Prepared in cooperation with the New Jersey Department of Environmental Protection

Implementation of MOVE.1, Censored MOVE.1, and Piecewise MOVE.1 Low-Flow Regressions with Applications at Partial-Record Streamgaging Stations in New Jersey

object with other necessary information
is a fair way to estimate values for x and y,
<- X@.Data[, 1L]
<- Y@.Data[, 1L]
cx <- X@censor.codes
cy <- Y@censor.codes
# Compute std err of y and of x from eqns of SLR, accounting
se_yaxis <- sqrt((1 - retcor[1L]^2)*retcor[5L]^2*retcor[9L]/(ret
se_xaxis <- sqrt((1 - retcor[1L]^2)*retcor[4L]^2*retcor[9L]/(retco
# These are the regression coefs, Bx is predict y from x
Bx <- retcor[1L]*sqrt(yvar/xvar)
By <- retcor[1L]*sqrt(xvar/yvar)
# Step through each one computing estimated values
for(i in seq(retcor[9L])) {
# Nothing to do if neither censored
if(cy[i] && !cx[i]) { # only cy is censored, use inverse Mill's ratio
ey <- retcor[3L] + Bx*(x[i] - retcor[2L]) # Expected value of y given
y[i] <- ey - se_yaxis*dnorm(y[i], ey, se_yaxis)/pnorm(y[i],
if(!cy[i] && cx[i]) { # only cx is censored
retcor[2L] + By*(y[i] - retcor[3L]) # Expected value

Open-File Report 2018–1089

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

Cover. Image showing R language code written by Dave Lorenz, U.S. Geological Survey on December 30, 2015.

Implementation of MOVE.1, Censored MOVE.1, and Piecewise MOVE.1 Low-Flow Regressions with Applications at Partial-Record Streamgaging Stations in New Jersey

By Susan J. Colarullo, Samantha L. Sullivan, and Amy R. McHugh

Prepared in cooperation with the New Jersey Department of Environmental Protection

Open-File Report 2018–1089

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

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2018

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

For an overview of USGS information products, including maps, imagery, and publications, visit https://store.usgs.gov.

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

Suggested citation:
Colarullo, S.J., Sullivan, S.L., and McHugh, A.R., 2018, Implementation of MOVE.1, censored MOVE.1, and piecewise MOVE.1 low-flow regressions with applications at partial-record streamgaging stations in New Jersey: U.S. Geological Survey Open-File Report 2018–1089, 20 p., https://doi.org/10.3133/ofr20181089.

ISSN 2331–1258 (online)

Contents

Abstract.....1

Introduction.....1

- Purpose and Scope.....3
- Previous Studies.....3

Methods.....3

- Ordinary Least Squares Regression.....4
- Maintenance of Variance Regression (MOVE.1).....5
- Censored MOVE.1.....7
- Piecewise MOVE.1.....10
- Application of the R Script.....10
  - Selection of Flows.....10
  - Index Station Selection.....12
  - Derivation of Weights for Averaging Predicted Flow Statistics.....14
    - Standard Error of Prediction for the 7Q10 Flow Statistic.....15
    - Time-Sampling Error for the 7Q10 Flow Statistic.....15
    - Standard Error of Estimate for the 7Q10 Flow Statistic.....15
  - Predicted Flow Statistics.....16
  - Flow Chart.....18

Summary.....18

Acknowledgments.....20

References Cited.....20

Figures

1. Scatterplot showing MOVE.1 regression line for base flows measured at partial-record station 01396190, South Branch Raritan River at Four Bridges, N.J., and daily flows measured at index station 01380450, Rockaway River at Main Street at Boonton, N.J 6
2. Scatterplot showing flow residuals for partial-record station 01396190, South Branch Raritan River at Four Bridges, N.J., from MOVE.1 regression on daily flows at index station 01380450, Rockaway River at Main Street at Boonton, N.J 6
3. Scatterplot showing percent flow residuals for partial-record station 01396190, South Branch Raritan River at Four Bridges, N.J., from MOVE.1 regression on daily flows at index station 01380450, Rockaway River at Main Street at Boonton, N.J 6
4. Scatterplots showing *A*, regular MOVE.1 regression and *B*, censored MOVE.1 regression for partial-record station 01397800, Neshanic River near Flemington, N.J., and index station 01401000, Stony Brook at Princeton, N.J 8
5. Scatterplots showing *A*, regular MOVE.1 regression and *B*, piecewise MOVE.1 regression for partial-record station 01381900, Passaic River at Pine Brook N.J., and index station 01379000, Passaic River at Millington, N.J 9
6. Scatterplots showing MOVE.1 regression lines used to evaluate the relation between base-flow measurements at partial-record station 01396190, South Branch Raritan River at Four Bridges, N.J., and daily flows at five index stations in New Jersey....13
7. Flow chart showing the MOVE.1 R script 19

Tables

1. Tests applied to flows for inclusion in MOVE.1 regressions 11
2. Percent standard errors of estimate at index stations used to weight flow statistics at partial-record station 01396190, South Branch Raritan River at Four Bridges, N.J 14
3. Flow statistics and regression metrics output by the R script for MOVE.1 16
4. Percent differences in estimated MOVE.1 low-flow statistics between the legacy code and the new R script 17
5. Minimum and maximum percent differences in regression metrics between the legacy code for MOVE.1 and the new R script for MOVE.1....17

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
	Area	
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
cubic foot (ft^3)	0.02832	cubic meter (m^3)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

Abbreviations

MOVE.1	Maintenance of Variance Extension, Type 1
SAS	Statistical Analysis System
R	R Foundation for Statistical Computing
NWIS	USGS National Water Information System
NJDEP	New Jersey Department of Environmental Protection

Implementation of MOVE.1, Censored MOVE.1, and Piecewise MOVE.1 Low-Flow Regressions with Applications at Partial-Record Streamgaging Stations in New Jersey

By Susan J. Colarullo, Samantha L. Sullivan, and Amy R. McHugh

Abstract

The U.S. Geological Survey (USGS) uses Maintenance of Variance Extension Type 1 (MOVE.1) regression to transfer streamflows measured at long-term continuous-record streamgaging stations to partial-record (PR) streamgaging stations where intermittent base-flow measurements are available. MOVE.1 regression is used widely throughout the hydrologic community to extend historical low flows and low-flow statistics at continuous-record streamgaging stations to streamgaging stations that have access to only a partial record of low flows. The method correlates base-flow measurements at PR streamgaging stations with daily mean streamflows measured at index stations that exhibit similar streamflow characteristics.

Following changes in the computing platform for storing, processing, retrieving, and publishing National Water Information System (NWIS) hydrologic data, legacy Statistical Analysis System (SAS) code developed by the USGS to implement the MOVE.1 regression was no longer suitable for reading and processing NWIS streamflow data. To migrate the MOVE.1 program so that it could continue to read streamflow data using the new hydrologic data platform, the SAS code was re-written in R, an open source programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The work described in this report was performed in a study conducted by USGS in cooperation with the New Jersey Department of Environmental Protection.

During migration from SAS to R, graphical and tabular output generated by the R script was compared to output produced by the legacy SAS code to ensure that equations used to perform the MOVE.1 regression remained the same. An option to perform censored MOVE.1 regression was added to extend the MOVE.1 methodology to cases where one or more measured continuous-record or PR streamgaging station flows are zero valued. In addition to permitting censored regression, the new R script includes an option to perform piecewise MOVE.1 regression when the relation between PR station and index station low flows varies significantly across the range of index station streamflows.

Together with traditional MOVE.1 regression, censored and piecewise MOVE.1 regression methods implemented by the R script offer less biased estimates than ordinary least squares regression for the annual 7-day 10-year and other low-flow statistics at PR stations for a range of base-flow conditions. The R script is used to implement the MOVE.1 regression methods across a variety of computing platforms.

Introduction

The New Jersey Department of Environmental Protection (NJDEP) is responsible for issuing permits to surface-water users in New Jersey, based in part on low-flow statistics estimated by hydrologists at the U.S. Geological Survey (USGS). Low-flow statistics are needed by NJDEP to check for compliance with the law controlling diversions from surface waters in the State of New Jersey (N.J.S.A 58:1-13 et seq.) and to establish waste-load allocations under the New Jersey Pollutant Discharge Elimination System program (N.J.C.A. 7:14A) (State of New Jersey, Department of Environmental Protection, 2011; State of New Jersey, Department of Environmental Protection, Division of Water Quality, 2015). Many streams for which permits are required do not have streamflow data from continuous-record streamgaging stations but are instead measured for flows intermittently to provide at least a partial record of gaged flows. For such streams, low-flow information can be “borrowed” from continuous-record stations located on hydrologically similar streams by using widely accepted principles of statistical inference. These continuous-record streamgaging stations, hereafter referred to as index stations or reference stations, can provide useful information about missing flows at intermittently observed gaging stations, referred to in the remainder of this report as partial-record (PR) stations.

The cooperative NJDEP/USGS low-flow project, which has been active since 1959, charges USGS with providing NJDEP with low-flow statistics needed to meet permitting requirements. Owing to cost considerations, many streams for which permits are required do not have a minimum of 10 years of daily streamflow data from index stations. One goal of the long-term cooperative low-flow project is to establish a network of partial-record stations on ungaged streams and correlate low-flow measurements at the PR stations with streamflow measurements from index stations. Unlike index stations where streamflow is measured continuously in time, instantaneous streamflow measurements are made under base-flow conditions at PR stations. These measurements are an important source of cost-effective data for computing the low-flow statistics required for permitting at ungaged stream locations.

Until 2016, low-flow statistics were estimated by the USGS using legacy Statistical Analysis System (SAS) code that implements the Maintenance of Variance Extension Type 1 (MOVE.1) method to correlate instantaneous base-flow measurements at PR stations with continuously recorded daily mean flows at index stations. The legacy code estimated low-flow statistics for PR stations on the basis of the MOVE.1 regression line and low-flow statistics estimated at index stations. Typical low-flow statistics used in the State's point source and water allocation permitting process include the annual minimum daily flow, the 75th percentile daily mean flow duration, and the 1Q10, 7Q10, and 30Q10 low-flow flow statistics. The 1Q10, 7Q10, and 30Q10 low-flow frequency statistics represent the smallest average discharge over any consecutive 1-day, 7-day, and 30-day periods, respectively, that occur once every 10 years, on average. The annual 7Q10 is of particular importance because it provides the foundation for defining passing-flow requirements in New Jersey and is used to support NJDEP surface-water permitting decisions within the State (Hoffman and Domber, 2013). Because low-flow frequency statistics quantify base flows of varying durations over an average 10-year period, at least 10 years of continuous daily record should ideally be available to reliably estimate 10-year recurrence interval flow statistics (Searcy, 1959).

To provide sufficient data to reliably extend the low-flow record at PR stations, 2 to 3 measurements are made at a PR station each year under high and low base-flow conditions for a period of roughly 5 to 6 years. Ten to 12 base-flow measurements generally provide an adequate amount of data to establish a correlation between base flows intermittently measured at a PR station and low flows measured at nearby index stations. Regression techniques grounded in principles of statistical inference are then used to infer a statistical relation between low flows measured at index stations and instantaneous base flows measured at PR stations. This relation is then used to estimate various low-flow statistics for PR stations. In addition to enabling the transfer of low-flow statistics from index to PR stations, regression can also be used to fill gaps in the daily flow record at index stations where gaging has temporarily failed.

