USGS
science for a changing world

In cooperation with the Michigan Department of Environmental Quality and the Michigan Department of Natural Resources

A Regression Model for Computing Index Flows Describing the Median Flow for the Summer Month of Lowest Flow in Michigan

Scientific Investigations Report 2008–5096

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

A Regression Model for Computing Index Flows Describing the Median Flow for the Summer Month of Lowest Flow in Michigan

By David A. Hamilton, Richard C. Sorrell, and David J. Holtschlag

In cooperation with the Michigan Department of Environmental Quality and the Michigan Department of Natural Resources

Scientific Investigations Report 2008–5096

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

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

U.S. Geological Survey
Mark D. Myers, Director

U.S. Geological Survey, Reston, Virginia: 2008

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS--the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Suggested citation:
Hamilton, D.A., Sorrell, R.C., and Holtschlag, D.J., 2008, A regression model for computing index flows describing the median flow for the summer month of lowest flow in Michigan: U.S. Geological Survey Scientific Investigations Report 2008–5096, 43 p.

Contents

Figures

Tables

Conversion Factors and Abbreviations

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter (cm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
	Area	
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
cubic foot per second per square mile [(ft^3/s)/mi^2]	0.01093	cubic meter per second per square kilometer [(m^3/s)/km^2]
gallon per day (gal/d)	3.785	liters per day (liters per day)
inch per year (in/yr)	2.54	centimeter per year (cm/yr)
	Transmissivity*	
foot squared per day (ft^2/d)	0.09290	meter squared per day (m^2/d)

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

°C=(°F-32)/1.8

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

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

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

*Transmissivity: The standard unit for transmissivity is cubic foot per day per square foot times foot of aquifer thickness [(ft^3/d)/ft^2]ft. In this report, the mathematically reduced form, foot squared per day (ft^2/d), is used for convenience.

List of Symbols

Symbol	Name
y	A column vector containing the hydrologic response variable
X	The "design matrix," which, in general, is composed of *p* columns of basin and climatic characteristics augmented with a leading column of ones that serve as explanatory variables to estimate the hydrologic response
β	A column vector of parameters, $\beta_0, \beta_1, \ldots, \beta_p$, that relate the explanatory variables to the hydrologic response variable
β_{ols}	The ordinary least-square estimator of β is denoted and computed as $\beta_{ols} = (X' \cdot X)^{-1} \cdot X' \cdot y$, where the prime symbol implies a matrix transpose and the –1 power implies a matrix inverse operation
ε	A vector of residuals that is assumed to be normally distributed and independent with mean zero and constant variance σ^2, commonly written $\varepsilon \sim NI(0, \sigma^2)$.
$Cov(\varepsilon, X)$	The covariance matrix between the residual vector, ε, and the explanatory variables contained in the design matrix, X
$LPL_{\alpha/2}$, $UPL_{1-\alpha/2}$	The lower prediction limit and the upper prediction limit, respectively centered about the regression estimate of the hydrologic response, $\hat{y} = x_0 \cdot \beta_{ols}$. The interval $[LPL_{\alpha/2}, UPL_{1-\alpha/2}]$ is likely to contain the true hydrologic response, y_0, with a probability of $1-\alpha$
x_0	A row vector containing the basin and climatic characteristics at a specific site, augmented with a leading one, that serves as the explanatory variables to estimate the hydrologic response at that site
α	The specified alpha level for the confidence interval. For example, if alpha was specified as 0.2, there would be a 10-percent chance that the hydrologic response would be less than $LPL_{\alpha/2}$ and a 10-percent chance that it would be greater than $UPL_{1-\alpha/2}$, providing a total probability of 20 percent that the true hydrologic response would be outside the prediction interval.
$t_{n-p-1, 1-\alpha/2}$	The ordinate from the Student's '*t*' probability distribution for a specified degrees of freedom equal to the number of observations (*n*), minus the number of estimated parameter for explanatory variables (*p*), minus one for the intercept.
s^2	The sample estimate of the population error variance σ^2
SS_T	The total sum of squares, computed as $(y-\overline{y})'(y-\overline{y})$, where $\overline{y}$ is the sample mean
SS_E	The sum of squared errors, computed as $(y-\hat{y})'(y-\hat{y}) = \varepsilon'\varepsilon$
df_E	The degrees of freedom in the error term, computed as $n-p-1$
MS_E	The mean square error, computed as $s^2 = SS_E / df_E$
RMS_E	The square root of the mean square error
R_p^2	The Pearson multiple coefficient of determination, computed as $1 - SS_E / SS_T$
R_{Adj}^2	The Pearson multiple coefficient of determination adjusted for the number of estimated parameters

Symbol	Name
r_P	The Pearson product moment correlation coefficient, which the square root of R_p^2
r_S	The Spearman correlation coefficient, which is equal to the Pearson's correlation coefficient if it were computed on the ranks of the data
R_S^2	The Spearman coefficient of determination, which is the square of the Spearman correlation coefficient
$cov(\beta_{ols})_{i,j}$	The covariance matrix among ordinary least-square parameter estimates
$cor(\beta_{ols})_{i,j}$	The correlation matrix, computed as $\mathrm{cov}(\beta_{ols})_{i,j} \big/ \sqrt{\mathrm{cov}(\beta_{ols})_{i,i} \cdot \mathrm{cov}(\beta_{ols})_{j,j}}$

A Regression Model for Computing Index Flows Describing the Median Flow for the Summer Month of Lowest Flow in Michigan

By David A. Hamilton[1], Richard C. Sorrell[1], and David J. Holtschlag[2]

Abstract

In 2006, Michigan enacted laws to prevent new large-capacity withdrawals from decreasing flows to the extent that they would functionally impair a stream's ability to support characteristic fish populations. The median streamflow for the summer month of lowest flow was specified by state decision makers as the index flow on which likely impacts of withdrawals would be assessed. At sites near long-term streamflow-gaging stations, analysis of streamflow records during July, August, and September was used to determine the index flow. At ungaged sites, an alternate method for computing the index flow was needed. This report documents the development of a method for computing index flows at ungaged stream sites in Michigan. The method is based on a regression model that computes the index water yield, which is the index flow divided by the drainage area. To develop the regression model, index flows were determined on the basis of daily flows measured during July, August, and September at 147 streamflow-gaging stations having 10 or more years of record (considered long-term stations) in Michigan. The corresponding index water yields were statistically related to climatic and basin characteristics upstream from the stations in the regression model. Climatic and basin characteristics selected as explanatory variables in the regression model include two aquifer-transmissivity and hydrologic-soil groups, forest land cover, and normal annual precipitation. Regression-model estimates of water yield explain about 70.8 percent of the variability in index water yields indicated by streamflow-gaging station records. Index flows computed on the basis of regression-model estimates of water yield and corresponding drainage areas explain about 94.0 percent of the variability in index flows indicated by streamflow-gaging station records. No regional bias was detected in the regression-based estimates of water yield within seven hydrologic subregions spanning Michigan. Thus, the single regression model developed in this report can be used to produce unbiased estimates of index water yield and flow statewide. In addition, a technique is presented for computing prediction intervals about the index flow estimates.

Introduction

The Michigan Legislature (2006) passed Public Act 33 in 2006 (PA33–2006); it and related laws are the first state laws to regulate water withdrawals. The legislation seeks to prevent any new or increased large-capacity withdrawal (generally referring to withdrawals that average more than 100,000 gallons of water per day (0.1547 ft^3/s) in any consecutive 30-day period) from causing an adverse resource impact. This impact is defined as decreasing the flow of a stream by part of the index flow such that the stream's ability to support characteristic fish populations is functionally impaired. PA33–2006 further defines index flow as the 50 percent exceedance (median) flow for the lowest flow month of the flow regime (year), as determined over the period of record or extrapolated from analyses of the U.S. Geological Survey (USGS) streamflow-gaging-station records in Michigan.

In this report, the index flow is characterized as the median flow during the lowest flow in July, August, and September. The lowest monthly median summer flow was calculated by ranking the daily average flows at each USGS streamflow-gaging station (station) for the period of record, grouped by month. The median exceedance flow for each month was determined, and the lowest monthly value in the summer was selected as the index flow for each station. Summer is the time of greatest stress on the ecosystem from low flows and high temperatures.

Multiple linear regression models (Draper and Smith, 1966) are commonly used to transfer streamflow information from gaged to ungaged sites. The regression model includes an equation for estimating or predicting the index water yield, computed as the index flow divided by the drainage area

[1] Michigan Department of Environmental Quality.

[2] U.S. Geological Survey.

contributing to flow, using basin and climatic characteristics as explanatory variables. In this report, "estimation" refers to the process of computing the square root of water yield or the corresponding index flow for a gaged site that was used in model development, whereas "prediction" refers to the process of computing the square root of water yield or the corresponding index flow for an ungaged site. Unless ambiguity would result, the term "computation" is used when the distinction between estimation and prediction is unimportant.

In addition to an equation for predicting the hydrologic response, regression models provide a probability model that describes the uncertainties of predicted responses. This uncertainty is sometimes expressed as a range of responses with a specified probability that is likely to contain the true hydrologic response at a particular stream site. The lower limit of this range can be used to help avoid overestimating a response, such as the index flow.

Purpose and Scope

This report documents the development of a multiple linear regression model for predicting the expected magnitude and uncertainty of the index water yield. The index water yield is the water yield associated with the index flow, which is the median flow for the month of lowest summer streamflow in Michigan. For ungaged sites, the predicted index water yield can be multiplied by the corresponding drainage area upstream from the site to compute the index flow. In addition to the expected magnitude of the index flow, the uncertainty characterized by the regression model provides a basis for computing a range of flows within which the true index flow is likely to occur. An example computation is given to illustrate application of the regression model for predicting water yield and computing magnitude and uncertainty of the index flow. The regression model is applicable to Michigan streams where index flows are not significantly affected by existing water withdrawals, diversions, or augmentations.

Previous Investigations

Knutilla (1967) and Holtschlag and Croskey (1984) developed statistical models for predicting a variety of low-, average-, and peak-flow characteristics for Michigan streams. Neff and others (2005) developed multiple regression equations for predicting base flow throughout the Great Lakes. None of these studies, however, resulted in a method for estimating the median streamflow during the summer month of lowest flow in Michigan. Longer periods of record and additional streamflow-gaging sites, combined with improved methods for determining basin and climatic characteristics, created an opportunity to improve estimation of streamflow characteristics in Michigan and support implementation of the 2006 water-withdrawal legislation.

Description of the Study Area

Michigan is in the eastern north-central part of the United States and is surrounded by four of the five Great Lakes (fig. 1). Ontario, Canada lies to the north and east of Michigan. To the west and south, border states are Wisconsin, Indiana, and Ohio. Michigan is the 10th largest state in the Union with a total land area of 58,110 mi^2, 38,575 mi^2 of Great Lakes waters, and 1,305 mi^2 of inland waters (Michigan Library, 2006). According to the U.S. Census Bureau (2007), the population of Michigan in 2006 was estimated to be 10,095,643.

Michigan has a humid continental climate in which the average precipitation (rainfall plus water-equivalent snowfall depths) varies from about 28 to 38 in/yr. December through March tend to have slightly less precipitation, whereas July through September tend to have slighter more precipitation, than is typical for the rest of the year. Greater evapotranspiration during the summer, however, generally causes summer streamflows to be lower than those at other times of the year.

Michigan consists of two peninsulas separated by the Straits of Mackinac, a body of water that connects Lake Michigan with Lake Huron (fig. 1). The straits are spanned by the Mackinac Bridge, where the northern tip of the Lower Peninsula is within about 5 mi of the southern coast of the eastern Upper Peninsula. The Upper Peninsula is heavily forested and somewhat mountainous in the west. Bedrock is at or near the surface in much of the Upper Peninsula.

The Lower Peninsula is covered by a thick layer of glacial drift. The northern part is characterized by sandy material and is heavily forested. Trout streams, sustained by plentiful base flow, are common in that area. Much of the southeastern part of the Lower Peninsula is flat lakebed plains that are extensively agricultural or urban; base flow is meager. The southwestern part of the Lower Peninsula has a wide mixture of landforms, soil types, land uses, and stream types.

Figure 1. Michigan's Upper And Lower Peninsulas and surrounding states and province.

Regression Modeling

A multiple linear regression model was developed to predict index water yield. The model consists of a linear equation that is a function of selected hydrologic characteristics and model parameters estimated from index flow divided by drainage areas at gaged sites. This equation, plus the probability model underlying the error distribution, form the regression model. The following paragraphs describe the mathematical procedures used to estimate the model parameters from available data and assumptions underlying the probability model. Techniques are described for using the model uncertainty and site-specific climatic and basin characteristics to bound model predictions with a specified level of certainty.

The general form of a multiple linear regression equation is

$$y = X \cdot \beta + \varepsilon \quad (1)$$

where

- y is a column vector containing the hydrologic response variable;
- X is referred to as the "design matrix," which, in general, is composed of p columns of basin and climatic characteristics augmented with a leading column of 1's that serve as explanatory variables to estimate the hydrologic response;
- β is a column vector of parameters, $\beta_0, \beta_1, \ldots, \beta_p$, that relate the explanatory variables to the hydrologic response variable; the ordinary least-square estimator of β is denoted β_{ols} and computed as $\beta_{ols} = (X' \cdot X)^{-1} \cdot X' \cdot y$, where the prime symbol implies a matrix transpose and the –1 power implies a matrix inverse operation;
- ε is a vector of residuals that is assumed to be normally distributed and independent with mean zero and constant variance σ^2, commonly written $\varepsilon \sim NI(0,\sigma^2)$. In addition, it is assumed in the regression model that the covariance between ε and X, $Cov(\varepsilon, X)$, equals zero.

Along with the predicted value itself, the distributional characteristics of the regression model error and the hydrologic characteristics at the site of interest are a basis for assessing the uncertainty of the predicted value. Let $[LPL_{\alpha/2}, UPL_{1-\alpha/2}]$ be a prediction interval between the lower prediction limit $LPL_{\alpha/2}$ and the upper prediction limit $UPL_{1-\alpha/2}$ centered about the regression estimate that is likely to contain the hydrologic response, y_0, with a probability of $1-\alpha$. For example, if α was specified as 0.2, there would be a 10-percent chance that the hydrologic response would be less than $LPL_{\alpha/2}$ and a 10-percent chance that it would be greater than $UPL_{1-\alpha/2}$, providing a total probability of 20 percent that the true hydrologic response would be outside the prediction interval.

Computationally,

$$\left[LPL_{\alpha/2}, UPL_{1-\alpha/2}\right] = x_0 \cdot \beta_{ols} \pm t_{n-p-1,1-\alpha/2} \sqrt{s^2 \left(1 + x_0 \left(X'X\right)^{-1} x_0'\right)} ,$$

where x_0 is a row vector of corresponding basin characteristics at the site of interest augmented by a leading 1, $(X'X)^{-1}$ is a function of the design matrix used to estimate the model parameters, $s^2 = \varepsilon' \cdot \varepsilon / (n-p-1)$ and $t_{n-p-1,\alpha/2}$ is the inverse of Student's t cumulative distribution function with $n-p-1$ degrees of freedom at the specified alpha level divided by 2.

The assumption that the regression residuals are normally distributed is often difficult to satisfy with water yield or flow values. In particular, the density function of the normal distribution is symmetrical, whereas water yield and flow data tend to be positively skewed because these variables are bounded by zero on the left and unbounded on the right side of the distribution. Logarithmic and square-root transformations are commonly applied to water yield and flow values to produce a hydrologic response variable for which model residuals are likely to be normally distributed. Unlike the logarithmic transformation, the square-root transformation does not eliminate observations that have zero values.

As a convenience to the interested reader, the following key statistics are defined. The total sum of squares is $SS_T = (y-\bar{y})'(y-\bar{y})$, where $\bar{y}$ is the mean of the hydrologic response variable, the sum of squared errors is $SS_E = (y-\hat{y})'(y-\hat{y}) \varepsilon' \cdot \varepsilon$, and the model sum of squares is $SS_M = SS_T - SS_E$. The mean square total is $MS_T = SS_T / (n-1)$. Degrees of freedom for the error is $df_E = n-p-1$, which subtracts the number of model explanatory variables, p, plus 1 for the intercept term, to describe the effective number of observations associated with the model error. The mean square error is $MS_E \equiv s^2 = SS_E / df_E$ and the model mean square is $MS_M = SS_M / p$. The root mean square error is the square root of the mean square error, $RMS_E = \sqrt{MS_E}$. In addition, an F statistic, computed by dividing the MS_M by the MS_E, characterizes the overall statistical significance of the model. On the basis of the F probability distribution, a probability value (*p-value*) is computed with the F statistic, as well as the degrees of freedom in the model and error components, to assess the likelihood that the null hypothesis that all model parameters are zero is true. A small *p*-value, commonly (but not necessarily) less than 0.05, is used to reject the null hypothesis, thereby accepting the alternative hypothesis that, overall, the regression model is statistically significant.

The Pearson multiple coefficient of determination, here denoted as R_p^2, describes the fraction of the variability of y described by $\hat{y}$ where $R_p^2 = 1 - SS_E / SS_T$. R_p^2 is equal to the squared Pearson's product moment correlation coefficient between hydrologic response variables computed on the basis of streamflow-gaging station records and values estimated by the regression equation, $r_p(y, \hat{y})$, where

$$r_P(y,\hat{y}) = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(\hat{y}_i - \bar{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \cdot \sqrt{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{\hat{y}}\right)^2}} \quad (2)$$

Indicators of estimation accuracy, such as R_p^2 and RMS_E, reflect the model fit to the dataset used in development of the regression model. These indicators tend to show model improvement with increasing numbers of explanatory variables because the model is increasing fit to the specific characteristics of the available observations. Prediction accuracy, which is associated with the accuracy of predicting the hydrologic response from a basin not used in model development, is more difficult to quantify with small datasets. Prediction accuracy, however, improves with the addition of explanatory variables only up to a point. Beyond this point, prediction accuracy may decrease with the further addition of explanatory variables because a model that too closely fits the specific characteristics of available observations may not generalize well.

To lessen the inflation of the model fit sometimes indicated by the R_p^2 value, the adjusted coefficient of determination, $R_{adj}^2 = 1 - \left(SS_E/n-p-1\right)/\left(SS_T/(n-1)\right)$, accounts for the number of parameters in the model. Spearman's correlation coefficient is a more robust measure of association than Pearson's correlation coefficient when the data distributions are skewed. The Spearman's correlation coefficient is computed similarly to Pearson's correlation coefficient except that the original data are replaced by their ranks, $r_s(y,\hat{y}) = r_p\left(rank(y), rank(\hat{y})\right)$, where the rank of the smallest value in the set is 1 and the rank of the largest value is n. Finally, Spearman's coefficient of determination, symbolized as R_S^2, is the square of Spearman's correlation coefficient.

Like the response estimates, estimated parameters β_{ols} associated with the individual explanatory variables are uncertain. As the sample size (n) becomes large, estimated parameters are unbiased and normally distributed about their true values, assuming that $\varepsilon \sim NI(0,\sigma^2)$. The covariance of the parameter estimates, $\mathrm{cov}(\beta_{ols})$ is equal to $\sigma^2(X' \cdot X)^{-1}$, which is commonly written $\beta_{ols} \sim N_{p+1}\left(\beta, \mathrm{cov}(\beta_{ols})\right)$. Diagonal elements of the covariance matrix describe the variance of the corresponding estimated parameters; off-diagonal elements describe the covariance among parameters. A large covariance among parameters indicates a coupling between one or more parameter estimates because of an approximate linear dependency among explanatory variables. Such a coupling complicates interpretation of parameter magnitudes associated with specific explanatory variables. The magnitude of parameter covariances is commonly evaluated on the basis of their correlations, computed as $\mathrm{cor}(\beta_{ols})_{i,j} = \mathrm{cov}(\beta_{ols})_{i,j} / \sqrt{\mathrm{cov}(\beta_{ols})_{i,i} \cdot \mathrm{cov}(\beta_{ols})_{j,j}}$. If the magnitudes of these correlations $\left|\mathrm{cor}(\beta_{ols})_{i,j}\right|$ exceed 0.95 (Poeter and others, 2005), the independence of the paired parameter estimates is uncertain.

A t statistic computed from the data can be used to assess the significance of individual parameters as $t = (\beta_{ols})_{i,j} / \sqrt{\mathrm{cov}(\beta_{ols})_{i,i}}$. This t statistic is used to compute the probability that the null hypothesis, $\beta_1 = 0$, is true. If this computed p-value is small, say less than 0.05, the null hypothesis is commonly rejected, and the alternative hypothesis that $\beta_1 = \beta_{ols,i}$ is effectively accepted. Rejecting the null hypothesis implies that the parameter $\beta_{ols,i}$ and corresponding explanatory variable are needed in the regression model.

