noaa.gov/hdsc/pfds/>.
- Nelms, D.L., Harlow, G.E., Jr., and Hayes, D.C., 1997, Base-flow characteristics of streams in the Valley and Ridge, the Blue Ridge, and the Piedmont Physiographic Provinces of Virginia: U.S. Geological Survey Water-Supply Paper 2457, 48 p., 1 pl.
- Nelms, D.L., and Moberg, R.M., Jr., 2010a, Preliminary assessment of the hydrogeology and groundwater availability in the metamorphic and siliciclastic fractured-rock aquifer systems of Warren County, Virginia: U.S. Geological Survey Scientific Investigations Report 2010-5190, 74 p.
- Nelms, D.L., and Moberg, R.M., Jr., 2010b, Hydrogeology and groundwater availability in Clarke County, Virginia: U.S. Geological Survey Scientific Investigations Report 2010-5112, 119 p.
- Nuckels, E.H., 1970, Virginia streamflow data program analysis: U.S. Geological Survey Open-File Report, 54 p.
- Paybins, K.S., 2008, Basin characteristics for selected streamflow-gaging stations in and near West Virginia: U.S. Geological Survey Open-File Report 2008-1087, 9 p. (Also available at <http://pubs.usgs.gov/of/2008/1087/>.)
- PRISM Climate Group, Oregon State University, 2007, Mean annual precipitation from 1971 to 2000 and average monthly rainfall for January, February, and March, accessed May 5, 2008, at <http://prism.oregonstate.edu/products/matrix.phtml>.
- Ries, K.G., III, and Friesz, P.J., 2000, Methods for estimating low-flow statistics for Massachusetts streams: U.S. Geological Survey Water-Resources Investigations Report 00-4135, 81 p.
- Sall, John, Greighton, L., and Lehman, A., 2007, JMP start statistics: A guide to statistics and data analysis using JMP (4th ed.): Cary, North Carolina, SAS Institute Inc., 607 p.
- SAS Institute Inc., 2012, JMP 10 modeling and multivariate methods: Cary, NC, 703 p.
- Smith, R.W., 1981, Rock type and minimum 7-day/10-year flow in Virginia streams: Virginia Water Resources Research Center Bulletin 116, 43 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. 1421-1432.
- Stedinger, J.R., and Thomas, W.O., Jr., 1985, Low-flow frequency estimation using base-flow measurements: U.S. Geological Survey Open-File Report 85-95, 22 p.
- Tasker, G.D., 1975, Combining estimates of low-flow characteristics of streams in Massachusetts and Rhode Island: Journal of Research of the U.S. Geological Survey, v. 3, no. 1, p. 107-112.
- Tasker, G.D., 1987, A comparison of methods for estimating low flow characteristics of streams: Water Resources Bulletin, v. 23, no. 6, p. 1077-1083.
- Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: Journal of Hydrology, v. 111, p. 361-375.
- Trainer, F.W., and Watkins, F.A., Jr., 1975, Geohydrologic reconnaissance of the upper Potomac River basin: U.S. Geological Survey Water-Supply Paper 2035, 68 p.
- U.S. Geological Survey, 2012, National Water Information System (NWIS), accessed May 17, 2012, at [http://nwis.waterdata.usgs.gov/nwis/dv/?referred\\_module=sw](http://nwis.waterdata.usgs.gov/nwis/dv/?referred_module=sw).
- Wetzel, K.L., and Bettendorff, J.M., 1986, Techniques for estimating streamflow characteristics in the eastern and interior coal provinces of the United States: U.S. Geological Survey Water-Supply Paper 2276, 80 p.



For additional information regarding this publication, contact:

Director  
USGS Virginia Water Science Center  
1730 East Parham Road  
Richmond, Virginia 23228  
email: [dc\\_va@usgs.gov](mailto:dc_va@usgs.gov)

Or visit the USGS Virginia Water Science  
Center Web site at:  
<http://va.water.usgs.gov>

Prepared by:  
USGS Science Publishing Network  
Raleigh Publishing Service Center  
3916 Sunset Ridge Road  
Raleigh, NC 27607



ISBN 978-1-4113-3855-5



9 781411 338555

ISSN 2328-0328 (online)

ISSN 2328-031X (print)

<http://dx.doi.org/10.3133/sir20145145>

Printed on recycled paper