A variety of regression methods have been used to infer base flows and low-flow statistics for PR stations, each with their own advantages and limitations. When ordinary least squares regression (OLS) is used to extend the low-flow record from index to PR stations, the variance is underestimated because PR flow estimates do not include the variability of flows about the regression line (Hirsch, 1982). As a result, the variance of estimated flows at the PR station is always biased low when OLS is used to extend the flow record. To overcome this bias, Hirsch (1982) developed the MOVE.1 regression methodology. MOVE.1 regression preserves the mean and variance of flows by minimizing the sum of squared error about the regression line along both axes of the regression plot. This approach provides an estimate of the mean flow that is unbiased, as well as a flow variance estimate that is unbiased for large sample sizes, at the PR station.

Prior to 2016, estimation of low-flow frequency statistics at PR stations using the MOVE.1 regression was implemented by running SAS code on a UNIX platform. Changes in the operating system that supports the current National Water Information System (NWIS) database required a change to the program used to access streamflow data. A decision was made to convert the existing SAS program to R, an open-source interpreted programming language widely used in the physical sciences for statistical computing that is supported by the R Foundation for Statistical Computing. Initial goals of the conversion were to duplicate the way in which instantaneous base flow and continuous daily flow data were input and processed, reproduce MOVE.1 regression line intercept and slope, match low-flow statistics predicted using MOVE.1 regression lines, replicate text and graphical output, and ensure proper calculation of the percent standard error of estimate (*%SEE*). The *%SEE* is used to assign weights for estimating weighted averages of predicted low-flow statistics at PR stations when more than one reference station is available to estimate a low-flow or flow-frequency statistic.

The next goal of script conversion was to add new features that allow the user to opt for alternate regression strategies when traditional MOVE.1 regression is not suited to estimating low flows or low-flow frequency statistics. These new features include incorporation of MOVE.1 censored regression for correlating flows that are zero valued or smaller than the assumed measurement error of 0.05 ft^3/s, as well as an option for performing piecewise MOVE.1 regression for correlating flows when flow characteristics change across the range of measured flows. Where one or more flows at PR or index stations are measured as less than or equal to 0.05 cubic foot per second (ft^3/s), censored MOVE.1 regression offers unbiased estimates of frequency statistics and other flow metrics at PR stations. The piecewise MOVE.1 regression option can easily accommodate abrupt changes in the mean of measured PR station base flows across the range of measured index station flows. Such changes typically arise when a physical process disproportionately affects large and small base flows. An example of such a process is water use, which has a disproportionate effect on small base flows relative to larger base flows. Like censored MOVE.1 regression, piecewise MOVE.1 regression provides unbiased estimates for low flows and frequency statistics.

The new R script duplicates the data processing, statistical analysis, and graphical output of the legacy MOVE.1 program, using the R programming language to maintain functionality with the new USGS database system. The script is compatible with the new Aquarius database system and features improved portability that enhances ease of use by scientists and planners across multiple operating systems. Operations performed by the legacy code and reproduced by the new R script include (1) reading a PR station file listing all index stations that are most hydrologically similar to the PR station to be used in the MOVE.1 regression, (2) reading flow data for the PR station and each of its index stations directly from the NWIS website, (3) performing a MOVE.1 regression for each index station believed to share the same streamflow characteristics as the PR station, (4) estimating frequency statistics using results of the regressions, (5) weighting flow statistics at each index station using the %SEE of its associated MOVE.1 regression, (6) estimating the PR flow statistic as the weighted sum across all index stations, and (7) reproducing all text, spreadsheet, and graphical output. An incremental approach was used to move functionality from the legacy code into R in the order of the sequence of operations listed above. As the conversion progressed, output from the legacy code and new R script were compared, and differences in output were used to identify where further revisions should be made in the R script. Once R script revisions duplicated all operations of the legacy code, the script was finalized.

Other improvements to the script include limited trapping and handling of errors triggered by improper user input, as well as enhanced organization, formatting, and archiving of output. Since the R script runs on a wide variety of platforms, including UNIX, Windows, and macOS operating systems, the revised script is expected to remain operational well into the future, regardless of further modifications made to software used to support the NWIS database.

Purpose and Scope

The purpose of this report is to document a new R script that replaces SAS legacy code used at the USGS to estimate low flows at partial-record stations. The SAS code and R script automate MOVE.1 regression to estimate base-flow frequency statistics at PR stations on the basis of daily flows measured at hydrologically similar index stations located throughout the State. The legacy code was run on a UNIX platform for 25 years, but a decision to no longer maintain a SAS UNIX license, coupled with a change in database software used to retrieve, store, process, and publish NWIS hydrologic data, required that the legacy code be re-written using a more portable programming language that can be executed on a Windows operating system. The SAS program was re-written using R, an open source programming language and software environment for statistical computing and graphics that is widely used by environmental scientists to develop statistical tools. This report highlights and documents key features of the R script relevant to performing operations needed to estimate low-flow statistics, including criteria used to select index and PR station flows, the statistical basis for determining weights used to estimate average flow statistics at a PR station as the weighted sum of flow statistics inferred from regressions performed at multiple index stations, and conditions for which the three distinct types of MOVE.1 regression are best suited. This study was conducted by the USGS in cooperation with NJDEP.

Previous Studies

The concept of maintaining the variance of streamflow data was first alluded to by Riggs (1968), who recommended use of a graphical procedure referred to as the “structural line” method that, in essence, averages the effects of regressing *x* on *y* with the effects of regressing *y* on *x*. Gillespie and Schopp (1982) published the first study related to estimating low-flow statistics at PR stations in New Jersey using this graphical approach. The MOVE.1 methodology, first developed by Hirsch (1982), is an extension of this “structural line” method. Using a mathematically rigorous framework, Hirsch proved that the MOVE.1 method of flow-record extension produces flow estimates at PR stations that are unbiased in the mean and, for sufficiently large sample sizes, also produces unbiased variance. The first application of the MOVE.1 methodology to estimate streamflows in New Jersey was published by Watson and others (2005), who documented use of the SAS MOVE.1 program to estimate low-flow statistics at 500 PR stations, using flow data collected at 66 reference stations with less than 20 years of flow record and 111 reference stations with a minimum of 20 years of flow record. No previous studies documenting use of censored MOVE.1 or piecewise MOVE.1 regression for estimating base flows or low-flow statistics at PR stations in New Jersey have been published. Note that, although Watson and others (2005) reference the regression methodology as “MOVE1,” the methodology is referred to as “MOVE.1” in this report to maintain consistency with Hirsch (1982) and most USGS publications.

Methods

The R script includes a variety of MOVE.1 regression options that can be used to estimate low flows and flow statistics at PR stations on the basis of daily flows measured at index stations. MOVE.1 regression overcomes many limitations associated with traditional OLS regression, including low bias in estimated variance. Unlike OLS regression, which seeks to find the regression line that minimizes the sum of squared PR flow residuals, the goal of MOVE.1 regression is to maintain the mean and variance of PR flows. For a large enough sample size, MOVE.1 regression will predict PR flows with unbiased variance (Hirsch, 1982). In addition to traditional MOVE.1 regression, the R script includes options for performing censored and piecewise variants of MOVE.1 regression that can be used when zero-valued flows are present in the streamflow dataset or when streamflow data exhibit marked changes in the relation between PR and index station flows.

Not all index station or PR station flow measurements are suitable for inclusion in the MOVE.1 regression. Only flows measured at index stations that are hydrologically similar to the PR station are used to predict flows and flow statistics at a PR station. Index stations that exhibit hydrologic similarity to a PR station share common streamflow characteristics, making them ideal candidates for filling gaps in the streamflow record for that PR station. In addition to using only measured flows from index stations that exhibit similar streamflow signatures as the PR station, only flows collected at the PR station under base-flow conditions are used in the regression. A series of base-flow tests, originally established in the legacy SAS code, are imposed on instantaneous PR measurements and daily flows at index stations. These tests are used to exclude concurrent flow measurements collected when base-flow conditions likely did not prevail at the PR and index stations.

Performing MOVE.1 regression on the basis of low flows measured at a single index station can provide reliable base-flow estimates for extension of the flow record at a PR station. However, when streamflows at multiple, hydrologically similar index stations are strongly correlated to flows measured at the PR station, it makes sense to exploit flow information available at all available index stations. The use of multiple index stations can better condition flow statistics at a PR station by transferring different facets of flow behavior to the PR station. Using flows measured at more than one index station also protects predicted PR flow estimates from potential bias introduced by index station flows that may not be truly representative of flows at the PR station. To incorporate flows measured at multiple index stations, all index station flows could, in theory, be pooled and used to perform the MOVE.1 regression. However, this might obscure the unique relation between a PR station and each of its associated index stations. Instead, separate MOVE.1 regressions are performed for each index station, with predicted low-flow statistics at the PR station estimated as a weighted average of low-flow statistics inferred from each individual regression.

Statistics from each MOVE.1 regression are weighted on the basis of the *%SEE* of the regression, a measure of the accuracy in PR flow estimates made using the regression line. Since regressions characterized by high standard errors are more uncertain than those associated with low standard errors, the reciprocal of the *%SEE*, estimated for each index station regression, is used to weight predictions of low-flow statistics before summing them across all index stations. Using the reciprocal of the *%SEE*, more certain PR flow estimates are weighted more heavily so that they contribute a disproportionately larger effect on predicted PR flows than those produced by regressions that have a high degree of uncertainty attached to them.

Ordinary Least Squares Regression

OLS regression has been used for several decades to extend instantaneous flows at a PR station ($\hat{y}(i),\ i = 1,\ldots,N_1$) on the basis of continuously measured flows at an index station ($x(i),\ i = 1,\ldots,N_1 + N_2$) where N_1 denotes the number of PR base flows measured on dates shared by the PR and index stations, N_2 is the number of low flows measured at the index station on days when flow at the PR station is missing, and $N_1 + N_2$ is the total number of measured low flows at the index station. Flows are typically transformed by computing the logarithm to force the probability of flows to be more normal and to produce a more linear relation between $\hat{y}(i)$ and $x(i)$. Values of missing log-transformed flows at the PR station ($\hat{y}(i),\ i = N_1 + 1,\ldots,N_1 + N_2$) have traditionally been estimated using the OLS regression equation

$$\hat{y}(i) = a + bx(i)\ . \qquad (1)$$

The values of intercept *a* and slope *b* in equation 1 are estimated by minimizing the sum of squared residuals for the N_1 log-transformed base-flow measurements available for the PR station that have concurrent log-transformed measured daily flows at the index station:

$$\sum_{i=1}^{N_1}\left[\hat{y}(i) - y(i)\right]^2 \qquad (2)$$

where $\hat{y}(i)$ and $y(i)$ are predicted and measured log-transformed base flows at the PR station. Solving equation 1 for values of *a* and *b* that yield the minimum sum of squared residuals produces the solution

$$\hat{y}(i) = m(y_1) + r\frac{S_{y_1}}{S_{x_1}}\left[x(i) - m(x_1)\right] \qquad (3)$$

where S_{x_1} and S_{y_1} are the standard deviations of the N_1 concurrently measured log-transformed index and PR flows, *r* represents the correlation coefficient describing the strength of the linear relation between the N_1 concurrent log-transformed index and PR base flows, and $m(x_1)$ and $m(y_1)$ are the means of the N_1 concurrently measured log-transformed index and PR flows (Hirsch, 1982).

Use of OLS regression provided by equation 3 to estimate base flows or low-flow statistics at PR stations on the basis of measured low flows at index stations has been shown by Matalas and Jacobs (1964) to produce variances in predicted PR low flows that are biased low by a factor of r^2, the squared correlation coefficient for concurrent index station and PR station base flows. As a consequence of this low bias in flow variance, variability in PR flows estimated using OLS regression is underestimated, causing the frequency of estimated PR flows in both tails of the flow distribution to likewise be underestimated. This tendency for the frequency of predicted extreme low-flow events to be underestimated by OLS regression generally produces overestimated low flows and underestimated high flows, a known drawback of traditional OLS regression (Helsel and Hirsch, 2002). To better reproduce the tails of the underlying PR flow distribution and obtain more reliable estimates of low flows at PR stations, an alternative regression strategy is needed to transfer low flows and low-flow statistics from index stations.