Development of a Regression Model for Index Flow Estimation

Regression models are a statistical means of transferring flow information obtained at streamflow-gaging stations to ungaged sites in the same hydrologic region. The process of transferring flow information from gaged to ungaged basins is commonly referred to as "flow regionalization." The transfer is facilitated by identifying climatic and basin (hydrologic) characteristics in the gaged basins that are statistically related to the flow statistics computed from gaging-station records. Once this statistical relation is identified and regression parameters in the model equation are estimated, only the selected climatic and basin characteristics upstream from the ungaged site are needed to estimate the flow statistic for that location by use of the regression equation.

The regression model includes this equation and a set of assumptions pertaining to the model errors, which are the discrepancies between estimates of the flow statistics computed from gaging-station records and estimates computed by use of the regression model. It is often necessary to transform the streamflow statistics being estimated to satisfy assumptions associated with the model error. In regional flow analysis, a square-root or logarithmic transformation of the streamflow statistics is commonly applied prior to estimating regression-model parameters. The inverse transform is commonly applied to regression estimates to compute the flow statistics of interest. The spatial distribution of model error is investigated to assess whether any bias occurs among the hydrologic subregions forming the region. If no subregional bias is detected, the regression model is considered appropriate for estimating the flow characteristic of interest throughout the region. The regression model error characteristics also are a basis for computing an interval about the regression estimate in which the true, but unknown, value of the streamflow statistic is likely to occur.

Selection of Streamflow-Gaging Stations

Development of the regression model for regional flow characterization includes selection of streamflow-gaging stations where (1) no trends occur in the mean and variance of flow, (2) the period of record is sufficiently long to accurately characterize flow conditions of interest through statistical analysis of station records, (3) flow characteristics of interest are not substantially affected by water withdrawals, diversions, or regulation, and (4) streamflow represents the natural hydrologic response to climatic conditions and basin characteristics that are typical of the area.

Streamflow data from the USGS network of continuous-record streamflow-gaging stations operated in Michigan through water year 2005 were used for this analysis. A water year is the 12-month period from October 1 to September 30 and is identified by the calendar year in which it ends. Stations were selected for the regression analysis with respect to the following criteria:

1. A minimum of 10 years of continuous-record data was required to reduce the temporal sampling variability of the flow statistic.
2. Estimates of daily flow were not thought to be appreciably affected by water withdrawal, diversion, or augmentation.
3. Effects of regulation, either from natural storage in lakes or retention in regulated surface-water bodies, were not thought to substantially mask the hydrologic response from precipitation.

From these evaluations, 147 streamflow-gaging stations were selected for inclusion in the analyses (figs. 2 and 3). Among selected stations, the average length of record was 40.2 years, and the range was from 11 to 91 years. The first water year of record used in the analysis was 1901, and 88 stations included data from water year 2005.

Identification of the Hydrologic Response Variable

The regression equation described in this report is a basis for computing an estimate of the index flow, which is defined as the median streamflow for the summer month of lowest flow in Michigan. The statistical distribution of index flows, however, is not consistent with assumptions underlying the regression model. To find a metric of index flow that is consistent with these assumptions and one in which climatic and basin characteristics physically associated with the streamflow response are more readily identified, mathematical transformations of index flow values were investigated. As a result, the response variable used in the regression equation was formed as the square root of the quotient of index flow divided by the drainage area of its associated basin. In this report, the response variable is referred to as the "hydrologic response variable." The inverse transformation is applied to the regression estimates of the hydrologic response to compute index flows.

Index Flow

In accordance with PA33–2006, the median flow during the lowest summer flow month was the index flow and is represented symbolically as IQ_{50}. A statistic was calculated to estimate IQ_{50} by ranking the daily mean flows measured at each selected gaging station by month for the entire period of record available and selecting the 50th percentile. The median flow for each summer month (July, August, and September) was determined, and the summer month with the lowest median flow was used to estimate IQ_{50} at that gaging station. To distinguish the true index flow IQ_{50} from the value of the flow response computed by use of the finite period of gaging-station record, the gaging station statistic is symbolized as $\tilde{I}Q_{50}$. The value of $\tilde{I}Q_{50}$ is assumed to converge to IQ_{50} as the length of gaged record increases. This assumption requires that there is no trend in the streamflow data (the expected value of IQ_{50} does not vary with time) and that $\tilde{I}Q_{50}$ is an unbiased estimator of IQ_{50}. No trends were detected in streamflow data at the selected stations.

For the 147 stations selected for the analysis, the lowest median flow occurred in July at 5 stations, in August at 92 stations, and in September at 50 stations. The index flow ranged from zero at stations 04157500, Sebewaing River State Drain near Sebewaing, Mich., and 04158000, Columbia Drain near Sebewaing, Mich., to 1,850 ft^3/s at station 04101500, St. Joseph River at Niles, Mich. (Appendix A). The average index flow at selected stations was 116 ft^3/s, the standard deviation of these flow was 228 ft^3/s, and the (dimensionless) skewness was 4.5044.

Index Water Yield

Much of the variability in index flow is related to drainage area (fig. 4)[3]. In development of the predictive equation, there was concern that the dominant relation between index flow and drainage area indicated by the power equation $\hat{I}Q_{50} = \beta_0 \cdot DA^{\beta_1}$ could mask more subtle relations involving basin and climatic characteristics. Also, the estimated exponent in the power equation, $\hat{\beta}_1$ of 1.2301, implies a slightly nonlinear relation between drainage area and index flow (fig. 4). A nonlinear relation between index flow and drainage area is considered physically unlikely because much of the index flow is thought to be derived from ground-water sources, which would be approximately linearly related to drainage area ($\beta_1 \simeq 1$). To help identify the appropriate relations, the index water yield IY_{50} was selected as a preferred metric to the index flow. IY_{50} was estimated from station records by dividing $\tilde{I}Q_{50}$ by the drainage area upstream from the corresponding gaged site, and it is symbolized by $\tilde{I}Y_{50}$.

For the 147 selected sites, $\tilde{I}Y_{50}$ ranged from zero at the two stations with zero index flows, to 1.3087 $ft^3/s\text{-}mi^2$ at station 04139000, Houghton Creek near Lupton, Mich. The average $\tilde{I}Y_{50}$ value was 0.3302 $ft^3/s\text{-}mi^2$, the standard deviation was 0.2600 $ft^3/s\text{-}mi^2$, and the skewness was 1.3422. The positive

[3] Data for two of the selected streamflow gaging stations could not be included in this plot because they were zero and could not be represented on a logarithmic scale.

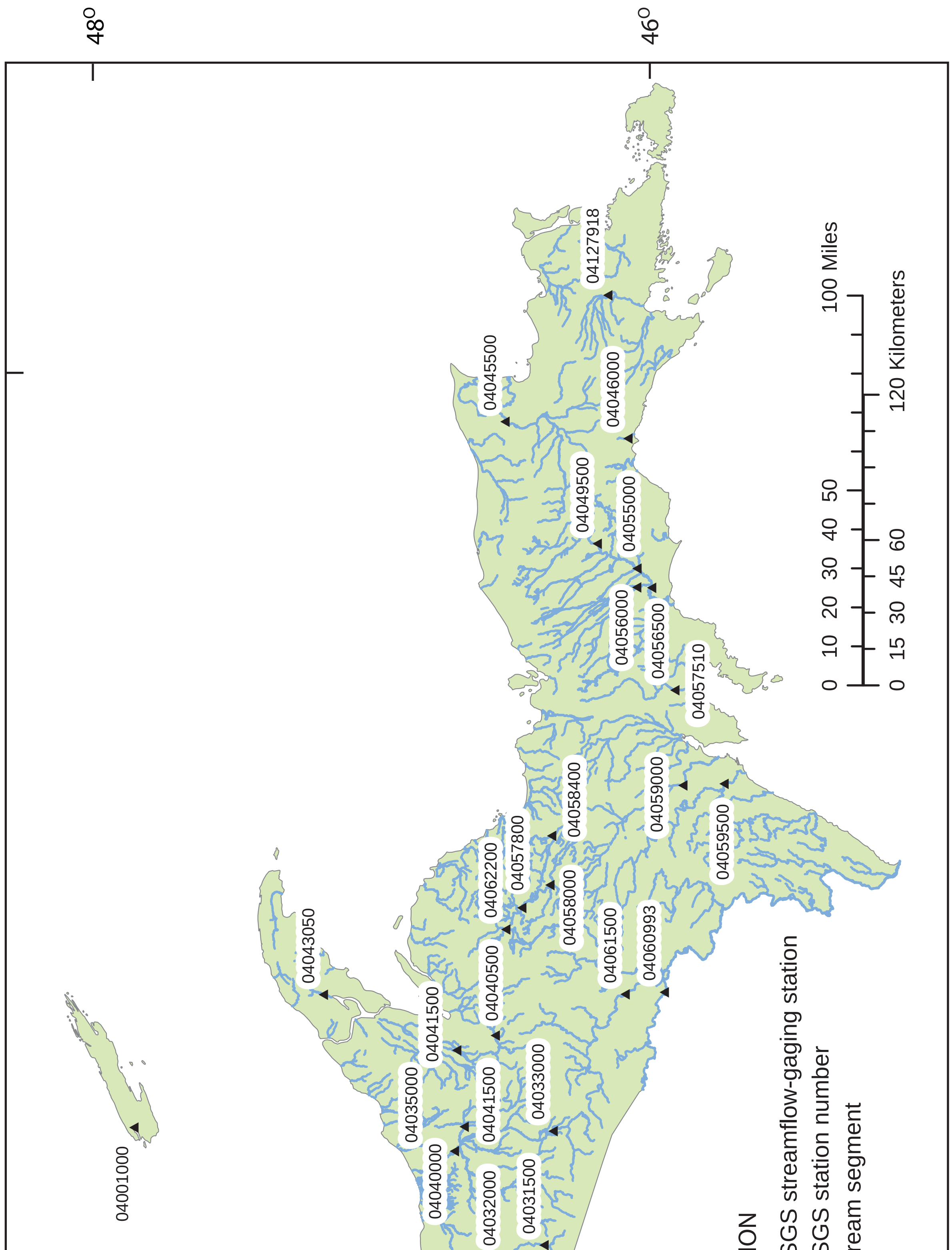


Figure 2. U.S. Geological Survey streamflow-gaging stations in Michigan's Upper Peninsula included in the analyses.

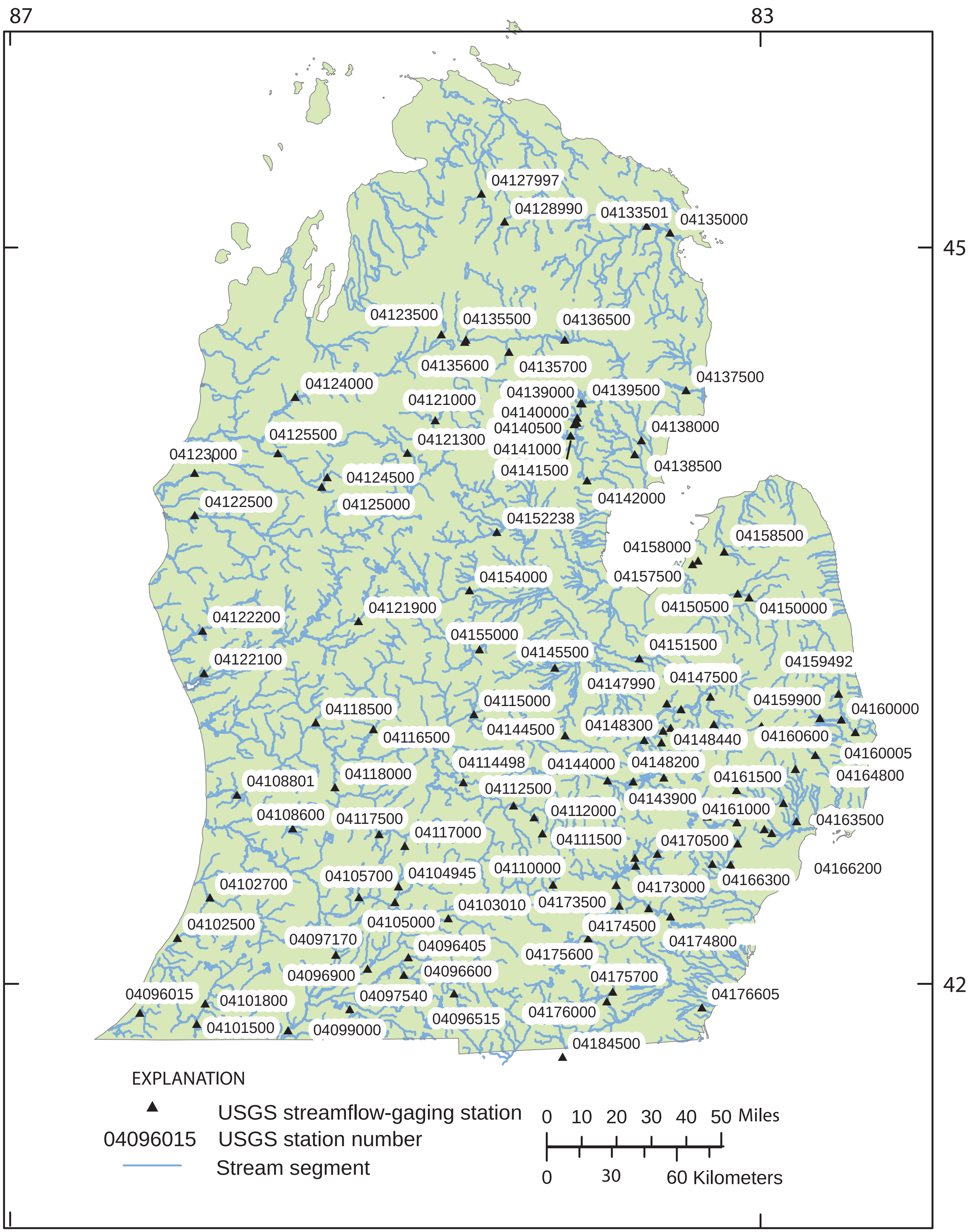


Figure 3. U.S. Geological Survey streamflow-gaging stations in Michigan's Lower Peninsula included in the analyses.

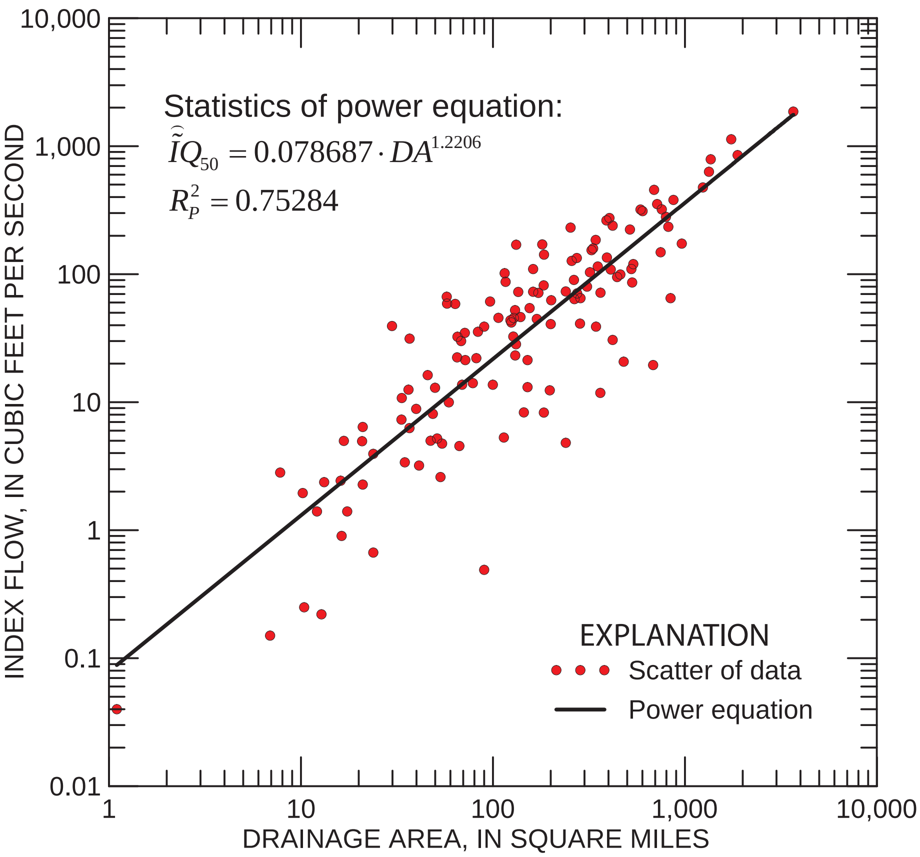


Figure 4. Relation between estimates of index flow from gaging station records, $\tilde{I}Q_{50}$, and drainage area (*DA*) [R_p^2, the Pearson coefficient of determination].

skewness indicates that index water yield values are spread out more to the right than to the left of the mean. Drainage areas range from 1.1 mi^2 at station 04141000, South Branch Shepards Creek near Selkirk, Mich., to 3,670 mi^2 at station 04101500, St. Joseph River at Niles, Mich.

A normal distribution fitted to the empirical $\tilde{I}Y_{50}$ data was inadequate to approximate the distribution of water-yield values (fig. 5), because the empirical distribution was frequently outside of the 95-percent confidence bounds of the fitted normal distribution. Formally, the Lilliefors test (Conover, 1980) rejected the null hypothesis that a normal distribution adequately approximated the distribution of $\tilde{I}Y_{50}$ at the 5-percent level ($p<0.001$) of significance. Similarly, the Lilliefors test rejected ($p<0.001$) the null hypothesis that a normal distribution adequately approximated the distribution of the common logarithm transform of the index yield ($L\tilde{I}Y_{50}$).

A square-root transformation was applied to the elements of $\tilde{I}Y_{50}$ to assess the effect on the empirical distribution of the resulting values. Based on a sample mean of 0.5274 (ft^3/s-mi^2)$^{1/2}$, variance of 0.0525 ft^3/s-mi^2, and a skewness of 0.1607, a normal distribution closely approximated the empirical distribution of square root (Root) transformed values symbolized as $R\tilde{I}Y_{50}$ (fig. 5). A Lilliefors Test did not reject the null hypothesis that $R\tilde{I}Y_{50}$ values were normally distributed at the 5-percent level of significance (p=0.5). Therefore, the square root transformation $R\tilde{I}Y_{50}$ was used as the hydrologic-response variable in the regression model.