Maintenance of Variance Regression (MOVE.1)

To produce unbiased estimates of the mean and variance at PR stations, Hirsch (1982) took a different regression approach to estimating low flows. Rather than minimize the sum of squared error, a new regression equation was derived to preserve the mean and variance of low-flow predictions. Hirsch (1982) developed the MOVE.1 regression by imposing the constraints that the mean and variance of the N_1 measured and predicted concurrent flows at a PR station be equal:

$$\sum_{i=1}^{N_1} \hat{y}(i) = \sum_{i=1}^{N_1} y(i) \tag{4a}$$

$$\sum_{i=1}^{N_1} \left[\hat{y}(i) - m(y_1)\right]^2 = \sum_{i=1}^{N_1} \left[y(i) - m(y_1)\right]^2 \ . \tag{4b}$$

When the ordinary least squares regression equation 3 is constrained by equations 4a and 4b rather than by minimizing the sum of squared residuals given by expression 2, the following MOVE.1 regression equation is derived:

$$\hat{y}(i) = m(y_1) + \frac{S_{y_1}}{S_{x_1}}\left[x(i) - m(x_1)\right] \tag{5}$$

where $m(y_1)$ and $m(x_1)$ are sample means for the N_1 concurrent low flows at the PR and index stations, respectively. Although MOVE.1 regression equation 5 produces an unbiased sample mean of PR flows (Stedinger and Thomas, 1985), the sample variance for PR flows, S_y^2, is an asymptotically unbiased estimate of the true variance that will be biased high for a finite sample size (Hirsch, 1982).

Note that, in comparing the second term in OLS regression equation 3 to the second term in MOVE.1 regression equation 5, it is evident that the variance of the PR low-flow estimate for the OLS regression is equal to the variance of the MOVE.1 estimate, scaled by a factor of r^2 (Helsel and Hirsch, 2002). For the general case where $|r|$ is less than 1, PR flows estimated by the OLS regression will be characterized by a smaller variance than PR flows estimated using the MOVE.1 regression. Because the MOVE.1 regression, by definition, preserves the variance of observed flows, the OLS regression will generally underestimate the variance of PR station flows. The OLS flow estimate will have the same variance as the MOVE.1 flow estimate only when flows at the PR station and index station are perfectly correlated, with $|r| = 1$.

One important property of the MOVE.1 regression is that it minimizes the sum of squared deviations of measured PR and index station flows from the regression line in both the *x* and *y* dimensions to define the regression line, also known as the line of organic correlation (LOC). The LOC lies between traditional OLS regression lines obtained by regressing *y* on *x* and *x* on *y* (Kritskiy and Menkel, 1968). MOVE.1 produces the same unique LOC, regardless of whether *x* or *y* is used as the dependent variable. Halfon (1985) refers to MOVE.1 as the geometric mean regression because the slope of the MOVE.1 regression line is equal to the geometric mean of the slopes associated with the traditional OLS regression line for *x* regressed on *y* (b_{xy}) and for *y* regressed on *x* (b_{yx}):

$$\sqrt{b_{xy} b_{yx}} \ . \tag{6}$$

Yet another way of viewing the MOVE.1 regression is that it is equivalent to minimizing the sum of the areas of right triangles, each defined by vertical and horizontal lines extending from measured flows to the regression line, indicating that the LOC minimizes errors along *x* and *y* axes (Hirsch and Gilroy, 1984). Another characteristic of MOVE.1 that makes it preferable to OLS for flow extension is the fact that the distributional properties of estimated PR flows produced by a MOVE.1 regression are similar to those of measured PR flows. Preservation of distributional properties is important when estimates of flow percentiles at PR stations, not just mean flow or a single flow estimate, are of interest (Helsel and Hirsch, 2002).

The MOVE.1 regression line estimated using low-flow data from PR station South Branch Raritan River at Four Bridges, N.J., (01396190) and index station Rockaway River at Main Street at Boonton, N.J., (01380450) are shown in figure 1. Note that the station at Four Bridges was operated as an index station from water years[1] 1999 through 2012. Despite the fact that the station has 13 years of continuous flow record, for purposes of this study, it is considered a PR station because it has less than 20 years of daily streamflow record. The points plotted in figure 1 represent concurrent log-transformed flows measured at the two stations on the same days. As one would expect, because both stations are affected by similar rainfall events and other physical processes, smaller PR station flows correspond to smaller flows measured at the index station. The points generally fall close to the regression line, indicating that daily flows measured at the index station explain much of the variation in daily flows at the PR station. The strong degree of correlation between PR station flows and index station flows is further supported by the high value of 0.96 estimated for R^2, the coefficient of determination for the MOVE.1 regression. On the basis of this value of R^2, the MOVE.1 regression line explains 96 percent of the variability of PR station flows about

[1] A water year is the 12-month period beginning October 1 and ending September 30 of the following year. It is designated by the year in which it ends.

their mean. Note that the coefficient of determination, R^2, is a measure of how much variance is explained by the regression, whereas the correlation coefficient r describes the degree of linear correlation between index and PR flows.

Also shown plotted in figure 1 and other regression plots are values of mean annual flow (QAvg) and the 7Q10 low-flow statistic. Mean annual flow can be considered an upper bound on base-flow statistics computed by the script, with the 7Q10 and other low-flow frequency statistics equal to QAvg only when streamflow measured at the index station is constant in time. The difference between 7Q10 and QAvg can be considered a measure of the degree of daily flow variability at the index station used for the regression.

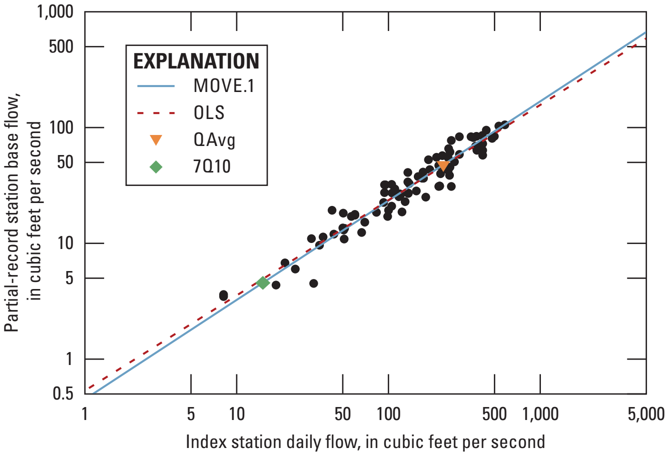


Figure 1. MOVE.1 regression line for base flows measured at partial-record station 01396190, South Branch Raritan River at Four Bridges, N.J., and daily flows measured at index station 01380450, Rockaway River at Main Street at Boonton, N.J.

(MOVE.1, Maintenance of Variance Extension, Type 1; OLS, ordinary least squares regression; QAvg, average daily flow; 7Q10, 7-day 10-year low flow)

The plot of PR flow residuals versus index flows output by the R script is shown in figure 2. These residuals represent deviations of measured PR flows from the MOVE.1 regression line presented in figure 1. Examination of the residual plot reveals that larger deviations from the MOVE.1 regression line tend to occur for larger index-station daily flows, indicating an increase in variance with increasing flows. This increase in flow variance with increasing values of flow, often referred to as the proportional effect, is typical for variables that follow a lognormal distribution (Journel and Huijbregts, 1978). To eliminate this effect, the R script normalizes all residuals by index-station flows to produce the plot of percent residual versus index flow shown in figure 3. The diffuse cloud of points in figure 3, free of obvious trends or discontinuities, indicates the absence of significant structural error in the MOVE.1 regression model for PR station 01396190 flows regressed on daily mean flows measured at index station 01380450. Absence of structural error implies that most physical processes affecting the relation between PR station flows and index station flows have been integrated into the regression and indicates that processes common to both stations exert roughly the same effect on both sets of flows.

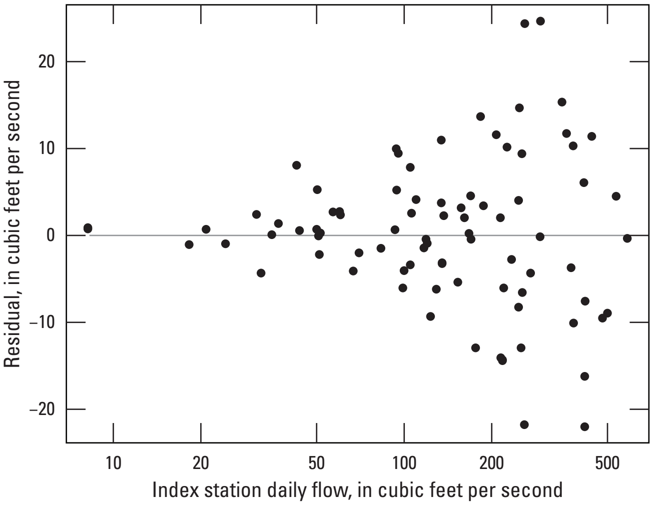


Figure 2. Flow residuals for partial-record station 01396190, South Branch Raritan River at Four Bridges, N.J., from MOVE.1 regression on daily flows at index station 01380450, Rockaway River at Main Street at Boonton, N.J.

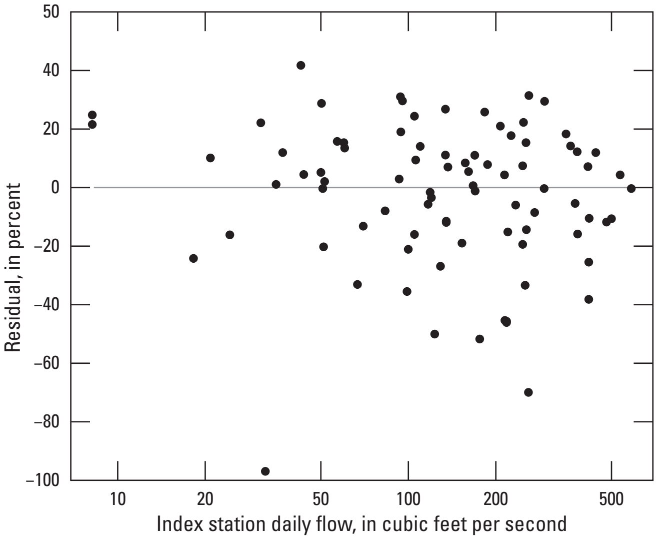


Figure 3. Percent flow residuals for partial-record station 01396190, South Branch Raritan River at Four Bridges, N.J., from MOVE.1 regression on daily flows at index station 01380450, Rockaway River at Main Street at Boonton, N.J.

Censored MOVE.1

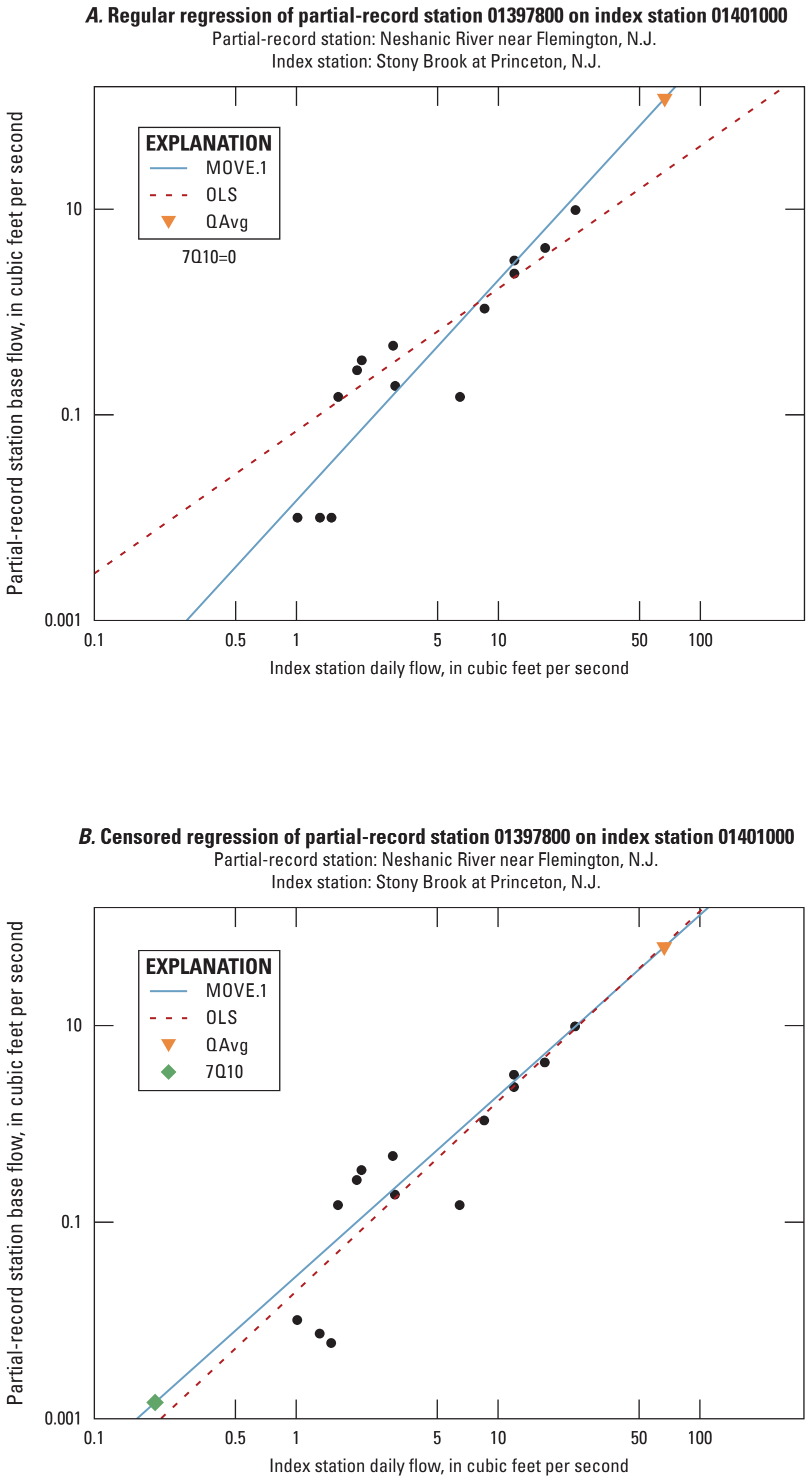
Under some circumstances, low-flow data cannot support the use of a MOVE.1 regression to predict the flow statistics. When zero-valued flows are part of the flow record at either the PR station or the index stations, a censored version of MOVE.1 ideally is implemented to avoid low bias in estimated PR flows. The new R script includes an option for performing left-censored MOVE.1 regression when flows equal to or less than 0.05 ft^3/s occur in the concurrent flow record. The censoring threshold of 0.05 ft^3/s was chosen because, for this study, estimated flows less than 0.05 ft^3/s were rounded to 0.0 ft^3/s to account for measurement errors that typically occur when collecting streamflow data.

Left-censored MOVE.1 regression was not included in the legacy code. Instead, a constant of 0.01 ft^3/s was added to all zero-valued flows prior to log transformation and, to prevent structural error in the regression, to all non-zero flows as well. Although this constant was subtracted from estimated flows, it introduced bias into the MOVE.1 regression because adding a value of 0.01 ft^3/s to small flows has a disproportionate effect on the regression compared with the addition of the same value to larger flows. With publication of the censored MOVE.1 regression R function censMove.1 developed by David Lorenz of USGS in 2014 (https://github.com/USGS-R/smwrQW/blob/master/R/censMove.1.R), it became possible to avoid this pitfall. The censMove.1 function retains all the advantages of the regular MOVE.1 regression methodology but allows for input of zero-valued concurrent flows measured at index stations, PR stations, or both. During implementation of the censored MOVE.1 regression, index and PR station flows are assumed to be characterized by a bivariate log-normal distribution. If all measured index station and PR station flows are uncensored, the actual measured flow values are input to the regular MOVE.1 regression. If a measured PR flow value is censored and its concurrent index flow is not censored, then the censored PR flow is set equal to the expected value of PR flows, given the censored PR flow value and the uncensored value of its concurrent measured index station flow. Likewise, if a measured index flow value is censored and its concurrent PR flow is not, then the censored index flow is assumed equal to the expected value of index flows, given the censored index flow value and the uncensored value of its concurrent measured PR station flow. Finally, if both measured PR and index flows are censored, then their values are set equal to their corresponding expected values, given their censored values, without regard for the censored value of the concurrent flow variable.

During implementation of the censored MOVE.1 regression, index and PR station flows are assumed to be characterized by a bivariate log-normal distribution. If all measured index station and PR station flows are uncensored, the actual measured flow values are input to the regular MOVE.1 regression. If a measured PR flow value is censored and its concurrent index flow is not censored, then the censored PR flow is set equal to the expected value of PR flows, given the censored PR flow value and the uncensored value of its concurrent measured index station flow. Likewise, if a measured index flow value is censored and its concurrent PR flow is not, then the censored index flow is assumed equal to the expected value of index flows, given the censored index flow value and the uncensored value of its concurrent measured PR station flow. Finally, if both measured PR and index flows are censored, then their values are set equal to their corresponding expected values, given their censored values, without regard for the censored value of the concurrent flow variable.

To illustrate, if index flow is less than 0.05 ft^3/s but flow at the PR station is not, then index station flow is estimated as the mean of the conditional distribution of index station flows, given an index station flow of 0.05 ft^3/s and the uncensored value of PR station flow. Likewise, for the case of censored index flow and uncensored PR flow, index flow is estimated as the expected value of index flow, given its censored value and the value of the concurrent PR flow. If index and PR station flows are censored, their values are estimated as expected values of distributions conditioned on their censored values, without regard to the censored value of the concurrent flow.

Results of regressions for PR station Neshanic River near Flemington, N.J., (01397800) and index station Stony Brook at Princeton, N.J., (01401000) using regular and censored MOVE.1 regression methodologies are shown in figure 4. When 0.01 ft^3/s was added to all flows so that the three zero-valued PR flows at station 01397800 were no longer equal to zero, the regular MOVE.1 regression produced a line with a smaller intercept and larger slope than the censored MOVE.1 regression with unmodified flows. It is evident from examination of the two regression plots that predicted PR flow statistics, based on the regular MOVE.1 line in figure 4*A*, will be biased high for large index station flows and biased low for small index station flows. Because all streamflows input to the regular MOVE.1 regression are modified by adding 0.01 ft^3/s and streamflows input to the censored MOVE.1 regression represent unaltered flows, the two types of regression operate on different sets of flows. Differences between flows in the two datasets are especially pronounced in lower ranges of flow, where the addition of 0.01 ft^3/s disproportionately affects flows relative to larger flows, and can dramatically change the estimate of *%SEE*. As a consequence of using fundamentally different streamflow datasets for the regular MOVE.1 and censored MOVE.1 regressions shown in figure 4, *%SEE* values from the two types of MOVE.1 regression cannot be used to objectively assess their relative goodness of fit.



***A.* Regular regression of partial-record station 01397800 on index station 01401000**

Partial-record station: Neshanic River near Flemington, N.J.

Index station: Stony Brook at Princeton, N.J.

***B.* Censored regression of partial-record station 01397800 on index station 01401000**

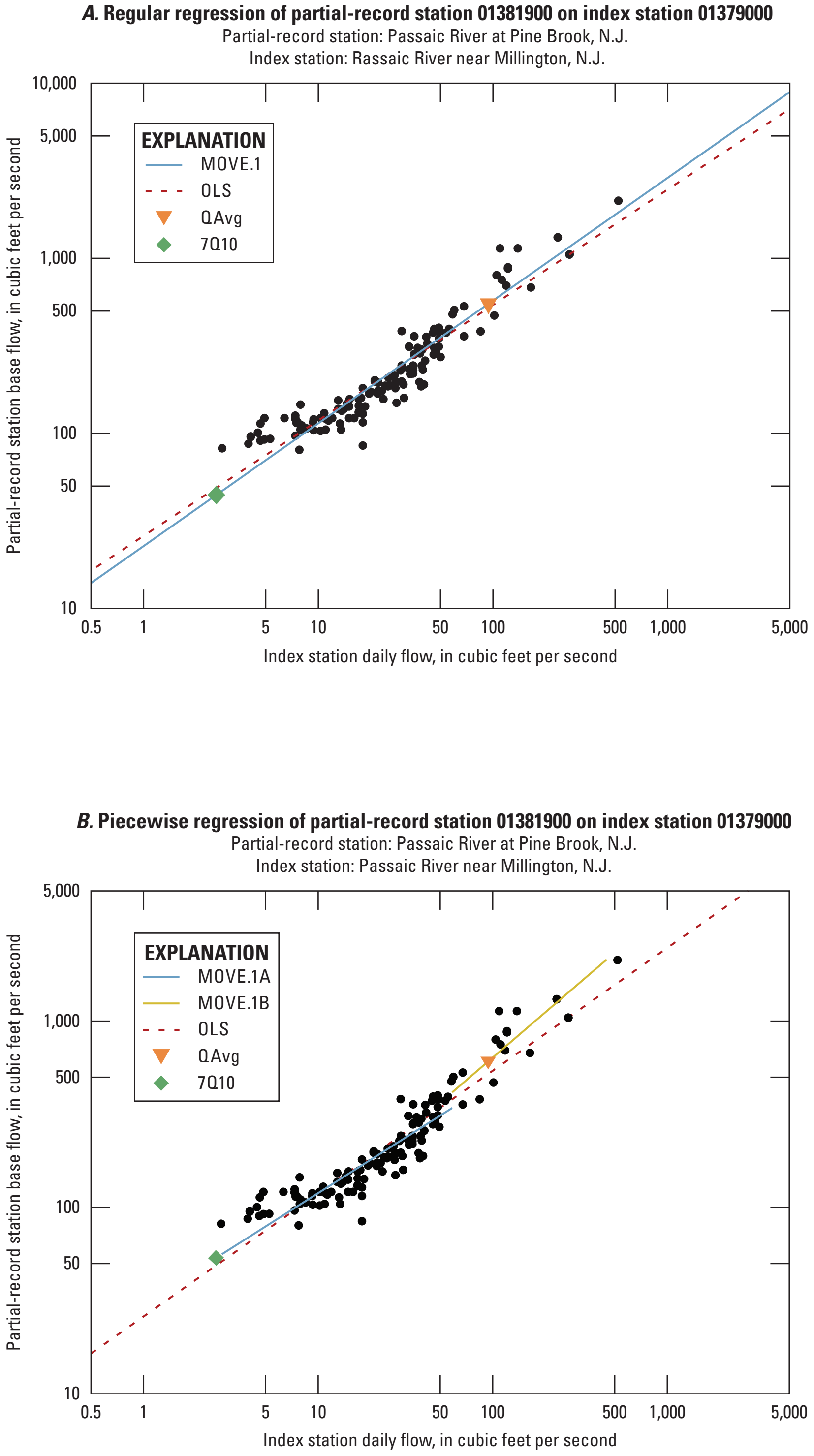
Partial-record station: Neshanic River near Flemington, N.J.

Index station: Stony Brook at Princeton, N.J.