Compilation of Hydrologic Characteristics for Use as Explanatory Variables

Hydrologic characteristics were compiled for the 147 stations used in these analyses. Compiled hydrologic characteristics include basin and climatic characteristics that are considered physically and statistically related to the IY_{50}. All hydrologic characteristics included as possible explanatory variables in the regression equation are available as Geographic Information System (GIS) files to facilitate computation of hydrologic characteristics. Basin characteristics included categories of aquifer transmissivity, forested area, and hydrologic soil group; climatic characteristics included normal (1971–2000) annual precipitation and annual snowfall amounts. The following paragraphs discuss the hydrologic

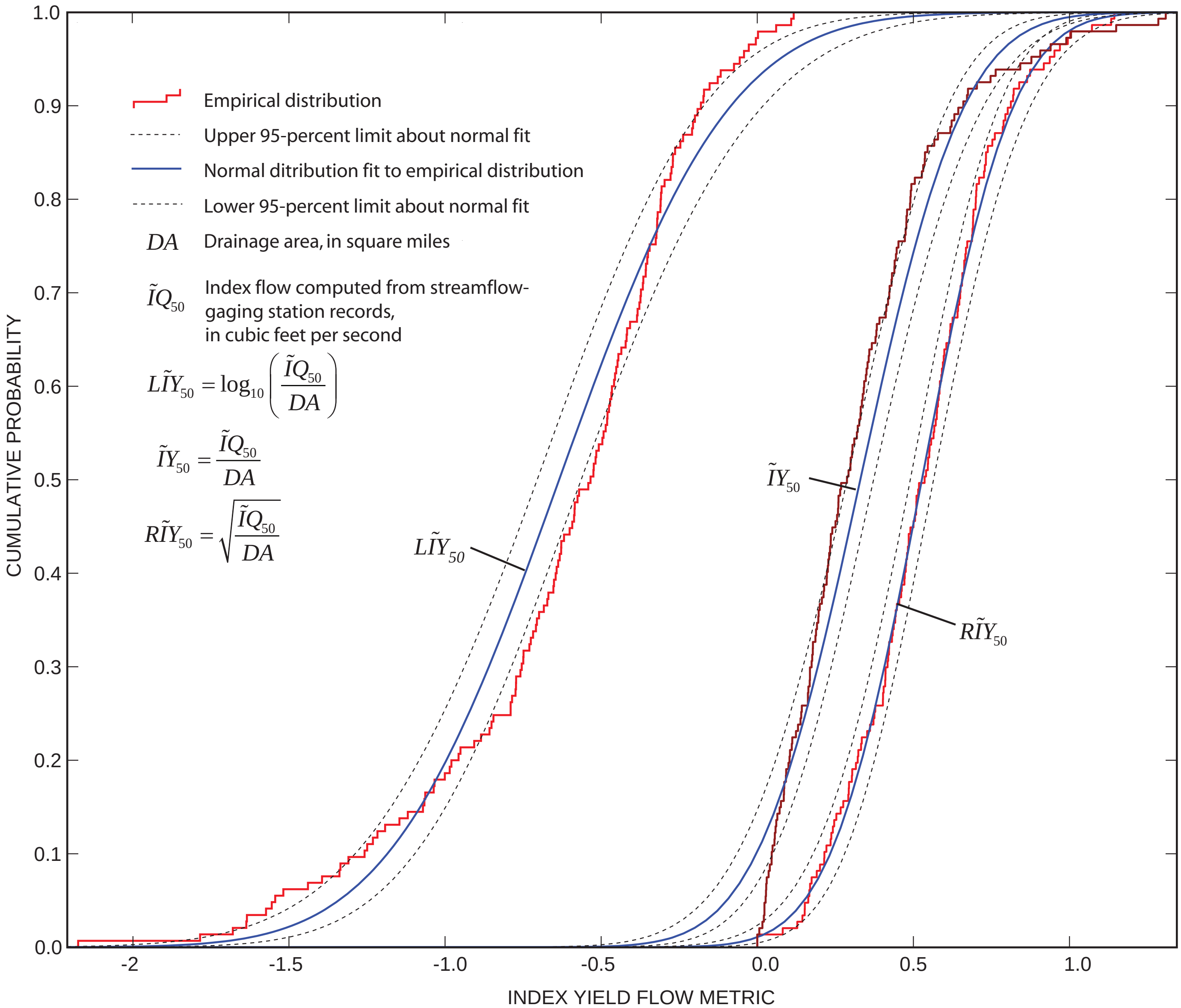


Figure 5. Empirical and fitted normal distributions for median-water-yield data from the month of lowest flow for selected streamflow-gaging stations in Michigan.

characteristics evaluated as possible explanatory variables in the regression equation.

Transmissivity is a measure of the capacity of an aquifer to transmit water. The transmissivity of an aquifer is equal to its hydraulic conductivity, commonly expressed in units of feet per day, multiplied by its saturated thickness, in feet. The Goundwater Mapping Project (*http://gwmap.rsgis.msu.edu/*), a multiagency study in Michigan, created a grid of the estimated transmissivities for the glacial deposits (Michigan Department of Information Technology, 2005a). The grid is composed of 1-km (0.621-mi) square elements and is based on an interpolation of transmissivities assigned to 270,000 water wells on the basis of lithologic information described in well logs prepared by well drillers. In areas of thin glacial deposits (less than 30 ft thick) the grid element was assigned a code of -1 to indicate that thin deposits prevented a reliable estimation of transmissivity at that element. Because of the uncertainty associated with interpolation over the highly heterogeneous aquifer transmissivity field, grid elements that were more than 2,000 m (6,560 ft) from a well were assigned a code of -2 to indicate that interpolation uncertainties prevented reliable estimation of transmissivity at that element. Otherwise, grid elements were assigned an estimated transmissivity value that ranged from 0 to 30,309 ft^2/d.

The Michigan Glacial Landsystems Coverage (Michigan Department of Information Technology, 2005b) classified the surface geologic deposits into 10 land systems. Each applicable land system was assigned to an aquifer transmissivity

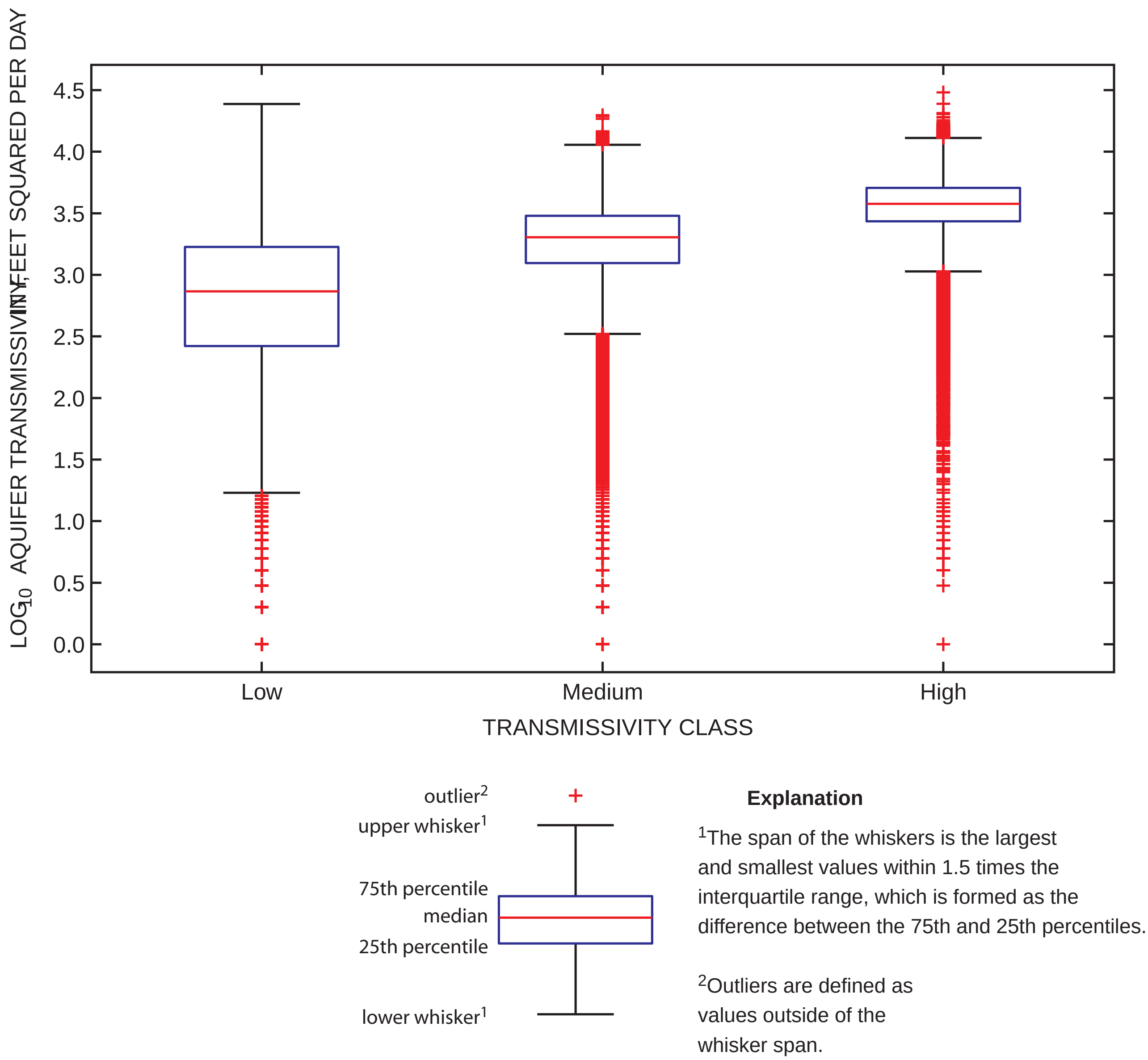
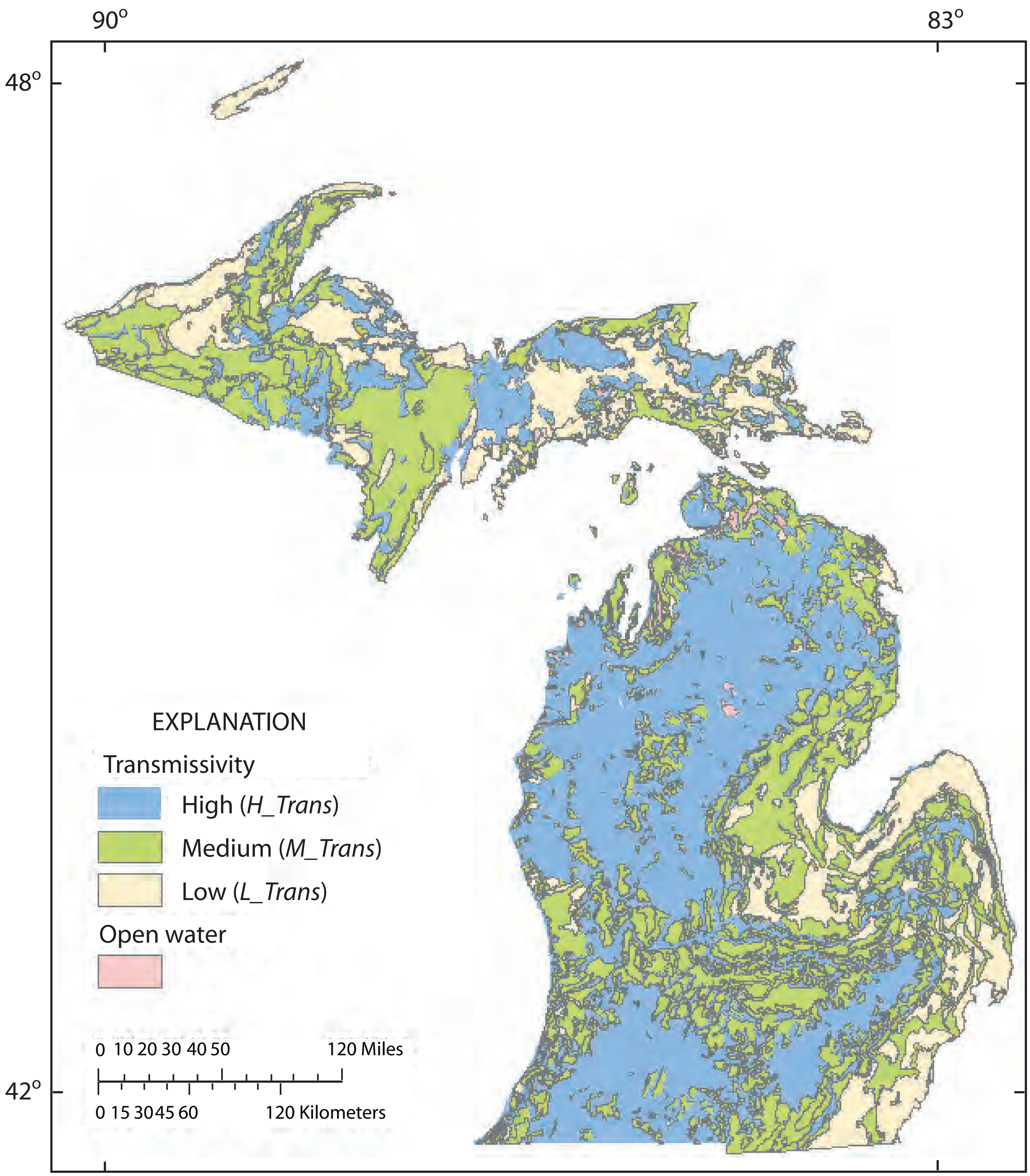


Figure 6. Distribution of estimated aquifer transmissivity within transmissivity classes in Michigan.

class. Bedrock, lacustrine fine, and thin drift over bedrock land systems were assigned to the low-transmissivity class; lacustrine coarse, lodgement till or fine supraglacial drift, and ice-marginal till land systems were classified as medium transmissivity; and coastal dunes, ice-contact outwash, and proglacial outwash were assigned to the high-transmissivity class. Land systems designated as lakes were not assigned a transmissivity class. About 0.25 percent of the elements were assigned aquifer transmissivities of zero and could not be displayed by means of a common logarithm transformation ($\log_{10}$). The $\log_{10}$ transformed distribution of aquifer transmissivities that were estimated to be greater than zero are shown for low, medium, and high classes of transmissivities in figure 6. Median estimated aquifer transmissivities increased from 723 ft^2/d in areas classified as low transmissivity, to 2,020 ft^2/d in areas classified as medium transmissivity, to 3,780 ft^2/d for areas classified as high transmissivity. The spatial distribution of estimated transmissivity classes in the glacial aquifers in shown in figure 7.

Land-use and land-cover characteristics affect hydrologic response primarily by affecting the rate at which water infiltrates into the soil and subsequently either drains to the ground-water system or flows overland to a nearby stream. As indicated by Anderson and others (1976), land use refers to “man’s activities on the land that are directly related to the land” (Clawson and Stewart, 1965), whereas land cover describes “the vegetative and artificial construction covering the land” (Burley, 1961).



Michigan aquifer transmissivity classes from Michigan Department of Information Technology, 2005b

Figure 7. Distribution of aquifer transmissivity classes in Michigan.

The State of Michigan uses the spatial data coverages in the Michigan Resource Information System (MIRIS) (1978) as the standard for hydrologic studies in Michigan. MIRIS contains land-use and land-cover data that had been compiled from county and regional planning commissions. The MIRIS data represent land-use and land-cover data in a grid that contains 26,319 rows and 25,247 columns of cells. Each cell represents a land area of 30 m square. The categories include Level I features (Anderson and others, 1976), which are coded in MIRIS as integers and are defined as follows: (1) urban or built-up land; (2) agricultural land; (3) rangeland; (4) forest land, which included Level II classification of deciduous, evergreen, and mixed forest lands; (5) water; (6) wetland; and (7) barren land. The code –9999 signifies no data or inapplicable, which occurs over areas such as the Great Lakes. The spatial distribution of forest land in the MIRIS coverage is shown in figure 8.

Four hydrologic soil groups have been defined by the U.S. Department of Agriculture Natural Resources Conservation Service (NRCS) (2007):

- Group A soils (basin characteristic variable *A_Soils*) have low runoff potential when thoroughly wet. Water is transmitted freely through the soil. Group A soils typically have less than 10 percent clay and more than 90 percent sand or gravel and have gravel or sand textures.

- Group B soils (*B_Soils*) have moderately low runoff potential when thoroughly wet. Water transmission through the soil is unimpeded. Group B soils typically have between 10 percent and 20 percent clay and 50 percent to 90 percent sand and have loamy sand or sandy loam textures.

- Group C soils (*C_Soils*) have moderately high runoff potential when thoroughly wet. Water transmission through the soil is somewhat restricted. Group C soils typically have between 20 percent and 40 percent clay and less than 50 percent sand and have loam, silt loam, sandy clay loam, clay loam, and silty clay loam textures.

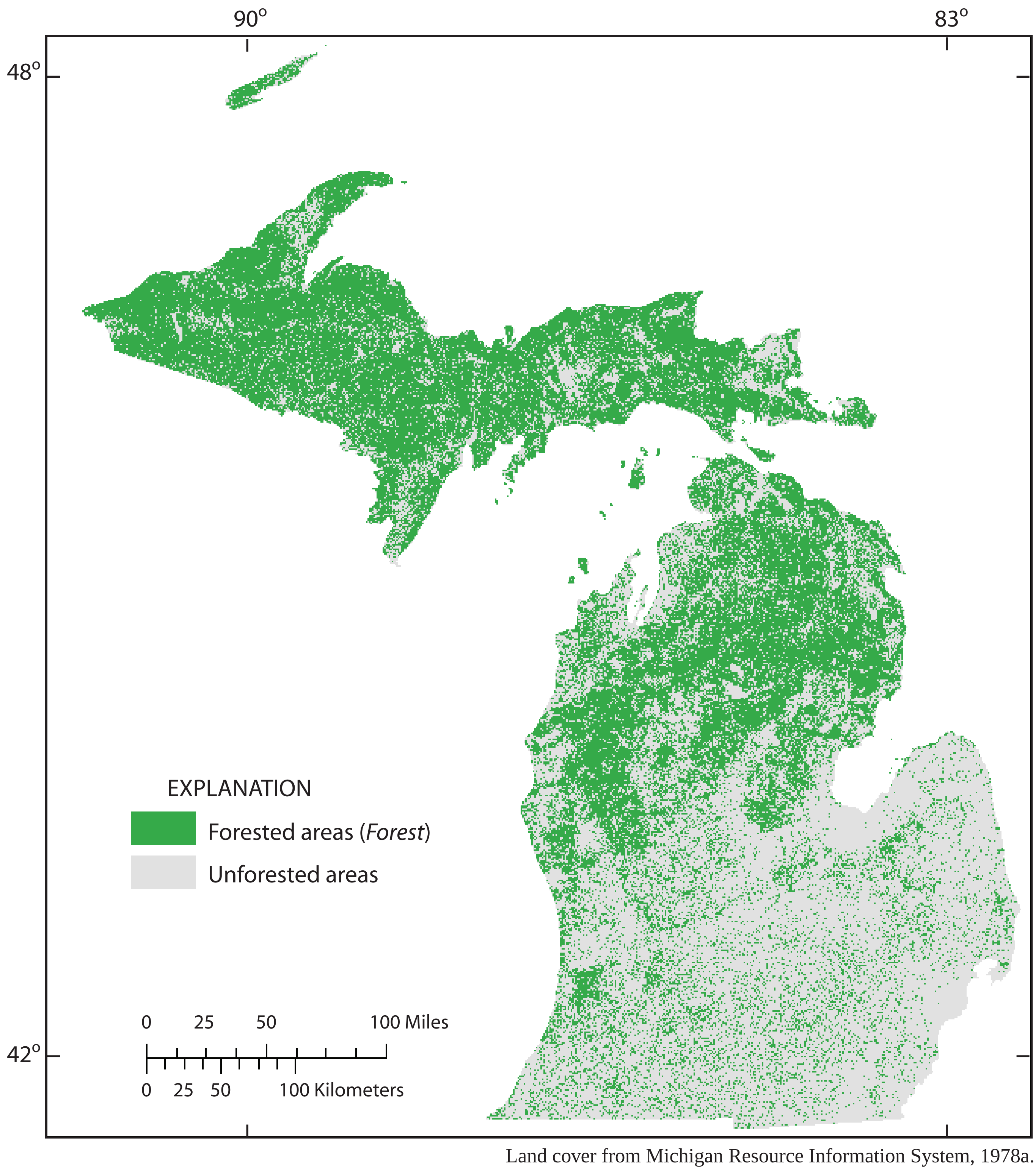
- Group D soils (*D_Soils*) have high runoff potential when thoroughly wet. Water movement through the soil is restricted or very restricted. Group D soils typically have greater than 40 percent clay, less than 50 percent sand, and have clayey textures.

The spatial distribution of hydrologic soil groups in Michigan is shown on figure 9 based on the MIRIS coverage. MIRIS represents soil data in a grid that contains 26,319 rows and 25,247 columns of cells. Each cell represents a land area of 30 m square. In MIRIS, hydrologic group A soils are coded as 1, group B soils are coded as 2, group C soils are coded as 3, group D soils are coded as 4, and no data or inapplicable areas are coded as -9999. The hydrologic-soil-group grid is georeferenced the same as the MIRIS grid for land use and land cover.

Runoff curve numbers (*RCN*) were developed by the U.S. Department of Agriculture National Resources Conservation Service (2004). Conceptually, *RCN* describes the direct runoff component of total flow that includes (1) the channel component representing precipitation falling directly on the stream channel, (2) the surface or overland flow component, which represents flow from precipitation that exceeds the infiltration rate on the land surface, and (3) the subsurface component, which represents infiltrated water that flows laterally underground to the stream without intercepting permanently saturated areas; this subsurface flow component is sometimes referred to as “interflow.” With reference to *RCN*, runoff does not include the base-flow component, which is likely the main component influencing IQ_{50}. Large direct-runoff components, however, are likely to be associated with smaller median or base-flow components. In general, greater *RCN* values are associated with soils with greater peak runoff potential, such as areas underlain the hydrologic soil group D; within each soil group, *RCN* increases with percentages of impervious areas, land covers that are prone to produce runoff, and basins that are considered to be in poor hydrologic condition. The MDEQ has developed GIS processing techniques for computing *RCN* from land-use and soil GIS coverages. From a possible range of 0 (no direct runoff) to 100, *RCN* ranged from 48 to 85, with an average of 70 for the selected basins. No statewide coverage is available to display the geographic variation of *RCN*, although it is similar to the hydrologic soil groups and land-use characteristics from which it is derived.