EXPLANATION
MOVE.1
OLS
QAvg
7Q10
10
0.1
0.001
0.1
0.5
1
5
10
50
100
Partial-record station base flow, in cubic feet per second
Index station daily flow, in cubic feet per second

Figure 4. *A*, Regular MOVE.1 regression and *B*, censored MOVE.1 regression for partial-record station 01397800, Neshanic River near Flemington, N.J., and index station 01401000, Stony Brook at Princeton, N.J.

(MOVE.1, Maintenance of Variance Extension, Type 1; OLS ordinary least squares regression; QAvg, average daily flow; 7Q10, 7-day 10-year low flow)



A. **Regular regression of partial-record station 01381900 on index station 01379000**
Partial-record station: Passaic River at Pine Brook, N.J.
Index station: Rassaic River near Millington, N.J.

B. **Piecewise regression of partial-record station 01381900 on index station 01379000**
Partial-record station: Passaic River at Pine Brook, N.J.
Index station: Passaic River near Millington, N.J.

EXPLANATION
MOVE.1A
MOVE.1B
OLS
QAvg
7Q10
Partial-record station base flow, in cubic feet per second
5,000
1,000
500
100
50
10
0.5
1
5
10
50
100
500
1,000
5,000
Index station daily flow, in cubic feet per second

Figure 5. *A*, Regular MOVE.1 regression and *B*, piecewise MOVE.1 regression for partial-record station 01381900, Passaic River at Pine Brook N.J., and index station 01379000, Passaic River at Millington, N.J.

(MOVE.1, Maintenance of Variance Extension, Type 1; MOVE.1A, regular; MOVE.1B, piecewise; OLS, ordinary least squares regression; QAvg, average daily flow; 7Q10, 7-day 10-year low flow)

Piecewise MOVE.1

In some cases, PR flows may exhibit a distinctly nonlinear relation with index flows, indicating that different processes affecting streamflow are in play across different ranges of index flows. Reasons for such statistically non-homogeneous behavior can be attributed to a number of causes, including water use that disproportionately reduces streamflows and causes the type of plot shown in figure 5*A*. Under such circumstances, a piecewise approach to MOVE.1 regression can help resolve the statistically non-stationary behavior by fitting two or more regression lines to the flow data.

At the time of publication of this report, an R function that performs piecewise MOVE.1 regression had not yet been developed. In place of a formal MOVE.1 piecewise regression, an iterative approach first described by Crawley (2013) was included as a regression option in the R script. The piecewise regression option allows the user to implement regular MOVE.1 regression by defining two separate regression lines that intersect at a breakpoint where a marked change in the relation between measured flows at the PR station and the index station occurs. To perform the piecewise regression, the user is prompted to enter upper and lower bounds on the index flow interval within which an obvious discontinuity in the relation between PR and index flows occurs. Only one breakpoint defining the shared endpoint for the two regression segments is allowed to occur within this interval. Additional R code was written to iteratively vary the location of this single breakpoint in fixed increments across the user-defined index flow interval and estimate standard error at each trial breakpoint. For an unbiased estimator, the mean squared error and the standard error of estimate (*SEE*) are equal. The trial breakpoint that produces the smallest *SEE* defines where the optimal breakpoint occurs along the index flow axis. Independent MOVE.1 regressions are then performed for flows greater than and less than the shared breakpoint, each characterized by a unique slope and intercept. Future improvements in the R script may include a piecewise regression methodology that doesn’t require the user to provide index flow interval bounds that define the upper and lower limits constraining where the true breakpoint is likely to be located.

Results of regular MOVE.1 and piecewise MOVE.1 regressions performed for PR station Passaic River at Pine Brook, N.J., (01381900) and index station Passaic River at Millington, N.J., (01379000) are shown in figures 5*A* and 5*B*. Although the regular MOVE.1 regression line does a fairly good job of minimizing error globally across the full range of index flows, it fails to locally minimize error in the lower ranges of index flow. Piecewise MOVE.1 regression does not provide a meaningful *%SEE* for both regression line segments that can be compared to the *%SEE* for regular MOVE.1 regression. However, the piecewise regression presented in figure 5*B* provides a smaller *%SEE* for index flows less than the 35 ft^3/s breakpoint than regular MOVE.1 regression does for the same range of smaller flows, indicating that piecewise regression explains more of the variability in measured PR flows than regular MOVE.1 regression for index flows smaller than 35 ft^3/s.

Application of the R Script

The R script begins execution by installing all libraries that support R functions needed to read NWIS flows, perform MOVE.1 regressions, and estimate flow statistics at the index stations. The user is prompted to select the location of the output directory using a folder selection dialog. The R script then warns the user that certain function files must be present for the script to run correctly. The script then asks the user to select the location where the text input file containing station identifiers for the PR stations and their index stations resides. The user is then offered the option of appending all output to existing output files. An interface with three radio buttons, each corresponding to one of the three MOVE.1 regression variants, allows the user to select the type of regression to be performed. The final interface allows the user to uncheck boxes corresponding to index stations that the user does not want to include in the regression analysis.

Once the user specifies the location of the station identifier input file and the type of MOVE.1 regression, the R script reads all station identifiers, extracts index station and PR station flows from the NWIS database, and loads historical low-flow statistics previously estimated for all New Jersey index stations on the basis of daily flows measured until water year 2001. The R script then loops through each index station associated with the PR station, performing a MOVE.1 regression using all index and PR station flows that meet a series of base-flow criteria, and computes the *SEE* for the regression corresponding to each index/PR station pair. Once all index station regressions have been performed for the PR station, the MOVE.1 regression line for each regression is used to predict flow statistics at the index station. The reciprocal of *%SEE* for each regression is then used to weight the flow statistic at each index station, and the weighted index station statistics are summed across all index stations to estimate the flow statistic at the PR station.

Selection of Flows

As part of an ongoing agreement between the U.S. Geological Survey Water Science Center in New Jersey and NJDEP, instantaneous streamflow measurements at PR stations are made under base-flow conditions 2 to 3 times a year at 80 designated low-flow PR stations throughout New Jersey. In total, there are more than 600 PR stations in New Jersey with enough instantaneous streamflow measurements to potentially be used to estimate low-flow statistics. However, because not all measurements are actually made under base-flow conditions, some of these flows must be excluded from the regression to obtain reliable estimates of base flows at the PR station.

To determine whether base-flow conditions prevailed at the time of measurement at the PR and index station, each flow record is subjected to a sequence of screening tests. If a given flow record satisfies all base-flow test criteria listed in table 1, it is included in the MOVE.1 regression. Note that, owing to the discrete nature of instantaneous base flows measured at PR stations, they were subject to a different set of tests than index daily flows measured at index stations. Although base-flow tests at PR stations rely solely on flags contained in the NWIS dataset, base-flow testing at index stations depends on changes in daily continuous flow over one or more consecutive days.

As shown in table 1, PR measurements flagged as not measured under base-flow conditions, with a BASEFLOW attribute NBAS indicating "not base flow" in the instantaneous measurement file, are assumed to have been measured during a time when base-flow conditions were known not to have occurred. These measurements were automatically excluded from the MOVE.1 regression without further consideration. Instantaneous PR flow measurements flagged with a BASEFLOW attribute of unspecified conditions (UNSP) or base-flow conditions (BASE), indicating that it was either unknown whether the flow measurement was made under base-flow conditions or that base-flow conditions were known to exist at the time of measurement, were included in the MOVE.1 regression. After filtering for base-flow conditions at the time of measurement, instantaneous flow measurements made at PR stations were then subject to a gage-height test to eliminate instantaneous flows made at times when stream stage was observed to be changing rapidly. For an instantaneous measurement at a PR station to be considered an accurate base-flow value, stream stage could not change by more than 0.02 foot (ft) during the time it took to make the flow measurement, unless the measurement was attributed as being made during tidal conditions, and the appropriate tidal adjustment was made to the measured flow.

Flows measured at index stations were subject to an entirely different sequence of base-flow tests than PR station flows. A number of flow criteria at index stations must be passed if the concurrent flow at the PR station has an attribute of UNSP or BASE. These criteria require comparison of daily flows on consecutive days, data that are typically unavailable at PR stations where instantaneous flow measurements are made. Index station daily flows are considered to be measured under valid base-flow conditions only when they drop by less than 30 percent and rise by less than 10 percent, relative to daily flow measured on the previous day.

In certain cases, instantaneous PR measurements indicate that base-flow conditions are present at the partial-record station despite a concurrent increase in measured daily flow at an index station relative to flow observed at the index station on the previous day. These conditions are generally encountered following late-day convective rainfall events that prevail during summer months. Because convective rainfall events are generally localized in spatial extent, they can increase flow at an index station without significantly influencing flow at the associated PR station. Under such circumstances, flow at the index station is estimated when the base-flow flag BASEFLOW is equal to BASE for the concurrent PR flow. To provide this estimate of index site flow, an additional test is subsequently performed to determine whether daily flow measured at the index station 2 days previously was greater than flow measured 1 day prior to the daily flow measurement at the index station. If true, index station flow is estimated to be equal to flow from the previous day, corrected by the rate of recession observed between 1 and 2 days prior. If not true, the index station measurement is excluded from the regression.

Table 1. Tests applied to flows for inclusion in MOVE.1 regressions.

[PR, partial-record]

Name of base-flow test	Type of station tested	Description	Exception
Base-flow flag check	PR station	Measurements flagged as BASEFLOW=NBAS in the measurement file are excluded.	None
Gage-height test	PR station	Gage-height change during the flow measurement (GHCHGF) must be less than or equal to 0.02 foot.	If a measurement is flagged as being adjusted for tide effect (BASEFLOW='TADJ') and appropriate flow adjustment was made.
Base-flow test	Index station	Daily flow must decrease by less than 30 percent or rise by less than 10 percent, relative to flow measured on the previous day.	1. If BASEFLOW='BASE' for concurrent PR flows, then index station daily flow is checked to determine whether it has been receding over the previous 2 days. If true, daily flow at the index station on the day of the measurement is estimated by linear interpolation. 2. If the absolute value of the change in daily flow is less than 0.5 cubic foot per second.

In a final base-flow test of index station flow, an exception is made for index stations with very low daily flows that do not pass criteria for percent change in flow from the previous day. If the absolute value of change in daily flow is less than 0.5 ft^3/s from the previous day, the index station flow is included in the regression. Owing to insensitivity of the stage-discharge rating for very low flows at some index stations, a change of only 0.01 ft in stage height can produce a change in index flow of 20 percent or more. Under these conditions, small fluctuations in continuous gage height record unrelated to rainfall events can produce relatively large percent differences at low flows computed from a rating curve. These index flows are included in the regression, as it is assumed that base-flow conditions persist.

The final set of concurrent PR and index station base-flow measurements included in the MOVE.1 regression excludes all PR station measurements with a BASEFLOW attribute of NBAS, all PR measurements that fail the gage-height test, and all index-station flows that fail the consecutive-day base-flow tests. Failure of any base-flow test performed on a PR station measurement automatically excludes the concurrent measurement made at the index station and vice versa. Exclusion of concurrent measurement pairs when a base-flow test failed for either station ensured that the MOVE.1 regression was performed on pairs of flows measured under similar base-flow conditions, helping to prevent a high bias for predictions of the annual 7Q10 and other low-flow statistics computed using the MOVE.1 regression line.