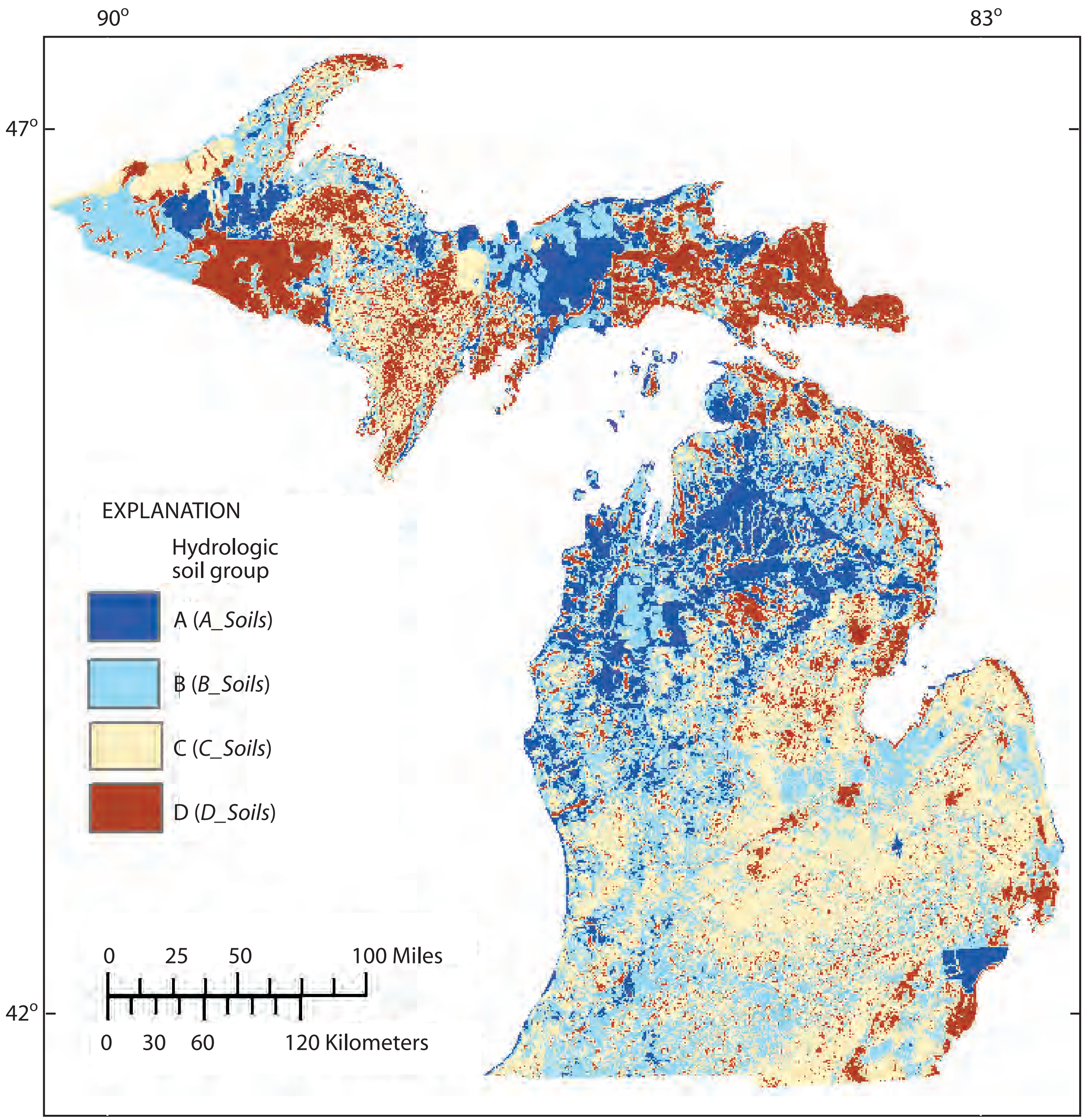
Normal annual precipitation for 1971–2000 ranged from about 28.5 in/yr in the northeastern part of the Lower Peninsula to about 38 in/yr in southeastern part of the Lower Peninsula (Michigan Climatological Resources Program, 2004). Precipitation in the far western part of the Upper Peninsula approaches 35 in/yr, whereas precipitation in the eastern part is about 32 in/yr (fig. 10).

Normal annual snowfall depths (fig. 11) for 1971–2000 in Michigan generally trend from a minimum of 40 in. in southeastern Lower Peninsula to a maximum of 220 in. in the northwestern tip of the Upper Peninsula (Michigan Climatological Resources Program, 2004). Evidence of lake-effect snow is apparent along the western coast of the Lower Peninsula and in a trend of increasing snowfall depths from south to north in the Upper Peninsula.



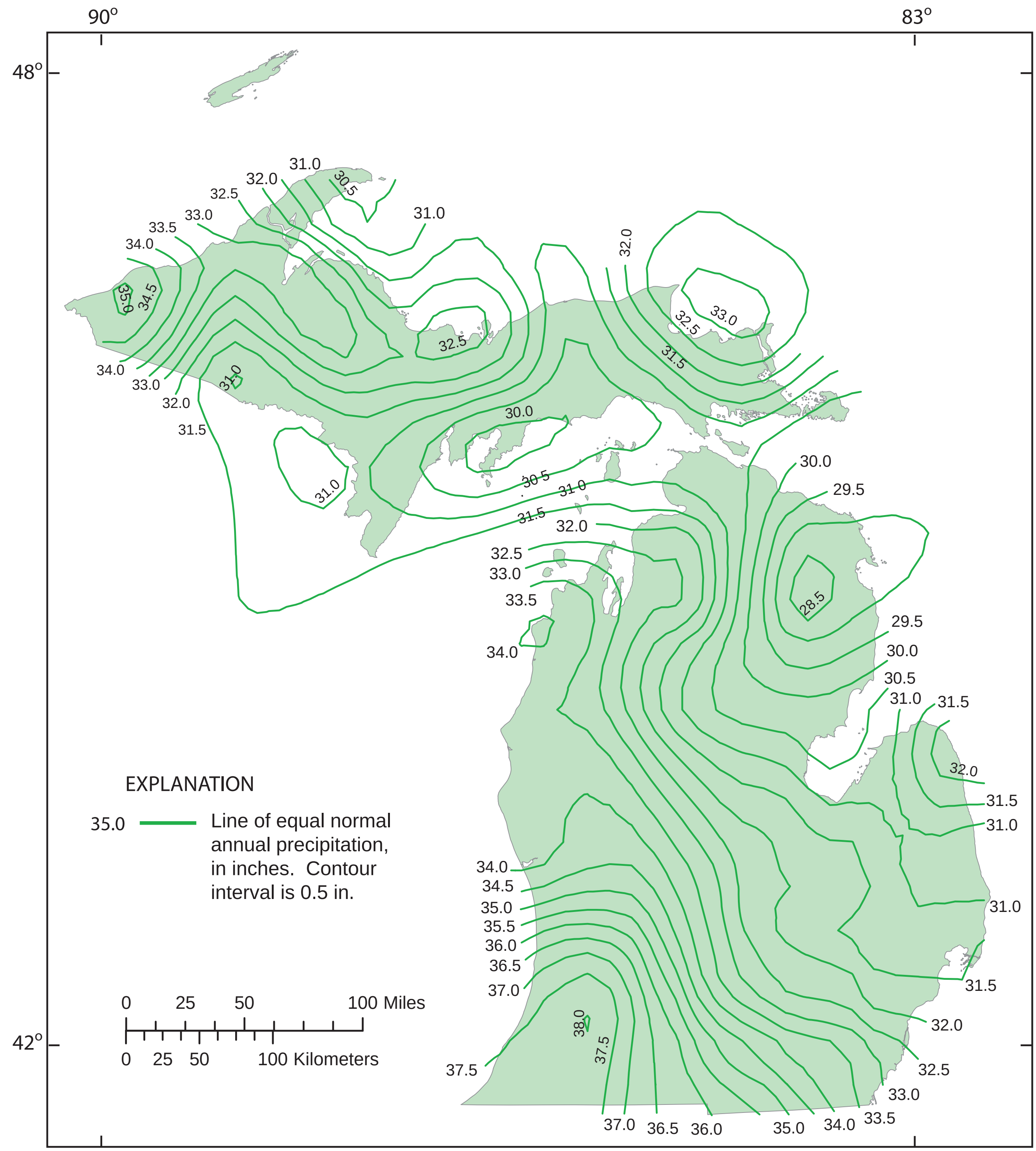
Land cover from Michigan Resource Information System, 1978a.

Figure 8. Distribution of forest cover in Michigan.



Hydrologic soil groups from Michigan Resource Information System, 1978a.

Figure 9. Distribution of hydrologic soil groups in Michigan.



Normal annual precipitation from Michigan Climatological Resources Program, 2004.

Figure 10. Distribution of normal annual precipitation in Michigan for 1971–2000.

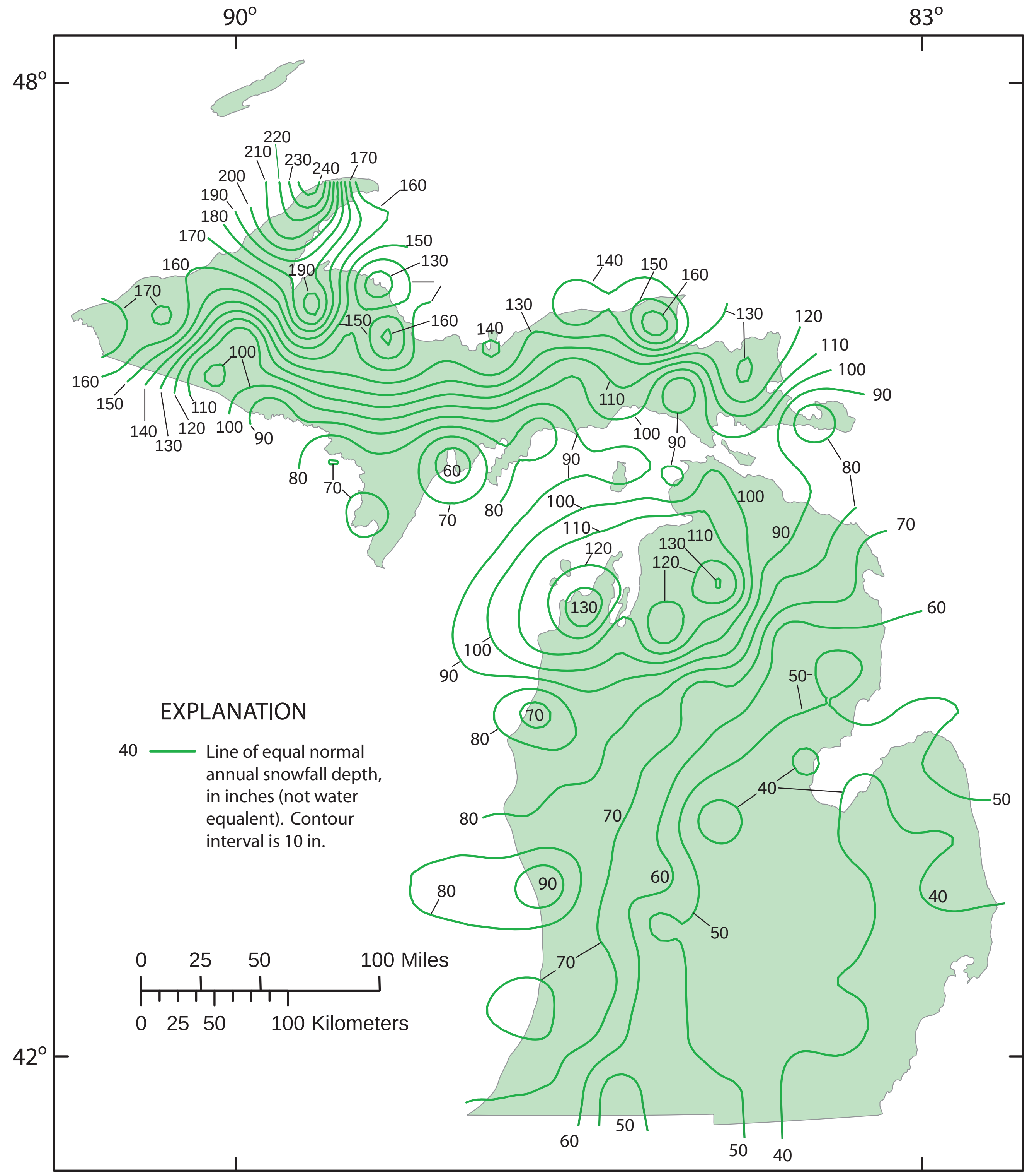


Figure 11. Distribution of normal annual snowfall depths in Michigan for 1971–2000.

Selection of Hydrologic Characteristics for Use as Explanatory Variables

Explanatory variables used in the regression equation were selected on the basis of both their statistical and hydrologic significance. One of the initial screening devices for assessing statistical associations was the matrix of correlation coefficients (table 1). Here, the maximum positive correlation (0.63) was found between $R\tilde{I}Y_{50}$ and forest (*Forest*); the maximum negative correlation (–0.72) was found between $R\tilde{I}Y_{50}$ and runoff curve numbers (*RCN*). Among explanatory variables, large negative correlations were detected between *RCN* and *A_Soils* (–0.90). A large positive correlation also was found between *Snowfall* and *Forest* (0.83).

Correlations between explanatory variables indicate some redundancy of information and result in some statistical ambiguity in identifying explanatory variables for inclusion in the regression equation. Percentages of land use classified within individual categories of both transmissivity and soil groups generally summed to 100 percent, except in some areas where soils or glacial drift were absent and the sum therefore was less than 100 percent. For these two sets of variables, intragroup categories were negatively correlated. Also, because the sums of all transmissivity and soil categories generally were 100 percent, all members of either the transmissivity or soil categories could not be included in the regression without special numerical constraints.

Initial development of the regression equation proceeded in an automated, stepwise manner. In particular, the variable most highly correlated with $R\tilde{I}Y_{50}$ was added to the equation first, followed by the variable that was most highly correlated given the presence of the first variable in the equation. The process continued until all the alternative explanatory variables were evaluated in turn. Introduction of new variables into the equation sometimes resulted in the elimination of variables previously included at an apparent significance level of 0.15.

Final selection of the regression equation was based on the following criteria:

- The model explained a significant amount of the variability in $R\tilde{I}Y_{50}$.
- The estimation error of the overall model was low.
- The number of selected explanatory variables was constrained so that model prediction error—the error applicable to sites not included in the development of the equation—would be similar to model estimation error.
- The signs and magnitudes of parameters associated with selected explanatory variables were generally consistent with the expected physical association between the individual explanatory variables and the hydrologic response.
- An apparent significance level of about 5 percent for individual parameters was generally maintained.

Estimation of the Hydrologic Response Variables

The regression equation for estimating the hydrologic response variable, $R\tilde{I}Y_{50}$, contains six explanatory variables and an intercept term. Based on the computed R^2_{adj} value, the regression model explains about 70.8 percent of the variability in $R\tilde{I}Y_{50}$ (fig. 12). The RMS_E was 0.12377, with corresponding MS_E or s^2 equal to 0.015320, and overall the *p*-value associated with the regression model was less than 0.0001 (p<0.0001). Based on the results of a Lilliefors test of normality, there was insufficient evidence to reject the normality of the residual distribution at the 0.01 level of significance (p=0.015). In this report, estimates of the index water yield, $\hat{I}Y_{50}$, were obtained by squaring estimates of $R\hat{I}Y_{50}$. After squaring individual values of $R\tilde{I}Y_{50}$ and $R\hat{I}Y_{50}$ to compute $\tilde{I}Y_{50}$ and $\hat{I}Y_{50}$ values, respectively, the $R^2_p(R\tilde{I}Y_{50}, R\hat{I}Y_{50})$ decreases from the 0.7080 determined in the regression to an $R^2_p(\tilde{I}Y_{50}, \hat{I}Y_{50})$ of 0.6128 because of the skewed distribution of the squared values. The coefficient of determination based on the ranks of the squared values $R^2_S(\tilde{I}Y_{50}, \hat{I}Y_{50})$, however, is 0.7498, which is slightly higher than the $R^2_p(R\tilde{I}Y_{50}, R\hat{I}Y_{50})$ of the more normally distributed $R\tilde{I}Y_{50}$ and $R\hat{I}Y_{50}$ values. Thus, the correlation between measured and estimated index water yield is preserved in the space appropriate to the distribution of the two variables. The mean and standard deviation of residuals between measured and estimated water-yield values are 0.0151, and 0.1622, respectively.

Explanatory variables included in the regression model, parameter estimates, and associated statistics are listed in table 2. Only the parameter associated with low transmissivity (*L_Trans*) was negatively associated with $R\tilde{I}Y_{50}$. In apparent contradiction to the suspected physical relation, the parameter associated with *D_Soils* is positively associated with $R\tilde{I}Y_{50}$ and is similar in magnitude to the parameter associated with *A_Soils*. The anomalous sign associated with *D_Soils* may be related to an association between *D_Soils* and other land-use and land-cover characteristics.

To investigate this possibility, a cross tabulation between the 1978 MIRIS land use-land cover areas with hydrologic soil groups was computed (table 1–3 of Appendix 1). The results of this tabulation indicate that 89.3 percent of the areas classified as water also were classified as group D soils and that 68.7 percent of the areas classified as wetlands also were classified as group D soils (table 3). Furthermore, 60.4 percent of the soils classified as group D also were classified as forest areas. Areas covered by water, wetlands, and forests would be expected to be associated with higher median flows than areas not associated with these land use-land cover characteristics. Thus, the positive sign of the parameter estimate for *D_Soils* is not considered physically anomalous.

Table 1. Lower triangular elements of the diagonally symmetric correlation matrix among candidate explanatory variables and the square root of median water yield for the summer month of lowest flow in Michigan.

[*H_Trans*, *M_Trans*, and *L_Trans* indicate the percentage of the land area underlain by high, medium, and low aquifer transmissivity classes, respectively; *Forest* indicates forest-covered lands; *A_Soils*, *B_Soils*, *C_Soils*, and *D_Soils* indicate the percent of land areas classified as hydrologic soil group A, B, C, and D, respectively; *RCN* indicates the runoff curve number; *Precip* indicates the normal annual precipitation for 1971–2000; *Snowfall* indicates the snowfall depths (not water equivalent); and $R\tilde{I}Y_{50}$ indicates the square root of the index water yield]

	H_Trans	M_Trans	L_Trans	Forest	A_Soils	B_Soils	C_Soils	D_Soils	RCN	Precip	Snowfall	$R\tilde{I}Y_{50}$
H_Trans	1.00	.	.	.	.	.	.	.	.	.	.	.
M_Trans	-0.69	1.00	.	.	.	.	.	.	.	.	.	.
L_Trans	-.57	-0.20	1.00	.	.	.	.	.	.	.	.	.
Forest	.13	-.24	0.09	1.00	.	.	.	.	.	.	.	.
A_Soils	.57	-.53	-.18	0.53	1.00	.	.	.	.	.	.	.
B_Soils	-.17	.29	-.09	-.57	-0.74	1.00	.	.	.	.	.	.
C_Soils	-.61	.41	.36	-.25	-.36	-0.19	1.00	.	.	.	.	.
D_Soils	-.05	-.01	.08	.41	-.11	-.19	-0.26	1.00	.	.	.	.
RCN	-.52	.44	.21	-.74	-.90	.60	.48	0.01	1.00	.	.	.
Precip	.16	-.02	-.18	-.18	-.23	.38	-.14	-.10	0.20	1.00	.	.
Snowfall	-.09	-.17	.31	.83	.26	-.37	-.08	.32	-.48	0.03	1.00	.
$R\tilde{I}Y_{50}$	.59	-.40	-.35	.63	.63	-.43	-.47	.20	-.72	.12	0.43	1.00

Correlations among parameter estimates for explanatory variables (excluding the intercept term) ranged from -0.6398 to 0.5075 (table 4), indicating no significant linear dependence among explanatory variables. Some ambiguity between the intercept term, which is associated with the leading column of 1's in the design matrix, and the parameter estimate associated with *Precip* is indicated by a correlation of –0.9881.

Values of the selected explanatory variables for all 147 observations used in regression model are in table 1–2 of Appendix 1. If a unit vector of equal length were appended before columns 3–8 in table 1–2, the table entries would be identical to the design matrix *X* used in the development of the regression model. Boxplots show the range and approximate distribution of the selected explanatory variables used in the regression equation (fig. 13).