Index Station Selection

New Jersey is fortunate to have over 120 index stations with more than 20 years of daily flows. To increase confidence in low-flow frequency statistics estimated at PR stations, only index stations with a minimum of 20 years of continuous record were used for this study. The MOVE.1 regression provides a means of transferring flow information from an index station for which continuous daily streamflow records are available to a PR station for which only instantaneous measurements of base flow are available or, in some instances, for which the continuous daily mean flow record is less than 20 years. Selection of index stations suitable for inferring flow estimates at a given PR station can be somewhat subjective. Only index stations that are not strongly affected by diversions, exhibit no significant trends in flow over time, drain basins characterized by similar physical properties affecting streamflow, and share the same physiographic province as the PR station are considered as candidates for inclusion in the MOVE.1 regression model. Such index stations tend to be hydrologically similar to the PR station and characterized by streamflow hydrographs with similar types of flow characteristics as the PR station.

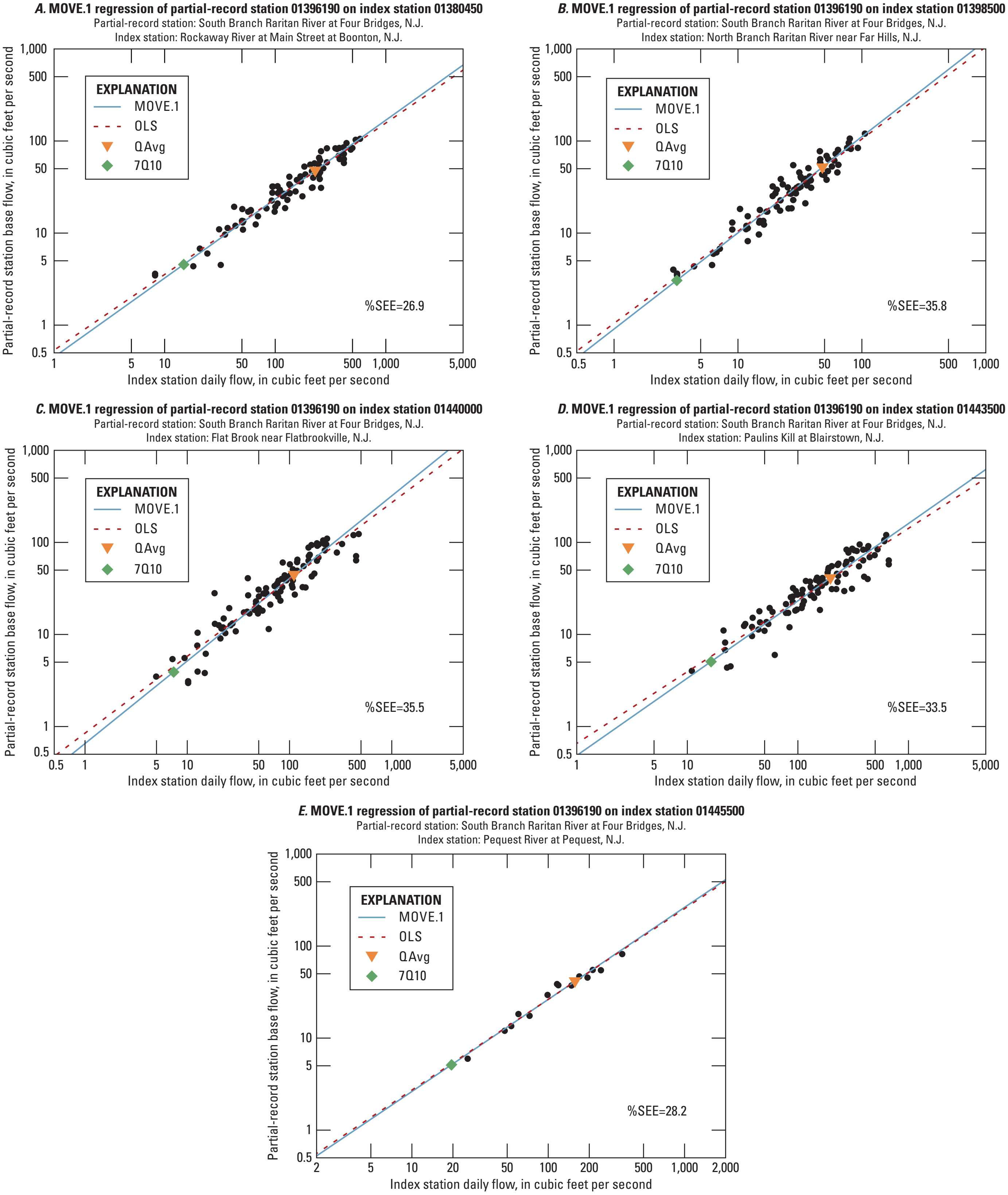
In New Jersey, there are often multiple index stations that are appropriate for use in transferring flow information to a single PR station. Use of daily flows measured at multiple index stations provides more reliable estimates of flow statistics at PR stations than use of daily flows at a single index station. Rather than pooling daily flows measured at the five index stations and performing a single MOVE.1 regression, separate regressions were performed for each of the five index stations, and estimated flows and flow statistics were weighted according to the *SEE* for each regression. This pairwise approach to regression preserves unique flow information shared between the PR station and each of its index stations. Because index station selection can be subjective even when adhering to established guidelines set forth to include only stations with similar hydrologic characteristics, incorporating flow information from multiple index sites based on the strength of their MOVE.1 regressions helps eliminate some of that subjectivity.

To illustrate how multiple index stations are used to infer flow statistics at a PR station, separate regressions were performed for each of the five hydrologically similar index stations associated with PR station 01396190. The five separate MOVE.1 regressions shown in figure 6 exhibit varying degrees of scatter about their regression lines. Differences in the amount of scatter of flow measurements about a regression line need to be accounted for when the line is used to predict a low-flow statistic for the PR station. Clearly, a flow statistic predicted using a MOVE.1 regression at an index station characterized by a large amount of flow variation about the regression line should have less effect on the estimate of the statistic at the PR station, and be assigned a smaller weight, than a statistic estimated using an index-station regression that shows a tighter degree of scatter about the line. To provide an example of how a low-flow statistic is estimated for a PR station as a weighted average of flow statistics predicted at each of the five index stations, the annual 7Q10 at the PR station is estimated using the equation

$$7Q10_{PR} = \sum_{j=1}^{L} \omega_j * 7Q10_j \qquad (7a)$$

$$\sum_{j=1}^{L} \omega_j = 1 \qquad (7b)$$

where L is the number of index stations used to estimate the low-flow statistic at the PR station, ω_j denotes the weight at the jth index station, $7Q10_j$ is the estimated annual 7Q10 at the jth index station, and $7Q10_{PR}$ is the annual 7Q10 at the PR station, estimated as the weighted average of the annual 7Q10 values across the L index stations, with all weights required to sum to 1.



A. **MOVE.1 regression of partial-record station 01396190 on index station 01380450**
Partial-record station: South Branch Raritan River at Four Bridges, N.J.
Index station: Rockaway River at Main Street at Boonton, N.J.

B. **MOVE.1 regression of partial-record station 01396190 on index station 01398500**
Partial-record station: South Branch Raritan River at Four Bridges, N.J.
Index station: North Branch Raritan River near Far Hills, N.J.

C. **MOVE.1 regression of partial-record station 01396190 on index station 01440000**
Partial-record station: South Branch Raritan River at Four Bridges, N.J.
Index station: Flat Brook near Flatbrookville, N.J.

D. **MOVE.1 regression of partial-record station 01396190 on index station 01443500**
Partial-record station: South Branch Raritan River at Four Bridges, N.J.
Index station: Paulins Kill at Blairstown, N.J.

E. **MOVE.1 regression of partial-record station 01396190 on index station 01445500**
Partial-record station: South Branch Raritan River at Four Bridges, N.J.
Index station: Pequest River at Pequest, N.J.

Figure 6. MOVE.1 regression lines used to evaluate the relation between base-flow measurements at partial-record station 01396190, South Branch Raritan River at Four Bridges, N.J., and daily flows at five index stations in New Jersey.

(%SEE, percent standard error of estimate; MOVE.1, Maintenance of Variance Extension, Type 1; OLS, ordinary least squares regression; 7Q10, 7-day 10-year low flow)

Derivation of Weights for Averaging Predicted Flow Statistics

The *SEE*, which quantifies the degree of scatter in PR flows about the regression line, offers a useful metric for assigning meaningful weights when combining flow statistics predicted from separate MOVE.1 regressions performed at *L* index stations, where *L* varies with the PR station. In the context of regression, the *SEE* is a measure of the accuracy of predictions made using the regression line. A larger value of *SEE* indicates a greater amount of unexplained variance in measured PR station flows when regressed on flows observed at an index station. Large values of *SEE* indicate that flow statistics estimated for the PR station are highly uncertain and should be assigned less weight than flow statistics estimated from regressions performed at index stations with smaller values of *SEE* that account for more variation in observed PR station flows.

The percent standard error of estimate, or *%SEE*, is the percent of unexplained variance relative to the value of the estimate. *%SEE* is a better measure of uncertainty than *SEE* because it accounts for the proportional effect encountered in natural streamflow processes, where greater variances are observed for larger flows. To lessen the weight of predictions made using MOVE.1 regressions characterized by a high degree of unexplained variability in flows and increase the weight of predictions made using regressions that exhibit a small amount of unexplained flow variability, weights ω_j , j=1, 2, …, *L* are defined as the reciprocal of *%SEE* for the *j*th index station regression:

$$\omega_j = \frac{1}{\%SEE_j} \quad j=1, 2, \ldots, L \tag{8}$$

where $\%SEE_j$ is the percent standard error of estimate at the *j*th index station. By assigning weights as the reciprocal of the *%SEE*, flow statistics at index stations that do a poor job of explaining measured PR flows will be provided less weight in the calculation of PR station flow statistics than index stations with flows that explain more of the variation in measured PR flows. Use of the reciprocal of the *%SEE* penalizes highly uncertain estimates characterized by large standard errors while assigning greater weight to estimates associated with less uncertainty, without disproportionately penalizing estimates of PR low-flow statistics made using regressions performed at index stations characterized by larger flows. Weighting of regression estimates by the reciprocal of their percent standard errors of estimate is typically used in statistical analysis to combine independent estimates (Hartung and others, 2008).

To illustrate how *%SEE*s from MOVE.1 regressions are used to weight index-station flow statistics, consider the effect of low-flow statistics predicted on the basis of MOVE.1 regressions for base-flow measurements made at the PR station at South Branch Raritan River at Four Bridges, N.J., (01396190) on the basis of daily flows at *L*=5 index stations. The regression line for index station 01380450 (fig. 6*A*) explains more of the variance in measured flows at the PR station, showing a smaller amount of scatter in measurements about the regression line and, as indicated in table 2, the smallest *%SEE* of all *L* index stations. As a result of this small *%SEE*, flow statistics estimated by regressing on mean daily flows at index station 01380450 are expected to be more accurate than statistics predicted using the regression on mean daily flows at the other index stations. Because a weight equal to the reciprocal of this smaller *%SEE* will be larger than the weight associated with the *%SEE* values for the other index stations, the weighted average will favor estimates of low-flow statistics made using the regression for index station 01380450.

The *%SEE* for a *t*-year flow statistic is calculated using its standard error of prediction (SEP_t), along with its time-sampling error (SE_t). For ease of computation, the R script calculates SEP_{10} and SE_{10} , the standard error of prediction and time-sampling error corresponding to the 10-year recurrence interval, rather than computing them on the basis of the recurrence interval of the particular flow statistic of interest. The rationale for using values of SEP_{10} and SE_{10} to calculate the *%SEE* for weighting all flow statistics at PR stations can be traced to the widespread use of the annual 7Q10 as a common statistic for computing waste load allocations and for defining minimum passing flows required to preserve downstream water rights and maintain the health of aquatic ecosystems in the State of New Jersey (Hoffman and Domber, 2013). Equations used to compute SEP_{10} and SE_{10} for each of the *L* MOVE.1 regressions, along with the *L* values of *%SEE* needed to estimate the *L* weights, are provided in the following sections.