Spatial Distribution of the Regression-Model Error

Taken over all streamflow-gaging stations in the analysis, the multiple linear regression equation developed in the report provides an unbiased estimator, $R\hat{I}Y_{50}$, of $R\tilde{I}Y_{50}$. Estimation of spatially referenced quantities without corresponding spatially referenced gaging-station coordinates as explanatory variables, however, can result in spatial patterns in the regression error. A significant spatial pattern in the distribution of regression errors would indicate that estimates could be locally biased.

To investigate the potential for local bias in regression estimates, each selected gaging station was assigned to a subregion within Michigan (fig. 14). The subregions used in this report are similar to subregions defined on USGS hydrologic unit maps (Seaber and others, 1987). So that similar numbers of streamflow-gaging stations would be included in each subregion, however, individual cataloging units shown on USGS hydrologic unit maps were grouped somewhat differently in this report than cataloging units grouped by the USGS to define subregions. In addition, the cataloging units forming the subregions in this report were clipped to the State's boundaries.

Notched boxplots show the distribution of model residuals by subregion (fig. 15). For each boxplot, the width of the notch is computed so that boxplots whose notches do not overlap would have different medians at the 5-percent level of significance. By examining the intervals spanned by the notches, however, the boxplots indicate no significant difference in median residual among hydrologic subregions. Similarly, a Kruskal-Wallis test (Conover, 1980), which compares the median residuals for each subregion, found no significant differences among subregions ($p = 0.3515$). The lack of geographic bias among subregions implies that the regression equation is applicable for all hydrologic subreaches, which together span the State of Michigan. The median residual of -0.0438 in Michigan hydrologic subregion 7 is slightly less than zero. A bootstrap analysis of residuals in subregion 7 alone, however, did not indicate that the median residual was biased at the 5-percent level of significance.

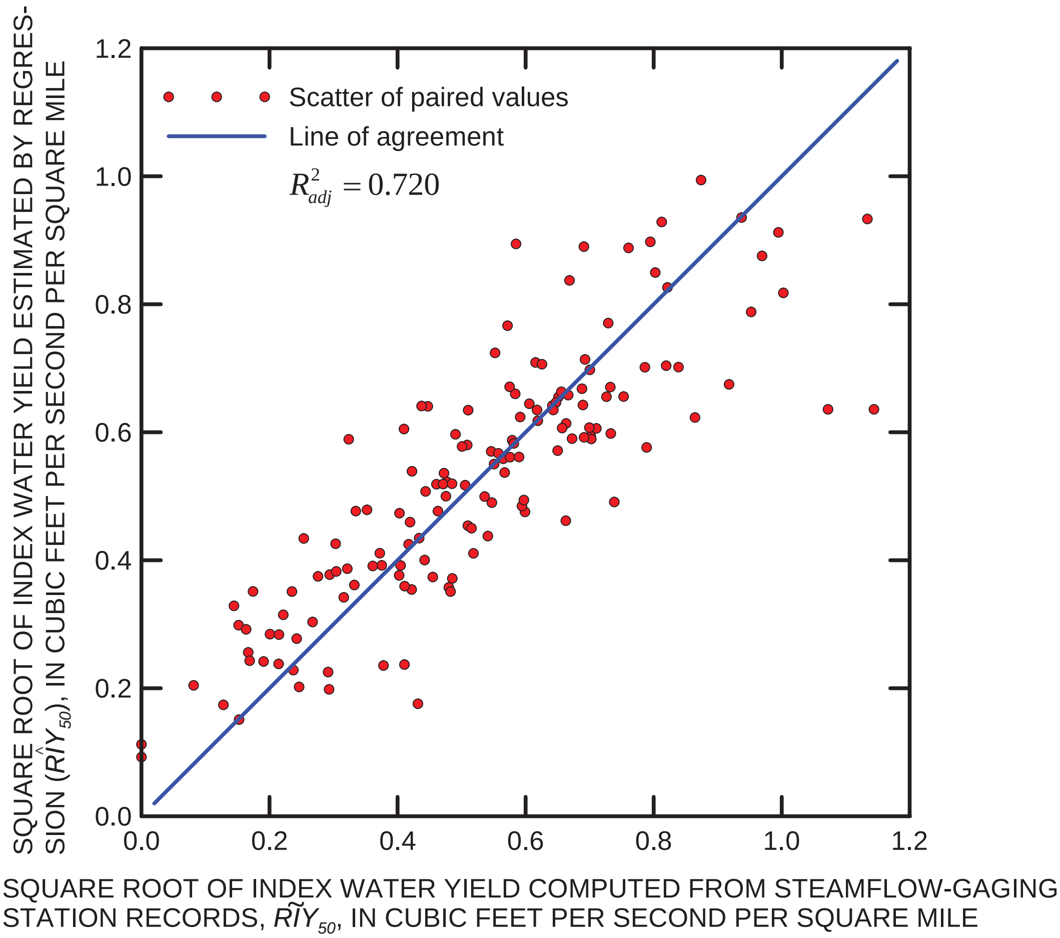


Figure 12. Relation between $R\hat{I}Y_{50}$ (the index of water yield estimated by regression) and $R\tilde{I}Y_{50}$ (the index of water yield computed on the basis of the streamflow-gaging station records) [R^2_{adj}, the adjusted Pearson coefficient of determination].

Table 2. Regression-model parameters for estimating the hydrologic response variable.

[*Intercept* refers to a leading column of ones in the design matrix; *L_Trans* refers to the percentage of the basin classified as having low ground-water transmissivity; *H_Trans* refers to the percentage of the basin classified as having high ground-water transmissivity; *Forest* refers to the percentage of the basin where land cover is classified as forest; *Precip* refers to the normal annual precipitation for the period 1971–2000, in inches; and *A_Soils* and *D_Soils* refer to the percentage of the basin classified in the A and D hydrologic soil groups, respectively]

Index *i*	Hydrologic characteristic	Parameter estimate β_i	Standard error of the parameter estimate	Student's *t* statistic	*p*-value
0	Intercept	-0.541982	0.1910	-2.838	0.0052
1	L_Trans	-.00136258	.0005397	-2.524	.0127
2	H_Trans	.00204796	.00051078	4.010	<.0001
3	Forest	.00402190	.0005452	7.377	<.0001
4	Precip	.0236424	.005778	4.092	<.0001
5	A_Soils	.00225536	.0007683	2.935	.0039
6	D_Soils	.00162107	.001136	1.427	.1557

Equation for predicting the hydrologic response variable:

$$R\hat{I}Y_{50} = \beta_0 + \beta_1 \cdot L\_Trans + \beta_2 \cdot H\_Trans + \beta_3 \cdot Forest + \beta_4 \cdot Precip + \beta_5 \cdot A\_Soils + \beta_6 \cdot D\_Soils$$

Table 3. Cross-tabulation of land use-land cover areas with hydrologic soil groups for land areas within Michigan.

Land use/ land cover	Hydrologic soil group					
	A	B	C	D	No data[1]	Percent
	Percentages of soil group by land use/land cover					
Urban	41.4	42.3	10.8	5.3	0.1	100
Agriculture	16.7	54.5	24.7	4.1	0.0	100
Range land	39.4	31.7	10.0	18.9	0.0	100
Forest	38.9	24.5	7.7	28.4	0.6	100
Water	3.9	5.1	0.6	89.3	1.1	100
Wetland	16.8	10.1	3.6	68.7	0.7	100
Barren	48.6	4.2	4.5	22.7	20.0	100
No data[1]	0.0	0.0	0.0	0.1	99.9	100
	Percentages of land use/land cover by soil group					
Urban	8.5	7.9	5.4	1.4	0.0	--
Agriculture	15.9	47.3	57.1	5.2	0.0	--
Range land	10.3	7.5	6.3	6.6	0.0	--
Forest	61.9	35.4	29.6	60.4	0.1	--
Water	0.3	0.3	0.1	8.8	0.0	--
Wetland	2.8	1.5	1.5	15.5	0.0	--
Barren	0.2	0.0	0.0	0.1	0.0	--
No data	0.1	0.0	0.0	1.8	99.9	--
Percent	100	100	100	100	100	--

[1] “No data” indicates that a cell in the Michigan Resource Information System coverage was coded as -9999. The no-data codes typically represented areas outside the land areas in Michigan. Generally, 99.9 percent of the time, a no-data code for land use-land cover corresponded to a no-data code for hydrologic soil group. Occasionally, no-data codes did not match between coverages.

Table 4. Lower triangular elements of the diagonally symmetric correlation matrix among parameters of selected explanatory variables and the square root of median water yield for the summer month of lowest flow in Michigan.

[*Intercept* refers to a leading column of 1’s in the design matrix; *L_Trans* refers to the percentage of the basin classified as having low ground-water transmissivity; *H_Trans* refers to the percentage of the basin classified as having high ground-water transmissivity; *Forest* refers to the percentage of the basin where land cover is classified as forest; *Precip* refers to the normal annual precipitation for the period 1971–2000, in inches; and *A_Soils* and *D_Soils* refer to the percentage of the basin classified in the A and D hydrologic soil groups, respectively]

Parameter	*Intercept*	*L_Trans*	*H_Trans*	*Forest*	*Precip*	*A_Soils*	*D_Soils*
Intercept	1	.	.	.	.	.	.
L_Trans	-0.1252	1	.	.	.	.	.
H_Trans	.2255	0.5075	1	.	.	.	.
Forest	.1065	-.0545	0.2581	1	.	.	.
Precip	-.9881	.0432	-.3134	-0.1215	1	.	.
A_Soils	-.3706	-.0988	-.6260	-.6398	0.3780	1	.
D_Soils	-.2477	-.0231	-.2292	-.5779	.1899	0.4752	1

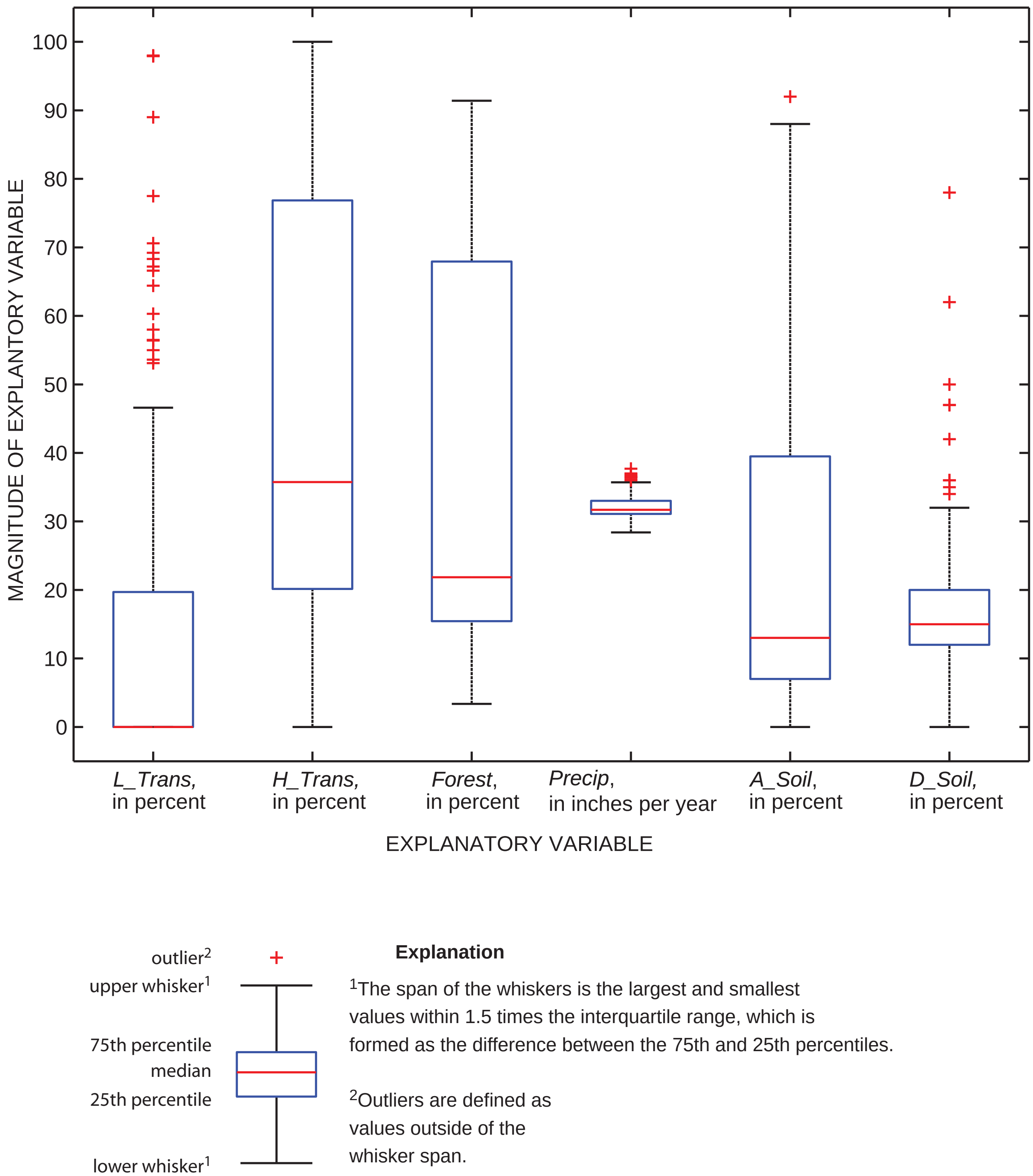


Figure 13. Distribution of explanatory variables selected for the regression model.

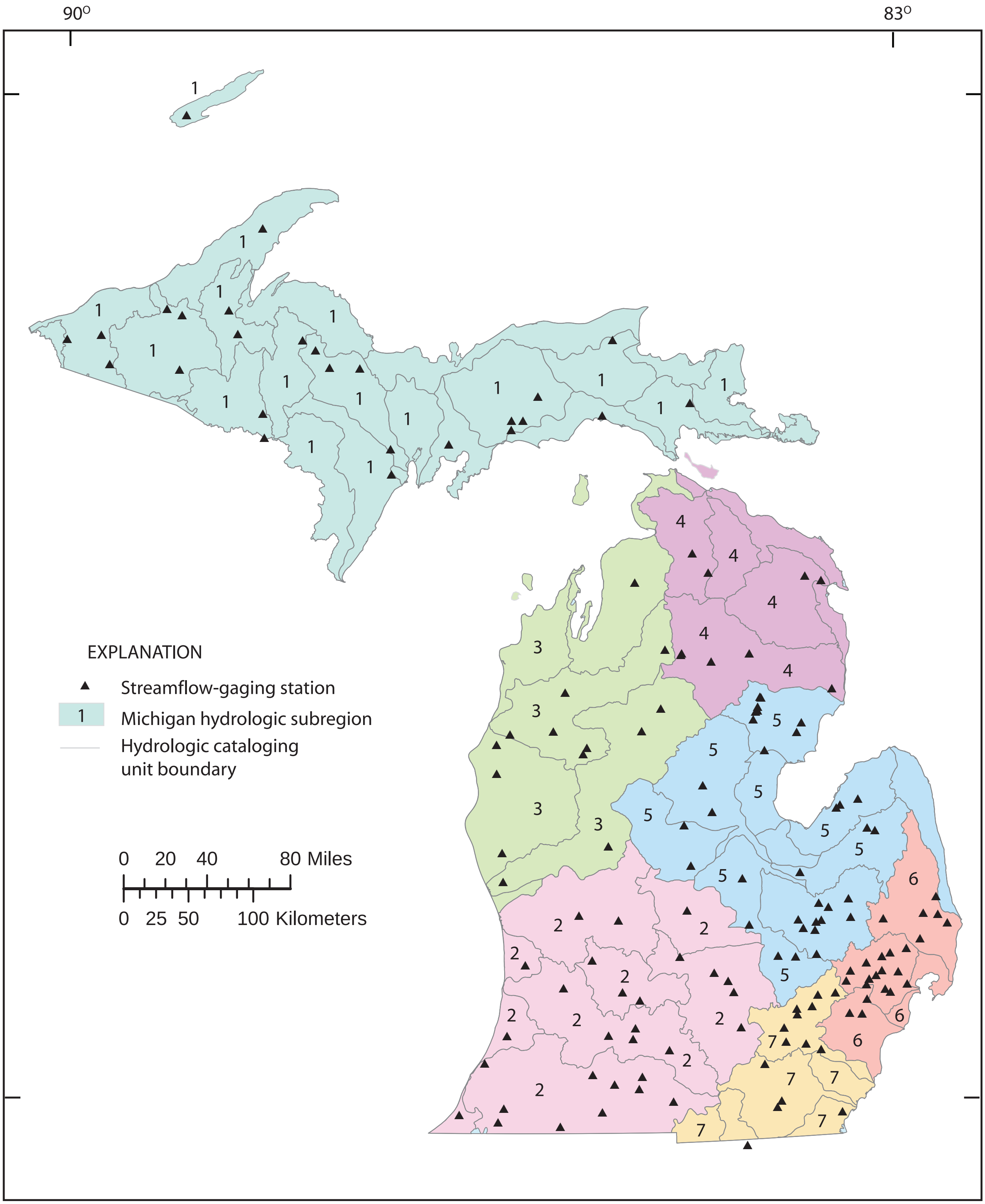
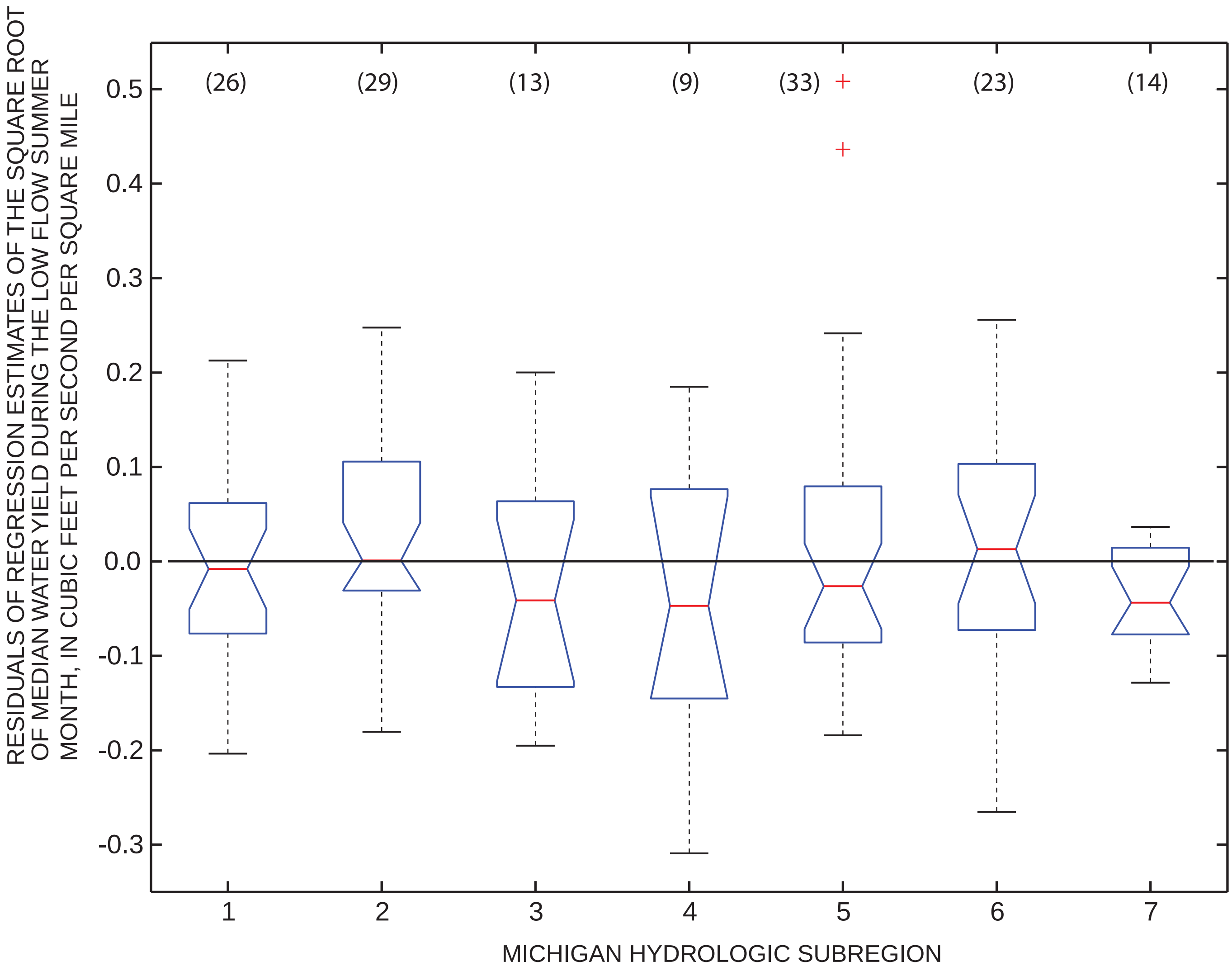
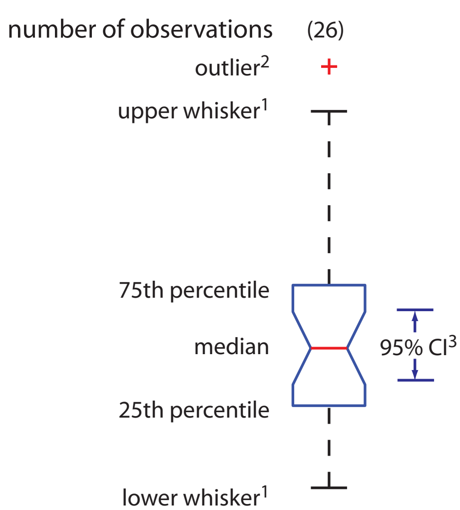


Figure 14. Hydrologic subregions used in the analysis of the spatial distribution of regression-model error.





Explanation

[1]The span of the whiskers is the largest and smmallest values within 1.5 times the interquartile range, which is formed as the difference between the 75th and 25th percentiles.

[2]Outliers are defined as values outside of the whisker span.

[3]The notch in the boxplot spans the 95% confidence interval (CI) about the median.

Figure 15. Regional distribution of regression-model errors for estimating median water yield during the summer month of minimum flow.

Computation of the Index Flow

The following sections describe computation of the index flow, $\hat{I}Q_{50}$, which involves squaring the hydrologic response variable (the estimated square root of the index water yield), symbolized as $R\hat{I}Y_{50} = \hat{I}Y_{50}$, and multiplying by the corresponding drainage area. Using $\hat{I}Y_{50}$, the assumption of linearity between drainage area and index flow is evaluated. Then, the match between the index flows determined from the analysis of streamflow-gaging station records and index flows computed on the basis of regression estimates are compared. Finally, an example is provided for computing the index flow, $\hat{I}Q_{50}$, given values for selected explanatory variables, and the upper and lower prediction limits, $\hat{U}PL_{1-\alpha/2}$ and $\hat{L}PL_{1-\alpha/2}$.

Index Water Yield and Flow

In developing a regression equation for estimating the (square root of) index water yield, a linear relation was assumed between the index flow and corresponding drainage area. In particular, the drainage area raised to the first power was assumed to be proportional to flow.

Two tests were done to evaluate the plausibility of this assumption. In the first test for unbiasedness, the estimated index flow was computed as $\hat{I}Q_{50} = \hat{I}Y_{50} \cdot DA$ and a residual series as $\xi_1 = \tilde{I}Q_{50} - \hat{I}Q_{50}$. Because ξ_1 was not normally distributed, the nonparametric two-sided sign test (Conover, 1980) was applied under the null hypothesis that the median residual ξ_1 did not differ significantly from zero. The resulting p-value was 0.4095, providing no statistical evidence to reject $\hat{I}Q_{50}$ as an unbiased estimator.

Secondly, the unbiasedness of $\hat{I}Q_{50}$ and the linearity of relation between drainage area and the index flow were tested. In this case, the form of the model evaluated was

$$\tilde{I}Q_{50} = \beta_0 + \hat{I}Y_{50} \cdot DA^{\beta_1} + \xi_2$$

where it is assumed that the estimated value of β_0, $\hat{\beta}_0$, was not significantly different from zero and that the estimated value of β_1, $\hat{\beta}_1$, was not significantly different from 1. Nonlinear estimation of the above equation resulted in parameter estimates of $\hat{\beta}_0 = -2.2913$ with an approximate standard error of $\check{s}_{\hat{\beta}_0} = 5.8644$ and $\hat{\beta}_1 = 1.0093$ with an approximate standard error of $\check{s}_{\hat{\beta}_1} = 0.00322$. Again, because ξ_2 was not normally distributed, the conventional interpretation that rejection of the null hypothesis at a probability level α required that the interval $[\hat{\beta}_0 - t_{1-\alpha/2.147-9} \cdot \check{s}_{\hat{\beta}_0}, \hat{\beta}_0 + t_{1-\alpha/2.147-9} \cdot \check{s}_{\hat{\beta}_0}]$ not include zero and the interval $[\hat{\beta}_1 - t_{1-\alpha/2.147-9} \cdot \check{s}_{\hat{\beta}_1}, \hat{\beta}_1 + t_{1-\alpha/2.147-9} \cdot \check{s}_{\hat{\beta}_1}]$ not include 1 could not be strictly applied. The value of $t_{1-\alpha/2,147-9}$ indicates the inverse of the Student's t cumulative distribution function with a specified probability level α, commonly 0.05, and degrees of freedom 147-9, reflecting the total number of observations used to develop the regression equation and the total number of parameters used in estimating the square root of the yield and the relation between the yield and flow. These intervals provide no evidence, however, to indicate that $\hat{\beta}_0$ is statistically different from 0 or that $\hat{\beta}_1$ differs substantially (more than 1 percent) from its hypothesized value of 1. The approximate correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ was –0.4202, which does not indicate significant ambiguity between the two parameter estimates. Other nonlinear models investigated, including $\tilde{I}Q_{50} = \beta_0 \cdot \hat{I}Y_{50} \cdot DA^{\beta_1} + \xi_3$ and $\tilde{I}Q_{50} = \beta_0 + \beta_1 \cdot \hat{I}Y_{50} \cdot DA^{\beta_2} + \xi_4$, resulted in one or more parameters having negative correlations less than –0.997, making the interpretations of individual parameter estimates unreliable. Therefore, $\hat{I}Q_{50}$ is considered an unbiased and physically plausible estimator of IQ_{50}.

Comparison of Index Flows

Index flows indicated by analysis of gaging-station records $\tilde{I}Q_{50}$ and computed on the basis of the statewide regression equation $\hat{I}Q_{50}$ were compared for 147 streamflow-gaging stations used in the development of the regression model. The resulting Spearman (rank) correlation was 0.97, and the corresponding coefficient of determination R_S^2 was 0.9351. Although data for the two sites where index flows determined on the basis of streamflow-gaging station records equaled zero could not be displayed, a logarithmic plot of the measured and computed index flows shows a close match about the line of agreement (fig. 16).

Example Computation

Following is an example computation to illustrate the procedure for estimating the index flow and computing the corresponding estimation interval. Station 04035000 is used to illustrate the computation. From table 1–2 Appendix 1, the explanatory variables for station 04035000 are *L_Trans* = 27.0 percent, *H_Trans* = 23.9 percent, *Forest* = 89.0 percent, *Precip* = 32.2 in., *A_Soils* = 14.0 percent, and *D_Soils* = 47.0 percent.

As an alternative to the matrix notation $x_0 \cdot \beta_{ols}$ used previously, the regression equation for predicting the water yield response can be written as

$$R\hat{I}Y_{50} = \beta_0 + \beta_1 \cdot L\_Trans + \beta_2 \cdot H\_Trans + \beta_3 \cdot Forest + \beta_4 \cdot Precip + \beta_5 \cdot A\_Soils + \beta_6 \cdot D\_Soils$$

Substituting the ordinary least square parameter estimates from table 4 for the beta coefficients and values of the explanatory variables for station 04035000, the regression equation can be written

$$R\hat{I}Y_{50} = -0.54198 + (-0.0013626 \cdot 27.0) + (0.0020480 \cdot 23.9) + (0.0040219 \cdot 89.0) + \cdots + (0.023642 \cdot 32.2) + (0.0022554 \cdot 14.0) + (0.0016211 \cdot 47.0)$$

At station 04035000, the drainage area is 273 mi², so the estimate of index flow, $\hat{I}Q_{50} = R\hat{I}Y_{50}^2 \cdot DA = 0.6972^2 \cdot 273 = 132.7$ ft³/s, in this case compares closely with the measured value of $\tilde{I}Q_{50} = 134$ ft³/s.

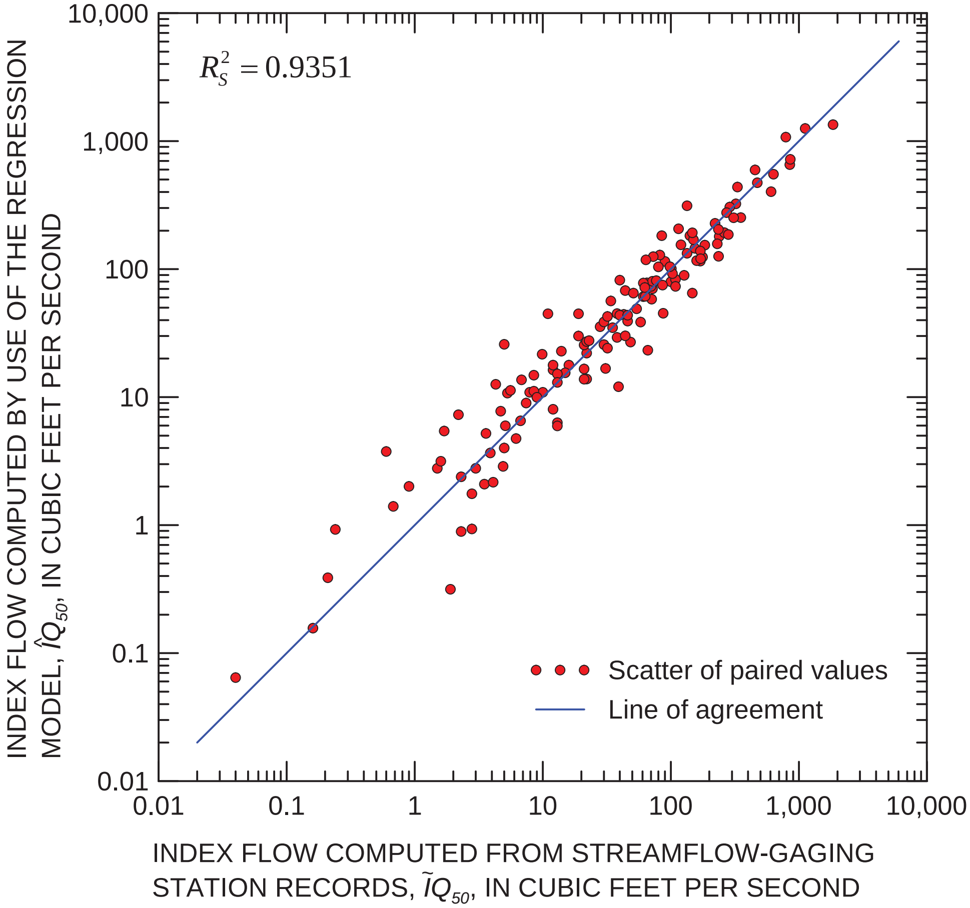


Figure 16. Relation between measured and computed index flows for selected streamflow-gaging stations in Michigan [R_S^2, the Spearman coefficient of determination].

The interval formed by the range of the lower and upper prediction limits is a measure of the uncertainty of the hydrologic response estimate. In particular, the prediction interval is likely to contain IQ_{50} with probability 1 minus alpha $(1-\alpha)$. The interval width will be smaller for a basin whose hydrologic characteristics are similar to those used to develop the regression than for basins whose characteristics are dissimilar.

The computation of a lower estimation limit about $R\hat{I}Y_{50}$ for $\alpha = 0.2$ will be shown with data from the site 04035000, as above. With this alpha value, the lower prediction limit will be less than IQ_{50} at a new site about 90 percent of the time. The lower prediction limit is computed as

$$LPL_{\alpha/2} = x_0 \cdot \beta_{ols} - t_{140,1-0.2/2} \cdot \sqrt{s^2\left(1 + x_0\left(X^{'}X\right)^{-1} x_0^{'}\right)}$$

where $t_{140,1-0.2/2} = 1.2876$, $s^2 = MS_E = 0.015320$, and $(X'X)^{-1}$ is from the entries in table 5, results in a lower 90-percent prediction limit of 0.5328. Similar computations resulted in an upper prediction limit of 0.8615 for $R\hat{I}Y_{50}$. The 90-percent prediction interval about $R\hat{I}Y_{50}$ corresponds to a 90-percent prediction interval about $\hat{I}Q_{50}$ of $[0.5238^2 \cdot 273.2, 0.8615^2 \cdot 273.2]$. Thus, the probability that IQ_{50} is contained within the estimation interval from [77.5, 202.6] ft³/s is 80 percent, or $\text{Prob}[77.5 < IQ_{50} < 202.6] = 0.8$.

Summary

In 2006, Michigan enacted legislation to prevent new large-capacity withdrawals from causing an adverse impact on a stream's ability to support characteristic fish populations. The median streamflow for the summer month of lowest flow was selected as the index flow against which possible withdrawals would be assessed. This report describes a method to predict the index flow at ungaged stream sites in Michigan. This study was conducted by the U.S. Geological Survey (USGS) in cooperation with the Michigan Department of Environmental Quality and the Michigan Department of Natural Resources.

A set of 147 USGS continuous streamflow-gaging stations were selected from among stations operated in Michi-

Table 5. The inverse of the X′X matrix needed to compute prediction limits.

[*Intercept* refers to a leading column of ones in the design matrix; *L_Trans* refers to the percentage of the basin classified as having low ground-water transmissivity; *H_Trans* refers to the percentage of the basin classified as having high ground-water transmissivity; *Forest* refers to the percentage of the basin where land cover is classified as forest; *Precip* refers to the normal annual precipitation for the period 1971–2000, in inches; and *A_Soils* and *D_Soils* refer to the percentage of the basin classified in the A and D hydrologic soil groups, respectively]

Explanatory variables in the regression model						
Intercept	***L_Trans***	***H_Trans***	***Forest***	***Precip***	***A_Soils***	***D_Soils***
2.38035E+00	-8.42611E-04	1.43560E-03	7.23794E-04	-7.11625E-02	-3.54982E-03	-3.50692E-03
-8.42611E-04	1.90162E-05	9.13204E-06	-1.04607E-06	8.79202E-06	-2.67491E-06	-9.25476E-07
1.43560E-03	9.13204E-06	1.70298E-05	4.69128E-06	-6.03695E-05	-1.60371E-05	-8.67801E-06
7.23794E-04	-1.04607E-06	4.69128E-06	1.93996E-05	-2.49764E-05	-1.74926E-05	-2.33578E-05
-7.11625E-02	8.79202E-06	-6.03695E-05	-2.49764E-05	2.17903E-03	1.09528E-04	8.13390E-05
-3.54982E-03	-2.67491E-06	-1.60371E-05	-1.74926E-05	1.09528E-04	3.85354E-05	2.70690E-05
-3.50692E-03	-9.25476E-07	-8.67801E-06	-2.33578E-05	8.13390E-05	2.70690E-05	8.42089E-05

gan for 10 or more years that were thought to represent the natural response of streamflow to precipitation. In particular, stations where median low flows were thought to have been appreciably affected by regulation or water withdrawals, augmentations, or diversions were excluded from the regression analysis. Of the 147 selected stations, minimum median flows occurred in July at 5 stations, in August at 92 stations, and in September at 50 stations. Index flows ranged from 0 to 1,850 ft^3/s. Index water yields, which were computed by dividing index flows by the corresponding drainage areas upstream from the stream measurement sites, ranged from 0 to 1.309 $ft^3/s\text{-}mi^2$. A square-root transformation was applied to the index water yields so that the transformed values were approximately normally distributed.

A multiple linear regression equation was developed to predict the square root of the index water yield at ungaged sites using selected basin and climatic characteristics as explanatory variables. Selected variables included percentages of land area underlain by low and high aquifer transmissivity, percentage of forest cover, normal annual precipitation, and percentages of land cover associated with hydrologic soil groups A and D (highly and poorly permeable soils, respectively). The regression model explains about 70.8 percent of the variability in the hydrologic response variable, which was the square root of the index water yield. No spatial bias in the regression estimates was detected among seven hydrologic subregions spanning Michigan. Therefore, the single regression equation developed in this report is appropriate for statewide application.

Index flows can be predicted at ungaged sites by squaring the predicted regression response and multiplying the result by the corresponding drainage area. The predicted index flow explains about 94.0 percent of the variability in index flows indicated by streamflow-gaging-station records. In addition, the report documents the technique and provides information needed to compute an interval about the predicted index flow. An example computation is provided.

Acknowledgments

The authors appreciate the valuable technical contributions the following individuals made in developing the regression model: Jon W. Allan, Manager, and Scott F. DeBoe, Senior Environmental Planner, Consumers Energy Company; Paul Seelbach, Research Director, Michigan Department of Natural Resources; and Howard W. Reeves, Research Hydrologist, U.S. Geological Survey. Arthur Cooper, Research Assistant, University of Michigan, did the GIS analysis needed to characterize the explanatory variables at the selected streamflow-gaging stations.

References Cited

Anderson, J.R., Hardy, E.E., Roach, J.T., and Witmer, R.E., 1976, A land use and land cover classification system for use with remote sensor data: U.S. Geological Survey Professional Paper 964, 41 p., accessed October 3, 2007, at *http://landcover.usgs.gov/pdf/anderson.pdf*

Conover, W.J., 1980, Practical nonparametric statistics (2d ed.): New York, Wiley, 493 p.

Burley, T.M., 1961, Land use or land utilization?: Professional Geographer, v. 13, no. 6, p. 18–20.

Clawson, Marion, and Stewart, C.L., 1965, Land use information—A critical survey of U.S. statistics including possibilities for greater uniformity: Baltimore, Md., Johns Hopkins Press for Resources for the Future, Inc., 402 p.

Holtschlag, D.J., and Croskey, H.M., 1984, Statistical models for estimating flow characteristics of Michigan streams: U.S. Geological Survey Water-Resources Investigations Report 84–4207, 80 p.

Draper, N.R., and Smith, Harry, 1966, Applied regression analysis: New York, Wiley, 407 p.

Knutilla, R.L., 1967, Flow characteristics of Michigan streams: U.S. Geological Survey Open-File Report 67–135, 55 p.

Michigan Climatological Resources Program, 2004, Michigan climatological normals, 1971–2000: East Lansing, Mich., State Climatologist's Office, Michigan State University, digital media.

Michigan Department of Environmental Quality [n.d.], Calculating curve numbers using GIS: Hydrologic Studies Unit, 10 p., accessed Oct. 3, 2007, at *http://www.michigan.gov/documents/deq/lwm-cn-calc-using-gis_202628_7.pdf*

Michigan Department of Information Technology, 2005a, Glacial deposits estimated transmissivity geographic theme: Lansing, Mich., Center for Geographic Information, accessed December 6, 2007, at *http://www.mcgi.state.mi.us/mgdl/?rel=ext&action=sext*

Michigan Department of Information Technology, 2005b, Michigan glacial landsystems coverage *by* D.P. Lusch: East Lansing, Michigan, Michigan State University, accessed on December 6, 2007, at *http://www.mcgi.state.mi.us/mgdl/?rel=ext&action=sext*

Michigan Library, 2006, Michigan in Brief—Information about the state of Michigan: Lansing, Mich., accessed December 6, 2007, at *www.michigan.gov/hal/0,1607,7-160-15481_20826_20829-56001--,00.html*

Michigan Legislature, 2006, Public Acts of 2006, Act No. 33: 93rd Legislature, Regular Session of 2006, accessed September 27, 2007, at *http://www.legislature.mi.gov/documents/2005-2006/publicact/pdf/2006-PA-0033.pdf*

Michigan Resource Information System, 1978, MIRIS Land-use/cover Polygon—Geographic data and metadata: Lansing, Michi., Michigan Department of Natural Resources, accessed December 6, 2007, at *http://www.mcgi.state.mi.us/mgdl/*

Natural Resources Conservation Service, 2004, Estimation of direct runoff from storm rainfall, Chapter 10 *of* Part 630, Hydrology, *of* National engineering handbook: Accessed December 7, 2007, at *http://policy.nrcs.usda.gov/media/pdf/H_210_630_10.pdf*

Natural Resources Conservation Service, 2007, Hydrologic soil groups, Chapter 7 *of* Part 630, Hydrology *of* National engineering handbook: Accessed December 6, 2007, at *http://policy.nrcs.usda.gov/media/pdf/H_210_630_7.pdf*

Neff, B.P., Day, S.M., Piggott, A.R., and Fuller, L.M., Base flow in the Great Lakes Basin: U.S. Geological Survey Scientific Investigations Report 2005–5217, 23 p.