Table 2. Percent standard errors of estimate at index stations used to weight flow statistics at partial-record station 01396190, South Branch Raritan River at Four Bridges, N.J.

[ID, identifier; %SEE, percent standard error of estimate]

Index Station ID	%SEE
01380450	26.9
01398500	35.8
01440000	35.5
01443500	33.5
01445500	28.2

Standard Error of Prediction for the 7Q10 Flow Statistic

The *%SEE* depends, in part, on the standard error of prediction (*SEP*), a measure of the variability in measured flows about the MOVE.1 regression line not caused by systematic error or bias. The *SEP* depends on the number of measured flows used in the regression, how far the regression relation is extended beyond measured flows to make predictions of flows or flow statistics at the PR station, and the standard error of index station statistics (Telis, 1991). The *SEP* for the 10-year recurrence interval associated with the annual 7Q10, or SEP_{10} , is determined using the equation

$$SEP_{10} = \left[SEE_{ols}^{2} \left(1 + \frac{1}{N_1} + \frac{\left(X_{10} - m(x_1) \right)^2}{\sum_{i=1}^{N_1} (x(i) - m(x_1)^2} \right) + \frac{S_y}{S_x}^{2} (1-r)^2 \left(x(i) - m(x_1)^2 \right) \right]^{0.5} \tag{9}$$

where SEP_{10} is the standard error of prediction in log units using the MOVE.1 regression for flows with 10-year recurrence interval (Telis, 1991), SEE_{ols} is the standard error of estimate in log units estimated from ordinary least-squares regression, N_1 is the number of concurrent discharge measurements at the PR and index stations used for the regression, X_{10} is the logarithm of the 10-year low-flow statistic at the index station, $m(x_1)$ is the mean of the logarithms of the N_1 concurrent flows for the index station, $x(i)$ is the logarithm of the *i*th discharge for the index station, S_y is the standard deviation of log-transformed measured flows at the PR station, S_x is the standard deviation of log-transformed measured flows at the index station, and r is the correlation coefficient for PR flows regressed on index station flows using OLS regression.

Time-Sampling Error for the 7Q10 Flow Statistic

In addition to being dependent on the 7Q10 *SEP*, the *%SEE* for the annual 7Q10 is a function of time-sampling errors attributed to the use of instantaneous flow measurements to derive flows of finite duration. Time-sampling errors, sometimes known as gage errors, are errors that arise when flow statistics for a given frequency and duration are estimated on the basis of computed mean daily flows. The accuracy of mean daily flows is a function of the accuracy of streamflow measurements, rating curves, and computation methods. Because the annual 7Q10 metric and flow-duration statistics predicted by the MOVE.1 regression line are derived using daily flows averaged over long periods of time, errors associated with such flow statistics will tend to be smaller than errors associated with daily flow measurements.

The time-sampling error for the annual 7Q10 is estimated using the equation developed by Kite (1988):

$$SE_{10} = \frac{S}{\sqrt{N}} \delta \tag{10}$$

where SE_{10} is the time-sampling error in logarithmic units, S is the standard deviation of log-transformed annual 7-day low flows, N is the number of years in the annual 7-day low-flow record, and δ is a function of the skew of the distribution of annual 7-day flows. Greater variability in 7Q10 estimates, a shorter period of record, and a highly skewed distribution of 7-day flows will produce large time-sampling errors. The time-sampling error can be converted to percent sampling error using the relation

$$\%SE_{10} = 100 \left[10^{2.3\, SE_{10}^{2}} - 1 \right]^{0.5} . \tag{11}$$

Standard Error of Estimate for the 7Q10 Flow Statistic

The *SEE* for flows with a 10-year recurrence interval is computed using the values of SEP_{10} and the SE_{10} , based on the equation for SEE_{10} provided by Telis (1991):

$$SEE_{10} = \left[SEP_{10}^{2} + \left(\frac{S_y}{S_x} \right)^2 SE_{10}{}^{2} \right]^{0.5} . \tag{12}$$

Use of the reciprocal of SEE_{10} to define index station weights for the weighted average of a flow statistic at the PR station disproportionately penalizes flow information at index stations that have larger flows because these larger flows generally have larger standard errors of estimate. To account for the fact that regressions involving greater flows tend to have larger values of SEE_{10}, the percent standard errors of estimate used to define values of ω_j in equation 8 were estimated as

$$\%SEE_{10} = 100\left[10^{2.3\,SEE_{10}^{2}} - 1\right]^{0.5} . \qquad (13)$$

$\%SEE_{10}$ values were used to define weights ω_j , $j = 1, 2, \ldots, L$ needed to calculate the average of flow statistics predicted across all L MOVE.1 regressions, regardless of the actual flow statistic being predicted. Values of ω_j for each PR station were standardized by summing them across all L index-station regressions, and dividing each weight by the sum before using them to weight each of the MOVE.1 predictions associated with the PR station.

Use of $\%SEE_{10}$ to define weights for arbitrary flow statistics predicted at PR stations can be justified if the *%SEE* associated with the flow statistic, normalized to sum to 1, is approximately equal to the percent standard error for prediction of the annual 7Q10. Under conditions of small variance in base flows, low-flow statistics derived for different durations and recurrence intervals will tend to become more similar, with percent standard errors of estimate that approach $\%SEE_{10}$ and index-station weights that are approximately equal to those estimated for the annual 7Q10. Future improvements in the R script may include explicit derivation of percent standard errors of estimate for low-flow statistics of arbitrary duration and recurrence, with index-station weights customized to the particular flow statistic of interest.

Predicted Flow Statistics

The MOVE.1 R script computes the same low-flow frequency statistics as the legacy SAS code. Flow statistics and regression parameters requested by NJDEP for permitting purposes and output by the R script include low-flow frequency statistics, arithmetic and harmonic means of daily flows, the standard error of the regression, and the Kendall tau trend statistic (table 3). Values of these statistics for all index stations associated with the PR station are read from a static file of flow statistics for New Jersey PR stations published in Watson and others (2005). These published flow statistics, compiled using daily streamflows through water year 2001, are used to predict equivalent statistics at the PR station on the basis of the L MOVE.1 regressions, weighted by normalized reciprocals of $\%SEE_{10}$ and summed to obtain a weighted average for the flow statistic at the PR station.

Table 3. Flow statistics and regression metrics output by the R script for MOVE.1.

[%, percent; ft^3/s, cubic foot per second; mi^2, square mile; yr, year]

Variable	Description
Q1.10	Annual 1-day 10-yr (1Q10) low flow (ft^3/s)
Q7.10	Annual 7-day 10-yr (7Q10) low flow (ft^3/s)
Q7.10.DA	Annual 7-day 10-yr (7Q10) low flow divided by drainage area (ft^3/s/mi^2)
WIN.Q7.10	Winter 7-day 10-yr (7Q10) low flow (ft^3/s)
Q30.5	Annual 30-day 5-yr (30Q5) low flow at the index station (ft^3/s)
Q30.10	Annual 30-day 10-yr (30Q10) low flow at the index station (ft^3/s)
WIN.Q30.5	Winter 30-day 5-yr (30Q5) low flow at the index station (ft^3/s)
WIN.Q30.10	Winter 30-day 10-yr (30Q10) low flow at the index station (ft^3/s)
QAVG	Arithmetic daily mean flows (ft^3/s)
HARMEAN	Harmonic mean of daily flows (ft^3/s)
DURA.25	Daily flow exceeded 25% of the time (ft^3/s)
DURA.50	Daily flow exceeded 50% of the time (ft^3/s)
DURA.75	Daily flow exceeded 75% of the time (ft^3/s)
PERCENT.PRED	Standard error of MOVE.1 prediction (%)
PERCENT.GAGE	Time-sampling, or gage, error (%)
PERCENT.EST	Standard error of MOVE.1 estimate (%)
CORRELATION	MOVE.1 correlation coefficient
POWER	Slope of MOVE.1 regression line
CONSTANT	Intercept of MOVE.1 regression line
KENDALLTAU	Kendall trend test statistic for MOVE.1 residuals
KENDALLPROB	Kendall trend test significance level for MOVE.1 residuals

Low-flow frequency statistics estimated at the PR station include the 1-day, 7-day, and 30-day low-flow frequencies; 25-percent, 50-percent, and 75-percent flow durations;

Table 4. Percent differences in estimated MOVE.1 low-flow statistics between the legacy code and the new R script.

[Variables are defined in table 3]

Variable	Q1.10	Q7.10	WIN.Q7.10	Q30.5	Q30.10	WIN.Q30.5	WIN.Q30.10	HARMEAN	DURA.25	DURA.50	DURA.75
Minimum percent difference	−3.67	−3.61	−1.27	−1.57	−2.35	−3.99	−1.79	−1.52	−2.79	−1.25	−1.53
Maximum percent difference	2.65	3.70	1.56	2.08	2.37	0.55	1.26	0.47	0.93	0.54	0.89

Table 5. Minimum and maximum percent differences in regression metrics between the legacy code for MOVE.1 and the new R script for MOVE.1.

[Variables are defined in table 3]

Variable	PERCENT.PRED	PERCENT.GAGE	PERCENT.EST	CORRELATION	POWER	CONSTANT	KENDALLTAU	KENDALLPROB
Minimum percent difference	−16.76	−0.47	−16.78	−0.41	−0.26	−4.30	−7.57	−12.14
Maximum percent difference	0.37	1.12	0.44	0.35	0.63	18.81	9.25	13.26

25-percent/75-percent flow duration ratio; annual mean flow; and harmonic mean flow. The 7-day and 30-day low-flow frequencies at the PR station are computed for annual and winter periods, with the winter period extending from November through March. The Kendall Tau statistic is of particular importance because it is used to determine the likelihood that significant trends in flow residuals occur over time, a situation that would give rise to non-stationary conditions and potential bias in predictions of flow statistics at PR stations. All flow statistics and regression metrics are output to a .csv file to facilitate further analysis.

In comparing the output of the R script to the legacy code, some minor differences were observed. As seen in table 4, computed differences in estimated flow statistics are less than 4 percent. Differences in output generally stem from the dissimilar ways that the legacy code and the R script handle data, including numerical rounding and differences in the streamflow values selected for the regression. These differences can alter results of the MOVE.1 regression enough to produce large percent differences in estimated flow statistics, particularly in lower ranges of flow. The table does not include differences for estimated flow statistics of less than 0.1 ft^3/s because the percent difference between estimates output by the legacy code and the R script can become large at such small values when estimated flow statistics made using log-transformed flows are back-transformed. Another reason for not including differences for estimates less than the 0.1 ft^3/s threshold is that estimates below 1 ft^3/s are typically rounded to the nearest 0.1 ft^3/s. Because this practice alters the difference between estimated flow statistics output by the MOVE.1 legacy program and statistics output by the MOVE.1 R script, it makes little sense to consider differences less than the 0.1 ft^3/s threshold.

Several factors contribute to the slight differences in flow estimates produced by the MOVE.1 legacy code and the new MOVE.1 R script. The principal factor contributing to differences in estimated flows can be largely attributed to differences in rounding that occur during input of flows. The legacy code retrieves mean daily flows directly from the NWIS database as unrounded values with a floating decimal, whereas the R script reads mean daily flows from NWISweb and is limited to three significant figures. A secondary factor contributing to differences in output is associated with errors discovered in the legacy code. In the process of converting from the legacy code to R, it was determined that the legacy code was not properly screening data when more than one measurement was made at a PR station during a single day. The legacy code failed to apply base-flow tests to subsequent measurements made on the same day and automatically included all daily measurements in the regression, whereas the new R script for MOVE.1 correctly imposes base-flow tests on each measurement made on the same day. Lastly, the new Aquarius database system has the ability to recognize data entry errors and, as a safeguard, does not include those measurements in the output file read by the new R script. For example, if any part of a flow record is considered to be in error as determined by the business rules of the new database, that measurement record is not included in the NWIS output file, and is therefore not read by the R script. To illustrate, if the start time of the flow measurement is incorrectly entered as beginning before the time of the site visit, that measurement is filtered from the output file of the database. Because no such safeguards were in place when the legacy code was run, any and all data input to the NWIS database were included in the MOVE.1 regression, regardless of whether the data were incorrectly input to the database.