Poeter, E.P., Hill, M.C., Banta, E.R., Mehl, Steffen, and Christensen, Steen, 2005, UCODE_2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation: U.S. Geological Survey Techniques and Methods 6–A11, 283 p.

Seaber, P.R., Kapinos, F.P., and Knapp, G.L., 1987, Hydrologic unit maps: U.S. Geological Survey Water-Supply Paper 2294, 63 p., accessed December 7, 2007, at *http://pubs.usgs.gov/wsp/wsp2294/*

U.S. Census Bureau, 2007, State & county QuickFacts, accessed December 6, 2007, at *http://quickfacts.census.gov/qfd/states/26000.html*

Appendix 1. Tables of streamflow-gaging station attributes, flow characteristics, and explanatory variables used in the development of the regression equation for estimating the index flow at ungaged streams in Michigan

Table 1–1. Flow, yield, and record characteristics for streamflow–gaging stations used in the regression analysis.—Continued.

[A water year is the 12–month period from October 1 to September 30. The water year is designated by the calendar year in which it ends]

U.S. Geological Survey station number	Station name	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square miles)	Minimum monthly median flow (cubic feet per second) (IQ_{50})	Minimum monthly median yield (cubic feet per second per square mile) (IY_{50})	Month of minimum flow	Years of record	Water years included in analyses
04001000	Washington Creek at Windigo, Mich.	47.92306	89.14500	13.2	2.3	0.17424	August	39	1965–2003
04031000	Black River near Bessemer, Mich.	46.51134	90.07462	200	40.0	0.20000	August	33	1955–1982, 2001–2005
04031500	Presque Isle River at Marenisco, Mich.	46.37217	89.69238	172	71.0	0.41183	August	38	1945–1982
04032000	Presque Isle River near Tula, Mich.	46.54689	89.77738	264	90.0	0.34078	August	29	1945–1973
04033000	Middle Branch Ontonagon River near Paulding, Mich.	46.35689	89.07736	162	100	0.61843	August	59	1943–1995, 2001–2005
04035000	East Br Ontonagon River near Mass, Mich.	46.68994	89.07347	273	134	0.49048	August	38	1942–1979
04040000	Ontonagon River near Rockland, Mich	46.72077	89.20709	1330	634	0.47530	August	64	1942–2005
04040500	Sturgeon River near Sidnaw, Mich.	46.58411	88.57597	169	44.0	0.26036	August	66	1913–1915, 1943–2005
04041500	Sturgeon River near Alston, Mich.	46.72632	88.66208	343	184	0.53629	August	72	1932–1940, 1943–2005
04043050	Trap Rock River near Lake Linden, Mich.	47.22854	88.38539	29.6	13.0	0.43919	August	39	1967–2005
04045500	Tahquamenon River near Paradise, Mich.	46.57501	85.26955	757	321	0.42410	August	52	1954–2005
04046000	Black River near Garnet, Mich.	46.11806	85.36537	33.5	10.0	0.29851	August	38	1952–1978, 1995–2005
04049500	Manistique River at Germfask, Mich.	46.23331	85.92791	420	238	0.56721	August	33	1938–1970
04055000	Manistique River near Blaney, Mich.	46.08609	86.05930	716	352	0.49183	August	33	1938–1970
04056000	West Branch Manistique River near Manistique, Mich.	46.08886	86.16125	326	154	0.47312	August	19	1938–1956
04056500	Manistique River near Manistique, Mich.	46.03053	86.16125	1,130	605	0.53716	August	68	1938–2005
04057510	Sturgeon River near Nahma Junction, Mich.	45.94302	86.70570	184	82.0	0.44662	August	39	1967–2005
04057800	Middle Branch Escanaba River at Humboldt, Mich.	46.49910	87.88652	45.7	16.0	0.35011	August	47	1959–2005
04058000	M Br Escanaba River near Ishpeming, Mich.	46.39438	87.75847	128	43.0	0.33515	August	22	1954–1975
04058400	Goose Lake Outlet near Sands Station, Mich.	46.39300	87.49375	36.3	12.0	0.33058	August	17	1966–1982

Table 1–1. Flow, yield, and record characteristics for streamflow–gaging stations used in the regression analysis—Continued.

[A water year is the 12–month period from October 1 to September 30. The water year is designated by the calendar year in which it ends]

U.S. Geological Survey station number	Station name	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square miles)	Minimum monthly median flow (cubic feet per second) (IQ_{50})	Minimum monthly median yield (cubic feet per second per square mile) (IY_{50})	Month of minimum flow	Years of record	Water years included in analyses
04059000	Escanaba River at Cornell, Mich.	45.90857	87.21375	871	330	0.37892	August	55	1951–2005
04059500	Ford River near Hyde, Mich.	45.75552	87.20152	444	85.0	0.19140	August	51	1955–2005
04060993	Brule River near Florence, Wisc.	45.96079	88.31597	378	236	0.62269	August	62	1944–2005
04061500	Paint River at Crystal Falls, Mich.	46.10578	88.33486	600	288	0.47976	August	52	1945–1996
04062200	Peshekee River near Champion, Mich.	46.55688	88.00263	132	28.0	0.21244	August	23	1961–1978, 2001–2005
04096015	Galien River near Sawyer, Mich.	41.87365	86.57502	80.8	21.0	0.25990	September	11	1995–2005
04096405	St. Joseph River at Burlington, Mich.	42.10282	85.04025	201	61.0	0.30348	September	43	1963–2005
04096515	South Branch Hog Creek near Allen, Mich.	41.94866	84.82774	48.7	7.9	0.16222	September	36	1970–2005
04096600	Coldwater River near Hodunk, Mich.	42.02921	85.10692	286	65.0	0.22759	September	27	1963–1989
04096900	Nottawa Creek near Athens, Mich.	42.05560	85.30832	162	72.0	0.44444	September	31	1967–1997
04097170	Portage River near Vicksburg, Mich.	42.11477	85.48555	68.2	30.0	0.43988	September	21	1946–1951, 1965–1979
04097540	Prairie River near Nottawa, Mich.	41.88838	85.40943	107	46.0	0.43152	September	43	1963–2005
04099000	St. Joseph River at Mottville, Mich.	41.80088	85.75610	1,880	850	0.45227	September	82	1924–2005
04101500	St. Joseph River at Niles, Mich.	41.82921	86.25973	3,670	1,850	0.50464	September	75	1931–2005
04101800	Dowagiac River at Sumnerville, Mich.	41.91338	86.21307	252	177	0.70378	August	45	1961–2005
04102500	Paw Paw River at Riverside, Mich.	42.18615	86.36836	390	262	0.67197	August	54	1952–2005
04102700	South Branch Black River near Bangor, Mich.	42.35420	86.18753	83.5	35.0	0.41916	September	40	1966–2005
04103010	Kalamazoo River near Marengo, Mich.	42.26171	84.85581	270	147	0.54545	September	19	1987–2005
04104945	Wanadoga Creek near Battle Creek, Mich.	42.39643	85.13166	48.3	15.0	0.31056	August	11	1995–2005
04105000	Battle Creek at Battle Creek, Mich.	42.33199	85.15416	274	70.0	0.25547	September	72	1934–2005
04105700	Augusta Creek near Augusta, Mich.	42.35337	85.35389	36.8	31.0	0.84239	August	41	1965–2005
04108600	Rabbit River near Hopkins, Mich.	42.64225	85.72197	65.1	22.0	0.33794	September	40	1966–2005
04108801	Macatawa River near Zeeland, Mich.	42.77919	86.01837	66.9	4.3	0.06428	September	45	1961–2005

Table 1–1. Flow, yield, and record characteristics for streamflow–gaging stations used in the regression analysis—Continued.

[A water year is the 12–month period from October 1 to September 30. The water year is designated by the calendar year in which it ends]

U.S. Geological Survey station number	Station name	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square miles)	Minimum monthly median flow (cubic feet per second) (IQ_{50})	Minimum monthly median yield (cubic feet per second per square mile) (IY_{50})	Month of minimum flow	Years of record	Water years included in analyses
04110000	Orchard Creek at Munith, Mich.	42.39365	84.26496	47.3	5.3	0.11205	September	13	1944–1956
04111500	Deer Creek near Dansville, Mich.	42.60837	84.32080	16.3	0.9	0.05521	September	52	1954–2005
04112000	Sloan Creek near Williamston, Mich.	42.67587	84.36386	10.4	0.2	0.02308	September	52	1954–2005
04112500	Red Cedar River at East Lansing, Mich.	42.72781	84.47775	344	38.0	0.11047	September	77	1902–1903, 1931–2005
04114498	Looking Glass River near Eagle, Mich.	42.82809	84.75943	284	40.0	0.14080	September	57	1944–1996, 2002–2005
04115000	Maple River at Maple Rapids, Mich.	43.10975	84.69305	420	30.0	0.07141	September	62	1944–2005
04116500	Flat River at Smyrna, Mich.	43.05281	85.26474	516	222	0.42998	August	36	1951–1986
04117000	Quaker Brook near Nashville, Mich.	42.56587	85.09361	7.8	2.8	0.35897	September	33	1954–1975, 1995–2005
04117500	Thornapple River near Hastings, Mich.	42.61587	85.23639	410	109	0.26553	September	61	1945–2005
04118000	Thornapple River near Caledonia, Mich.	42.81114	85.48335	795	281	0.35337	September	41	1952–1981, 1984–1994
04118500	Rogue River near Rockford, Mich.	43.08225	85.59086	257	127	0.49378	September	50	1952–1982, 1988–2005
04121000	Muskegon River near Merritt, Mich.	44.33557	84.89003	352	115	0.32689	August	27	1947–1973
04121300	Clam River at Vogel Center, Mich.	44.20057	85.05281	239	73.0	0.30506	August	40	1966–2005
04121900	Little Muskegon River near Morley, Mich.	43.50253	85.34254	136	72.0	0.53137	July	30	1967–1996
04122100	Bear Creek near Muskegon, Mich.	43.28863	86.22284	16.7	5.0	0.29940	August	40	1966–2005
04122200	White River near Whitehall, Mich.	43.46418	86.23257	404	273	0.67491	August	49	1957–2005
04122500	Pere Marquette River at Scottville, Mich.	43.94501	86.27869	689	455	0.65999	August	67	1939–2005
04123000	Big Sable River near Freesoil, Mich.	44.12028	86.28008	115	101	0.87826	August	32	1942–1973
04123500	Manistee River near Grayling, Mich.	44.69307	84.84726	132	170	1.28496	August	31	1943–1973
04124000	Manistee River near Sherman, Mich.	44.43639	85.69868	865	856	0.98971	August	86	1903–1916, 1934–2005

Table 1–1. Flow, yield, and record characteristics for streamflow–gaging stations used in the regression analysis—Continued.

[A water year is the 12–month period from October 1 to September 30. The water year is designated by the calendar year in which it ends]

U.S. Geological Survey station number	Station name	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square miles)	Minimum monthly median flow (cubic feet per second) (IQ_{50})	Minimum monthly median yield (cubic feet per second per square mile) (IY_{50})	Month of minimum flow	Years of record	Water years included in analyses
04124500	East Branch Pine River near Tustin, Mich.	44.10251	85.51728	58.9	9.9	0.16808	August	26	1952–1963, 1992–2005
04125000	Pine River near Leroy, Mich.	44.06279	85.54894	130	51.0	0.39140	August	12	1952–1963
04125500	Pine River near Hoxeyville, Mich.	44.20306	85.79951	254	230	0.90658	August	31	1952–1982
04126200	Little Manistee River near Freesoil, Mich.	44.18362	86.16758	185	141	0.76381	August	19	1957–1975
04127918	Pine River near Rudyard, Mich.	46.18585	84.59783	202	77.0	0.38138	August	34	1972–2005
04127997	Sturgeon River at Wolverine, Mich.	45.29890	84.61114	181	170	0.93975	August	64	1942–2005
04128990	Pigeon River near Vanderbilt, Mich.	45.15668	84.46669	57.7	58.0	1.00520	July	55	1951–2005
04133501	Thunder Bay River at Herron Road near Bolton, Mich.	45.12446	83.64721	586	309	0.52739	September	40	1945–1980, 2002–2005
04135000	Thunder Bay River near Alpena, Mich.	45.09418	83.49970	1,240	474	0.38288	August	22	1901–1908, 1980–1993
04135500	Au Sable River at Grayling, Mich.	44.65974	84.71253	96.6	61.0	0.63147	August	51	1943–1993
04135600	East Branch Au Sable River at Grayling, Mich.	44.66890	84.70558	71.2	34.0	0.47753	August	27	1958–1984
04135700	South Branch Au Sable River near Luzerne, Mich.	44.61474	84.45557	391	134	0.34236	August	38	1967–1989, 1991–2005
04136500	Au Sable River at Mio, Mich.	44.66001	84.13112	1,360	788	0.57894	August	54	1952–2005
04137500	Au Sable R near Au Sable, Mich.	44.43640	83.43386	1,740	1,120	0.64383	September	19	1987–2005
04138000	East Branch Au Gres River at McIvor, Mich.	44.23252	83.70082	89.9	38.0	0.42269	August	23	1951–1973
04138500	Au Gres River near National City, Mich.	44.17613	83.74248	151	21.0	0.13871	August	31	1951–1981
04139000	Houghton Creek near Lupton, Mich.	44.39585	84.04722	29.8	39.0	1.30872	August	23	1950–1972
04139500	Rifle River at "The Ranch" near Lupton, Mich.	44.39335	84.03833	57.4	66.0	1.14983	August	22	1950–1971
04140000	Prior Creek near Selkirk, Mich.	44.33502	84.06833	21.0	6.7	0.31905	August	22	1951–1972
04140500	Rifle River at Selkirk, Mich.	44.31335	84.06944	116	87.0	0.74742	August	32	1951–1982
04141000	South Branch Shepards Creek near Selkirk, Mich.	44.30780	84.08694	1.1	0.0	0.03636	July	27	1952–1978

Table 1–1. Flow, yield, and record characteristics for streamflow–gaging stations used in the regression analysis—Continued.

[A water year is the 12–month period from October 1 to September 30. The water year is designated by the calendar year in which it ends]

U.S. Geological Survey station number	Station name	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square miles)	Minimum monthly median flow (cubic feet per second) (IQ_{50})	Minimum monthly median yield (cubic feet per second per square mile) (IY_{50})	Month of minimum flow	Years of record	Water years included in analyses
04141500	West Branch Rifle River near Selkirk, Mich.	44.26113	84.10916	65.4	32.0	0.48930	July	12	1952–1963
04142000	Rifle River near Sterling, Mich.	44.07252	84.01999	333	159	0.47791	August	69	1937–2005
04143900	Shiawassee River at Linden, Mich.	42.81586	83.80190	81.9	22.0	0.26862	August	30	1968–1994, 2001–2003
04144000	Shiawassee River at Byron, Mich.	42.82364	83.94579	363	71.0	0.19543	September	36	1948–1983
04144500	Shiawassee River at Owosso, Mich.	43.01503	84.18108	530	86.0	0.16214	September	75	1931–2005
04145500	Bad River near Brant, Mich.	43.29669	84.22915	89.9	0.6	0.00667	August	11	1949–1959
04146000	Farmers Creek near Lapeer, Mich.	43.04475	83.33717	51.1	5.1	0.09980	August	73	1933–2005
04146063	South Branch Flint River near Columbiaville, Mich.	43.15947	83.35078	211	48.5	0.23029	August	26	1980–2005
04147500	Flint River near Otisville, Mich.	43.11114	83.51940	526	109	0.20715	August	52	1953–1989, 1991–2005
04147990	Butternut Creek near Genesee, Mich.	43.13586	83.59912	34.8	3.6	0.10345	August	14	1970–1983
04148140	Kearsley Creek near Davison, Mich.	43.03364	83.58134	99.7	13.0	0.13039	August	40	1966–2005
04148160	Gilkey Creek near Flint, Mich.	43.02419	83.62551	6.9	0.2	0.02319	August	14	1970–1983
04148200	Swartz Creek near Holly, Mich.	42.82753	83.62828	12.1	1.5	0.12397	September	20	1956–1975
04148300	Swartz Creek at Flint, Mich.	42.98781	83.73246	114	5.6	0.04904	August	14	1970–1983
04148440	Thread Creek near Flint, Mich.	42.97503	83.63579	54.4	4.7	0.08640	August	14	1970–1983
04148500	Flint River near Flint, Mich.	43.03892	83.77163	960	171	0.17805	August	73	1933–2005
04150000	South Branch Cass River near Cass City, Mich.	43.56696	83.11189	239	5.0	0.02090	September	32	1949–1980
04150500	Cass River at Cass City, Mich.	43.58419	83.17606	363	11.0	0.03034	September	55	1948–1997, 2001–2005
04151500	Cass River at Frankenmuth, Mich.	43.32780	83.74802	842	64.0	0.07597	September	69	1935–1936, 1939–2005
04152238	South Branch Tobacco River near Beaverton, Mich.	43.86697	84.54529	152	63.0	0.41366	September	19	1987–2005

Table 1–1. Flow, yield, and record characteristics for streamflow–gaging stations used in the regression analysis—Continued.

[A water year is the 12–month period from October 1 to September 30. The water year is designated by the calendar year in which it ends]

U.S. Geological Survey station number	Station name	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square miles)	Minimum monthly median flow (cubic feet per second) (IQ_{50})	Minimum monthly median yield (cubic feet per second per square mile) (IY_{50})	Month of minimum flow	Years of record	Water years included in analyses
04153500	Salt River near North Bradley, Mich.	43.70281	84.47056	145	8.5	0.05874	August	38	1934–1971
04154000	Chippewa River near Mount Pleasant, Mich.	43.62558	84.70779	409	150	0.36702	August	73	1933–2005
04155000	Pine River at Alma, Mich.	43.37948	84.65556	309	80.0	0.25882	August	75	1931–2005
04157500	State Drain near Sebewaing, Mich.	43.71196	83.42774	67.3	0	0.00000	August	15	1940–1954
04158000	Columbia Drain near Sebewaing, Mich.	43.72724	83.39607	33.9	0	0.00000	August	18	1940–1954, 1988–1990
04158500	Pigeon River near Owendale, Mich.	43.76363	83.24606	53.3	3.0	0.05629	September	30	1953–1982
04159492	Black River near Jeddo, Mich.	43.15253	82.62409	479	22.0	0.04589	September	62	1944–2005
04159900	Mill Creek near Avoca, Mich.	43.05447	82.73465	169	6.8	0.04033	September	31	1963–1975, 1988–2005
04160000	Mill Creek near Abbottsford, Mich.	43.04503	82.61381	184	8.5	0.04620	August	18	1947–1964
04160050	Black River near Port Huron, Mich.	42.99003	82.53770	683	19.0	0.02783	September	11	1933–1943
04160570	North Branch Belle River at Imlay City, Mich.	43.03031	83.06716	16.1	2.3	0.14286	August	36	1966–2001
04160600	Belle River at Memphis, Mich.	42.90086	82.76909	151	13.0	0.08587	September	43	1963–2005
04160800	Sashabaw Creek near Drayton Plains, Mich.	42.72003	83.35355	21.0	2.2	0.10476	September	46	1960–2005
04160900	Clinton River near Drayton Plains, Mich.	42.66031	83.39022	78.5	14.0	0.17834	August	46	1960–2005
04161000	Clinton River at Auburn Hills, Mich.	42.63337	83.22438	123	44	0.35685	August	34	1935–1938, 1940, 1957–1982, 2001–2002, 2004–2005
04161100	Galloway Creek near Auburn Heights, Mich.	42.66725	83.20049	17.4	1.6	0.09195	August	32	1960–1991
04161500	Paint Creek near Lake Orion, Mich.	42.76753	83.21994	39.8	9.0	0.22613	August	23	1956–1975, 1989–1991
04161540	Paint Creek at Rochester, Mich.	42.68837	83.14299	71.8	21.0	0.29248	August	46	1960–2005
04161580	Stony Creek near Romeo, Mich.	42.80086	83.09021	23.8	3.9	0.16387	August	41	1965–2005
04161800	Stony Creek near Washington, Mich.	42.71531	83.09188	69.1	13.0	0.18813	August	48	1958–2005

Table 1–1. Flow, yield, and record characteristics for streamflow–gaging stations used in the regression analysis—Continued.

[A water year is the 12–month period from October 1 to September 30. The water year is designated by the calendar year in which it ends]

U.S. Geological Survey station number	Station name	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square miles)	Minimum monthly median flow (cubic feet per second) (IQ_{50})	Minimum monthly median yield (cubic feet per second per square mile) (IY_{50})	Month of minimum flow	Years of record	Water years included in analyses
04163400	Plum Brook at Utica, Mich.	42.60142	83.07409	16.6	2.8	0.16867	August	40	1965–1998, 2000–2005
04163500	Plum Brook near Utica, Mich.	42.58365	83.03048	23.8	0.7	0.02857	July	13	1954–1966
04164100	East Pond Creek at Romeo, Mich.	42.82253	83.02021	20.8	4.9	0.23558	August	47	1959–2005
04164300	East Branch Coon Creek at Armada, Mich.	42.84586	82.88493	12.8	0.2	0.01641	August	47	1959–2005
04164500	North Branch Clinton R near Mount Clemens, Mich.	42.62920	82.88881	198	12.0	0.06073	September	59	1947–2005
04164800	Middle Branch Clinton River at Macomb, Mich.	42.70642	82.95909	41.2	3.5	0.08495	September	19	1963–1968, 1970–1982
04166000	River Rouge at Birmingham, Mich.	42.54587	83.22354	36.7	6.2	0.16894	September	56	1950–2005
04166200	Evans Ditch at Southfield, Mich.	42.45781	83.26743	10.2	1.9	0.18627	September	48	1958–2005
04166300	Upper River Rouge at Farmington, Mich.	42.46448	83.36966	17.6	4.1	0.23295	September	48	1958–2005
04169500	Huron River at Commerce, Mich.	42.59031	83.48466	49.9	12.0	0.24048	August	30	1946–1975
04170000	Huron River at Milford, Mich.	42.57892	83.62661	139	46.0	0.33141	August	57	1949–2005
04170500	Huron River near New Hudson, Mich.	42.51253	83.67633	155	54.0	0.34771	August	57	1949–2005
04171500	South Ore Creek near Brighton, Mich.	42.49781	83.80244	33.3	7.4	0.22222	September	18	1951–1968
04172000	Huron River near Hamburg, Mich.	42.46531	83.79994	320	103	0.32177	September	54	1952–2005
04173000	Huron River near Dexter, Mich.	42.38615	83.91106	538	120	0.22326	August	29	1946–1972, 1976–1977
04173500	Mill Creek near Dexter, Mich.	42.30004	83.89856	131	23.0	0.17598	September	42	1952–1982, 1995–2005
04174500	Huron River at Ann Arbor, Mich.	42.28615	83.73327	747	147	0.19684	August	91	1915–2005
04174800	Huron River at Ypsilanti, Mich.	42.24921	83.61244	817	235	0.28750	August	16	1974–1984, 1990–1994
04175600	River Raisin near Manchester, Mich.	42.16809	84.07606	128	32.0	0.25059	August	33	1970–1981, 1985–2005

Table 1–1. Flow, yield, and record characteristics for streamflow–gaging stations used in the regression analysis—Continued.

[A water year is the 12–month period from October 1 to September 30. The water year is designated by the calendar year in which it ends]

U.S. Geological Survey station number	Station name	Latitude (decimal degrees)	Longitude (decimal degrees)	Drainage area (square miles)	Minimum monthly median flow (cubic feet per second) (IQ_{50})	Minimum monthly median yield (cubic feet per second per square mile) (IY_{50})	Month of minimum flow	Years of record	Water years included in analyses
04175700	River Raisin near Tecumseh, Mich.	41.94310	83.94578	266	62.5	0.23532	August	24	1957–1980
04176000	River Raisin near Adrian, Mich.	41.90421	83.98050	460	98.5	0.21422	September	46	1954–1978, 1985–2005
04176605	Otter Creek at Lasalle, Mich.	41.86699	83.45354	63.7	1.7	0.02669	September	18	1988–2005
04184500	Bean Creek at Powers, Ohio	41.67755	84.23217	205	19.0	0.09255	September	65	1941–2005

Table 1 –2. Values of selected explanatory variables used in the development of the regression equation for estimating the index flow.

U.S. Geological Survey station number	Drainage area (square miles)	Percent of basin with low ground-water transmissivity (*L_Trans*)	Percent of basin with high ground-water transmissivity (*H_Trans*)	Percent of basin with forest cover (*Forest*)	Normal annual precipitation for 1971–2000 (inches) (*Precip*)	Percent of basin with A hydrologic soil group (*A_Soils*)	Percent of basin with D hydrologic soil group (*D_Soils*)	Michigan hydrologic subregion
04001000	13.2	98.0	0.0	91.4	31.0	0.0	0.0	1
04031000	200	7.7	6.8	85.5	34.7	.0	9.0	1
04031500	172	.0	9.7	86.6	33.8	.0	10.0	1
04032000	264	0.8	11.3	90.3	34.2	.0	5.0	1
04033000	162	.0	21.0	86.6	31.7	1.0	62.0	1
04035000	273	27.0	23.9	89.0	32.2	14.0	47.0	1
04040000	1330	34.1	8.9	84.2	32.6	25.0	29.0	1
04040500	169	32.4	14.1	84.7	33.0	7.0	34.0	1
04041500	343	17.0	20.2	85.1	32.7	9.0	36.0	1
04043050	29.6	36.2	3.6	59.5	31.4	19.0	13.0	1
04045500	757	55.0	30.9	79.0	31.2	32.0	50.0	1
04046000	33.5	27.7	.0	75.4	30.7	32.0	30.0	1
04049500	420	31.5	48.9	68.4	30.5	54.0	14.0	1
04055000	716	46.6	36.4	60.8	30.5	62.0	12.0	1
04056000	326	60.3	29.9	77.9	31.6	72.0	5.0	1
04056500	1,130	53.4	32.1	66.4	30.8	32.0	49.0	1
04057510	184	12.7	85.5	79.8	31.4	45.0	35.0	1
04057800	45.7	56.5	24.3	80.9	33.0	5.0	47.0	1
04058000	128	68.3	20.7	77.1	33.0	6.0	47.0	1
04058400	36.3	66.6	33.4	78.7	32.5	53.0	19.0	1
04059000	871	22.4	29.7	81.7	32.4	30.0	36.0	1
04059500	444	.1	5.4	86.1	31.5	10.0	36.0	1
04060993	378	.0	29.7	72.6	31.2	3.0	13.0	1
04061500	600	1.8	19.7	84.5	31.5	3.0	78.0	1
04062200	132	97.9	.0	84.9	33.1	1.0	42.0	1
04096015	80.8	15.4	8.3	17.0	37.6	11.0	11.0	2
04096405	201	.0	69.6	18.5	35.3	8.0	14.0	2
04096515	48.7	.0	32.5	15.4	35.7	11.0	11.0	2
04096600	286	.0	57.2	16.0	36.0	4.0	14.0	2
04096900	162	.0	96.7	24.5	36.1	7.0	21.0	2
04097170	68.2	.0	86.8	19.0	37.0	1.0	15.0	2
04097540	107	.0	80.3	18.5	36.5	10.0	15.0	2
04099000	1,880	.0	76.0	18.2	36.4	9.0	14.0	2
04101500	3,670	.0	77.7	18.3	36.9	9.0	14.0	2
04101800	252	.0	94.5	20.8	37.8	23.0	13.0	2

Table 1 –2. Values of selected explanatory variables used in the development of the regression equation for estimating the index flow—Continued.

U.S. Geological Survey station number	Drainage area (square miles)	Percent of basin with low ground-water transmissivity (*L_Trans*)	Percent of basin with high ground-water transmissivity (*H_Trans*)	Percent of basin with forest cover (*Forest*)	Normal annual precipitation for 1971–2000 (inches) (*Precip*)	Percent of basin with A hydrologic soil group (*A_Soils*)	Percent of basin with D hydrologic soil group (*D_Soils*)	Michigan hydrologic subregion
04102500	390	1.9	71.5	25.8	37.7	36.0	16.0	2
04102700	83.5	1.1	36.1	32.5	37.7	23.0	26.0	2
04103010	270	.0	54.7	16.0	34.9	6.0	11.0	2
04104945	48.3	.0	55.4	24.9	35.5	12.0	18.0	2
04105000	274	.0	44.7	22.4	34.8	12.0	17.0	2
04105700	36.8	.0	99.1	28.2	36.6	4.0	16.0	2
04108600	65.1	.0	48.4	21.1	36.7	23.0	13.0	2
04108801	66.9	1.7	25.0	8.3	36.4	11.0	5.0	2
04110000	47.3	.0	39.5	19.6	32.7	25.0	18.0	2
04111500	16.3	.0	4.4	15.1	32.6	14.0	13.0	2
04112000	10.4	.0	.0	13.6	32.5	2.0	8.0	2
04112500	344	.0	13.7	14.2	32.6	11.0	14.0	2
04114498	284	.0	28.9	14.8	32.4	9.0	18.0	2
04115000	420	29.3	20.1	10.7	32.4	5.0	15.0	2
04116500	516	.0	79.0	26.7	34.0	45.0	19.0	2
04117000	7.8	.0	30.8	21.9	35.3	4.0	14.0	2
04117500	410	.0	30.3	18.4	34.9	5.0	12.0	2
04118000	795	.0	31.3	21.8	35.2	11.0	11.0	2
04118500	257	.0	46.6	29.8	34.5	36.0	12.0	2
04121000	352	.0	85.2	63.1	30.7	57.0	16.0	3
04121300	239	3.1	77.2	53.0	32.1	62.0	.0	3
04121900	136	.0	96.0	41.4	33.4	60.0	15.0	3
04122100	16.7	70.6	25.8	43.5	34.0	32.0	15.0	3
04122200	404	.0	81.7	57.9	34.0	62.0	15.0	3
04122500	689	.0	91.3	74.7	33.8	70.0	16.0	3
04123000	115	.0	91.1	79.7	33.5	64.0	21.0	3
04123500	132	.0	100.0	73.0	32.2	92.0	5.0	3
04124000	865	.0	94.4	76.3	31.9	80.0	12.0	3
04124500	58.9	.0	31.2	40.1	32.6	54.0	18.0	3
04125000	130	.0	54.5	51.7	32.7	61.0	11.0	3
04125500	254	.0	72.1	61.6	32.9	63.0	9.0	3
04126200	185	.0	99.3	83.3	33.5	84.0	10.0	3
04127918	202	44.5	26.6	71.3	32.1	30.0	43.0	1
04127997	181	.0	98.6	69.2	31.4	77.0	13.0	4
04128990	57.7	.0	90.6	64.4	30.9	66.0	22.0	4

Table 1 –2. Values of selected explanatory variables used in the development of the regression equation for estimating the index flow—Continued.

U.S. Geological Survey station number	Drainage area (square miles)	Percent of basin with low ground-water transmissivity (*L_Trans*)	Percent of basin with high ground-water transmissivity (*H_Trans*)	Percent of basin with forest cover (*Forest*)	Normal annual precipitation for 1971–2000 (inches) (*Precip*)	Percent of basin with A hydrologic soil group (*A_Soils*)	Percent of basin with D hydrologic soil group (*D_Soils*)	Michigan hydrologic subregion
04129000	586	1.1	55.4	71.4	28.5	38.0	24.0	4
04133501	1,240	2.8	54.6	67.5	28.4	29.0	27.0	4
04135000	96.6	.0	97.8	69.4	31.8	88.0	6.0	4
04135500	71.2	.0	100.0	70.9	31.2	85.0	8.0	4
04135600	391	.0	97.8	81.1	30.2	76.0	15.0	4
04135700	1,360	.5	94.7	80.1	30.2	81.0	11.0	4
04136500	1,740	2.0	89.5	80.9	29.1	79.0	12.0	4
04137500	89.9	.9	21.4	62.4	29.7	42.0	14.0	4
04138000	151	15.2	4.4	41.1	29.9	18.0	32.0	5
04138500	29.8	.0	48.7	58.1	29.6	54.0	14.0	5
04139000	57.4	.0	53.5	58.2	29.5	47.0	19.0	5
04139500	21.0	.0	46.7	45.7	29.8	41.0	15.0	5
04140000	116	.0	53.4	55.8	29.6	44.0	20.0	5
04140500	1.1	.0	.0	13.1	29.9	5.0	8.0	5
04141000	65.4	5.0	50.8	50.8	30.0	52.0	13.0	5
04141500	333	3.6	36.2	56.2	30.0	43.0	20.0	5
04142000	81.9	.0	49.8	14.6	31.5	8.0	18.0	5
04143900	363	.0	34.5	19.4	31.8	7.0	16.0	5
04144000	530	.0	32.0	16.6	31.7	6.0	14.0	5
04144500	89.9	56.4	.0	11.4	32.3	2.0	6.0	5
04145500	51.1	12.3	15.1	20.5	31.4	10.0	14.0	5
04146000	211	19.8	26.1	22.0	31.3	9.0	15.0	5
04146063	526	14.9	33.0	20.1	31.3	9.0	17.0	5
04147500	34.8	12.3	28.4	23.9	31.5	12.0	12.0	5
04147990	99.7	5.6	32.9	19.3	31.5	11.0	16.0	5
04148140	6.9	67.2	.0	6.5	31.6	2.0	4.0	5
04148160	12.1	.0	52.9	21.7	31.5	17.0	26.0	5
04148200	114	6.1	13.1	14.7	31.6	4.0	14.0	5
04148300	54.4	.5	26.9	18.1	31.5	11.0	14.0	5
04148440	960	15.3	27.5	17.7	31.5	9.0	15.0	5
04148500	239	.0	26.6	10.1	31.0	11.0	11.0	5
04150000	363	6.2	31.1	12.4	31.1	12.0	16.0	5
04150500	842	9.8	25.9	20.7	31.0	14.0	18.0	5
04151500	152	1.3	64.3	42.1	31.5	46.0	18.0	5
04152238	145	39.3	4.2	19.6	31.7	6.0	14.0	5

Table 1 –2. Values of selected explanatory variables used in the development of the regression equation for estimating the index flow—Continued.

U.S. Geological Survey station number	Drainage area (square miles)	Percent of basin with low ground-water transmissivity (*L_Trans*)	Percent of basin with high ground-water transmissivity (*H_Trans*)	Percent of basin with forest cover (*Forest*)	Normal annual precipitation for 1971–2000 (inches) (*Precip*)	Percent of basin with A hydrologic soil group (*A_Soils*)	Percent of basin with D hydrologic soil group (*D_Soils*)	Michigan hydrologic subregion
04153500	409	.7	76.7	34.6	32.5	43.0	16.0	5
04154000	309	.7	69.4	24.2	32.9	36.0	15.0	5
04155000	67.3	77.5	.0	4.2	30.9	2.0	5.0	5
04157500	33.9	89.0	.0	3.4	30.9	3.0	3.0	5
04158000	53.3	33.6	1.0	12.2	31.2	5.0	10.0	5
04158500	479	36.3	14.4	7.0	31.0	10.0	10.0	5
041594920	169	32.3	25.6	11.9	30.9	9.0	12.0	6
04159900	184	31.4	23.5	12.5	30.9	9.0	12.0	6
04160000	683	35.0	16.4	9.6	31.0	10.0	11.0	6
04160050	16.1	46.5	.0	12.3	31.0	15.0	15.0	6
04160570	151	53.6	.0	11.4	31.0	6.0	13.0	6
04160600	21.0	.0	100.0	18.7	31.3	34.0	21.0	6
04160800	78.5	.0	95.8	15.4	31.3	26.0	15.0	6
04160900	123	.0	81.0	11.9	31.4	24.0	16.0	6
04161000	17.4	.0	52.5	11.1	31.3	12.0	30.0	6
04161100	39.8	.0	87.3	14.5	31.2	19.0	15.0	6
04161500	71.8	.0	61.9	15.0	31.0	15.0	16.0	6
04161540	23.8	.0	33.6	22.6	31.2	4.0	17.0	6
04161580	69.1	.0	56.7	20.0	31.2	8.0	15.0	6
04161800	16.6	49.3	2.5	8.6	31.4	17.0	16.0	6
04163400	23.8	43.5	1.7	9.3	31.4	14.0	18.0	6
04163500	20.8	4.2	35.4	19.8	31.1	4.0	14.0	6
04164100	12.8	46.2	.0	8.0	31.1	.0	7.0	6
04164300	198	64.4	6.3	11.4	31.2	2.0	19.0	6
04164500	41.2	53.1	3.5	11.0	31.2	13.0	13.0	6
04164800	36.7	4.3	35.2	7.7	31.4	13.0	20.0	6
04166000	10.2	58.0	.0	3.7	31.6	9.0	9.0	6
04166200	17.6	.2	17.2	15.9	31.6	10.0	15.0	6
04166300	49.9	.0	98.5	22.5	31.5	25.0	28.0	6
04169500	139	.0	90.1	18.0	31.7	25.0	25.0	7
04170000	155	.0	90.6	18.3	31.7	24.0	25.0	7
04170500	33.3	.0	76.5	24.6	32.1	7.0	19.0	7
04171500	320	.0	84.4	19.2	32.0	17.0	21.0	7
04172000	538	.0	79.2	20.6	32.2	16.0	22.0	7
04173000	131	.0	50.3	15.4	33.0	12.0	18.0	7

Table 1 –2. Values of selected explanatory variables used in the development of the regression equation for estimating the index flow—Continued.

U.S. Geological Survey station number	Drainage area (square miles)	Percent of basin with low ground-water transmissivity (*L_Trans*)	Percent of basin with high ground-water transmissivity (*H_Trans*)	Percent of basin with forest cover (*Forest*)	Normal annual precipitation for 1971–2000 (inches) (*Precip*)	Percent of basin with A hydrologic soil group (*A_Soils*)	Percent of basin with D hydrologic soil group (*D_Soils*)	Michigan hydrologic subregion
04173500	747	.0	70.8	19.2	32.3	14.0	20.0	7
04174500	817	.0	67.9	18.8	32.3	13.0	21.0	7
04174800	128	.0	90.5	20.0	34.1	9.0	17.0	7
04175600	266	.1	71.6	17.7	34.0	7.0	15.0	7
04175700	460	.0	53.8	15.9	34.2	6.0	14.0	7
04176000	63.7	54.0	0.0	13.0	33.5	23.0	7.0	7
04176605	205	1.4	9.6	14.4	35.0	3.0	9.0	7
04184500	13.2	98.0	.0	91.4	31.0	.0	.0	7

Table 1 –3. Cross-tabulation of cell counts and percentages for Michigan Resource Information System (MIRIS) 1978 land use-land cover and hydrologic soil groups in Michigan[1].

	Hydrologic soil group						
Land use-land cover	A	B	C	D	Outside of Michigan[2]	Percent	Adjusted percent[3]
Urban	4,343,953	4,432,438	1,137,269	555,570	14,563	1.6	6.3
Agriculture	8,163,149	26,685,700	1,2077,467	2,000,247	2,392	7.4	29.3
Range land	5,292,626	4,255,525	1,338,514	2,544,935	5,416	2.0	8.0
Forest	31,740,852	19,970,933	6,257,886	2,3151,323	453,893	12.3	48.8
Water	148,888	193,023	24,015	3,391,954	42,373	0.6	2.3
Wetland	1,457,774	873,403	314,675	5,950,392	59,920	1.3	5.2
Barren	105,535	9,180	9,802	49,231	43,448	0.0	0.1
Outside of Michigan	28,474	7,383	1,510	693,402	496,646,760	74.9	--
Percent	7.7	8.5	3.2	5.8	74.8	100	100
Adjusted percent	30.7	33.7	12.7	22.9	--	100	100

[1]The Michigan Resource Information System represents the 1978 land use-land cover and hydrologic soil groups in Michigan as a rectangular grid of integers that contain 26,319 rows and 25,247 columns. Each grid is identically referenced geographically. Each cell in the grid represents a land area of 30 meters square (900 square meters). Numeric codes for the land use-land cover grid are as follows: (1) urban or built-up land, (2) agricultural land, (3) rangeland, (4) forest land, which included Level II classification of deciduous, evergreen, and mixed forest lands, (5) water, (6) wetland, and (7) barren land. For hydrologic soil groups numeric codes are as follows: (1) group A soils, (2) group B soils, (3) group C soils, and (4) group D soils. For both coverages, the code -9999 signifies no data or inapplicable, which occurs over extensive areas of adjacent states, the Province of Ontario, Canada, and the Great Lakes.

[2] “Outside of Michigan” refers to land areas of adjacent states and the Province of Ontario, Canada, and water areas over the Great Lakes, both within and outside of Michigan.

[3] Adjusted percentage accounts only for the land areas within Michigan.

Hamilton and others—**Regression Model for Computing Index Flows Describing Median Flow…Michigan**—Scientific Investigations Report 2008–5096