Percent differences between regression metrics from the R script and the legacy code were as high as nearly 19 percent, as shown in table 5. Errors in the percent predicted (*PERCENT.PRED*) and percent estimated (*PERCENT.EST*) calculations are consistently high. These differences are a reflection of the differences in estimated flows caused by one or more of the three factors mentioned above. Large differences in the regression intercept (CONSTANT) and Kendall probability (KENDALL.PROB) were caused by the very small values used to compute them. Overall, the R script reproduced low-flow estimates output by the legacy code, and when it didn't, differences could generally be traced to the three principal factors mentioned above.

Flow Chart

The flow chart in figure 7 graphically illustrates the order of operations used by the R script to read PR station identifiers and associated index station identifiers; read base flows and daily flows; select flow records according to the sequence of base-flow tests; perform OLS regression and MOVE.1, censored MOVE.1, or piecewise MOVE.1 regressions; compute standard errors of estimate at each index station; calculate the weighted average of flow statistics at the PR station; make predictions of flow statistics at PR stations on the basis of regression results; and generate pdf files containing plots of regression lines and residuals for each index station. Finally, the R script outputs weighted averages of flow statistics predicted at the PR station to a comma delimited .csv file.

Summary

To accommodate changes in software used to maintain and serve data from the U.S. Geological Survey (USGS) National Water Information System (NWIS) database, the USGS successfully migrated SAS legacy code to estimate flows at partial-record (PR) stations to an R script. The R script performs MOVE.1 regressions to predict low-flow statistics at a PR station, based on daily flows measured at index stations believed to share similar streamflow characteristics. Low-flow statistics at the PR stations are used in support of the New Jersey Department of Environmental Protection (NJDEP) surface-water permitting process. MOVE.1 regression extends the record at a PR station by preserving the mean and variance of measured and predicted low flows. The R script can be executed on various computing platforms, including those with UNIX, Windows, and macOS operating systems.

The R script begins execution by reading a file containing the station identifiers of the PR station and all index stations considered to be hydrologically similar to the PR station. The R script then reads instantaneous base flows at the PR station and daily flows at one or more index stations, screens flows to ensure that all flow records meet a series of base-flow criteria, transfers flow information from index stations to the PR station by performing MOVE.1 regressions on concurrent flows, predicts low-flow statistics at the index stations on the basis of their regression lines, computes index station weights as reciprocals of the percent standard errors of estimate for index station regressions, estimates low-flow statistics at the PR station as the weighted sum of low-flow statistics estimated at multiple index stations, generates pdf files containing regression and residual plots, and outputs predicted low-flow statistics and their weighted averages to a .csv text file that can easily be imported to a spreadsheet. Low-flow statistics output by R script include the annual 7Q10, defined as the lowest annual 7-consecutive-day flow that occurs, on average, during a 10-year period. The annual 7Q10 provides the foundation for defining passing-flow requirements in New Jersey and is used to support NJDEP surface-water permitting decisions within the State.

The R script replicates MOVE.1 output generated by the legacy code and adds functionality by introducing new regression options for the user. These new options include censored MOVE.1 regression, which calls the USGS-developed censMove.1 R function to perform regression when zero-valued flows occur at index or PR stations. A piecewise MOVE.1 regression option was also added to be used where measured base flows at the PR station fail to exhibit statistically homogeneous behavior across the range of measured index station flows. The piecewise regression option permits two regression lines to be identified by iteratively changing the slopes and intercepts of the MOVE.1 regression line segments until the sum of squared error is minimized.

Equations used to derive the standard error of prediction, the time-sampling error, and the percent standard errors of estimate in the legacy code were reproduced in the R script. The standard errors of estimate provide measures of the degree of uncertainty in estimated low-flow statistics made at each index station using its associated regression line. Normalized reciprocals of the percent standard errors of estimate are used to weight terms in the average of low-flow statistics predicted at each of the index stations. The MOVE.1 regression methodology is used by the NJDEP to predict the annual 7Q10 at PR stations, using index station weights that are based on the 7Q10 low-flow statistic. Future improvements to the R script may include development of a formal framework for estimating other low-flow statistics at PR stations that explicitly calculates weights for the particular t-year recurrence statistic of interest, as well as development of a piecewise regression approach that doesn't require the user to enter an index flow interval where the breakpoint appears likely to be located.

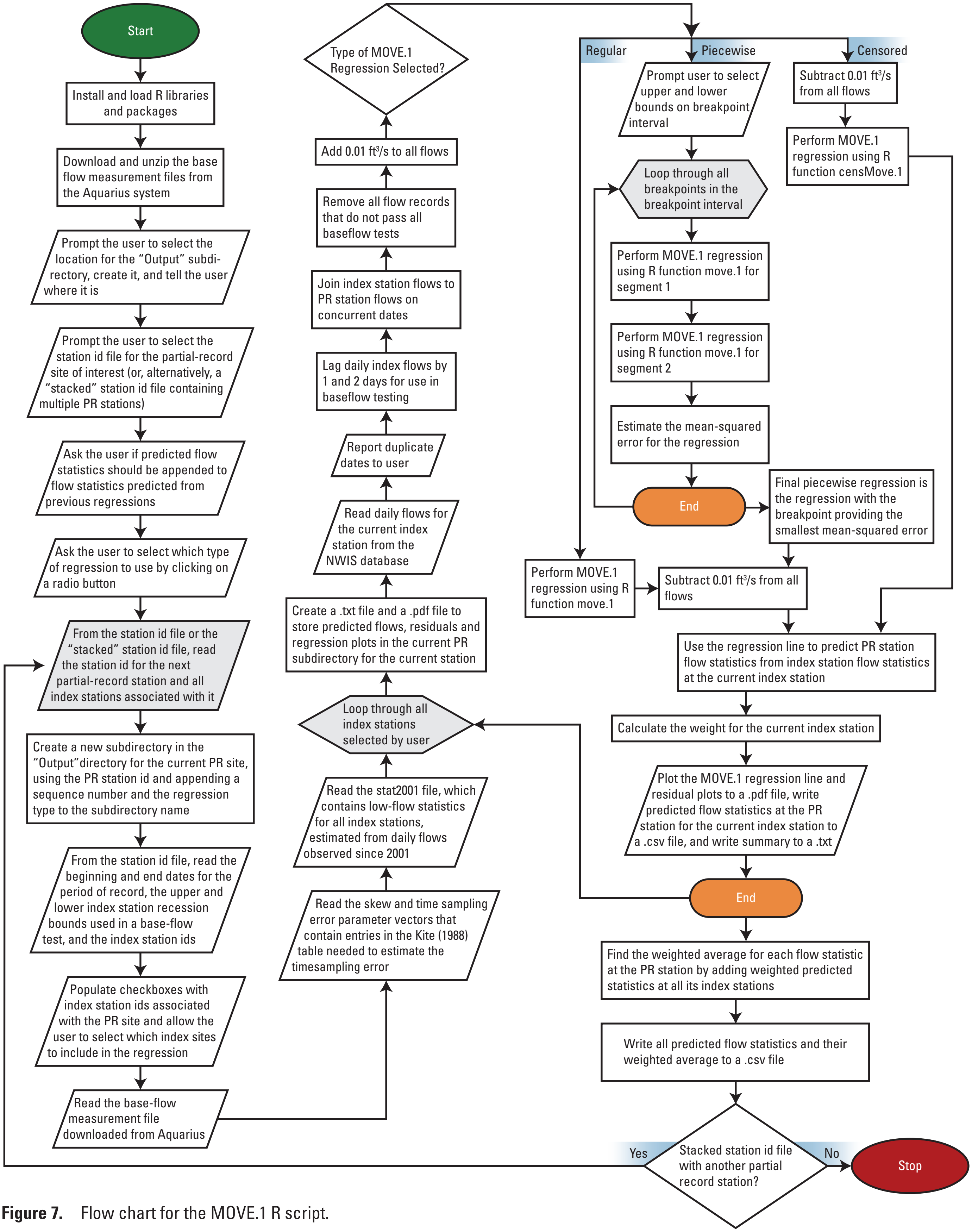

Start
Install and load R libraries and packages
Download and unzip the base flow measurement files from the Aquarius system
Prompt the user to select the location for the "Output" subdirectory, create it, and tell the user where it is
Prompt the user to select the station id file for the partial-record site of interest (or, alternatively, a "stacked" station id file containing multiple PR stations)
Ask the user if predicted flow statistics should be appended to flow statistics predicted from previous regressions
Ask the user to select which type of regression to use by clicking on a radio button
From the station id file or the "stacked" station id file, read the station id for the next partial-record station and all index stations associated with it
Create a new subdirectory in the "Output"directory for the current PR site, using the PR station id and appending a sequence number and the regression type to the subdirectory name
From the station id file, read the beginning and end dates for the period of record, the upper and lower index station recession bounds used in a base-flow test, and the index station ids
Populate checkboxes with index station ids associated with the PR site and allow the user to select which index sites to include in the regression
Read the base-flow measurement file downloaded from Aquarius
Read the skew and time sampling error parameter vectors that contain entries in the Kite (1988) table needed to estimate the timesampling error
Read the stat2001 file, which contains low-flow statistics for all index stations, estimated from daily flows observed since 2001
Loop through all index stations selected by user
Create a .txt file and a .pdf file to store predicted flows, residuals and regression plots in the current PR subdirectory for the current station
Read daily flows for the current index station from the NWIS database
Report duplicate dates to user
Lag daily index flows by 1 and 2 days for use in baseflow testing
Join index station flows to PR station flows on concurrent dates
Remove all flow records that do not pass all baseflow tests
Add 0.01 ft³/s to all flows
Type of MOVE.1 Regression Selected?
Regular
Piecewise
Censored
Prompt user to select upper and lower bounds on breakpoint interval
Loop through all breakpoints in the breakpoint interval
Perform MOVE.1 regression using R function move.1 for segment 1
Perform MOVE.1 regression using R function move.1 for segment 2
Estimate the mean-squared error for the regression
End
Final piecewise regression is the regression with the breakpoint providing the smallest mean-squared error
Subtract 0.01 ft³/s from all flows
Perform MOVE.1 regression using R function censMove.1
Perform MOVE.1 regression using R function move.1
Subtract 0.01 ft³/s from all flows
Use the regression line to predict PR station flow statistics from index station flow statistics at the current index station
Calculate the weight for the current index station
Plot the MOVE.1 regression line and residual plots to a .pdf file, write predicted flow statistics at the PR station for the current index station to a .csv file, and write summary to a .txt
End
Find the weighted average for each flow statistic at the PR station by adding weighted predicted statistics at all its index stations
Write all predicted flow statistics and their weighted average to a .csv file
Stacked station id file with another partial record station?
Yes
No
Stop

Figure 7. Flow chart for the MOVE.1 R script.

(PR, partial-record; ft³/s, cubic feet per second)

Acknowledgments

Special thanks go to Dave Lorenz of the USGS for developing the move.1, censMove.1 , and predict.move.1 R functions required to perform MOVE.1, censored MOVE.1, and piecewise MOVE.1 regressions invoked by the R script and to estimate flow statistics based on results of the regressions (https://github.com/USGS-R/smwrStats/blob/master/R/move.1.R, https://github.com/USGS-R/smwrQW/blob/master/R/censMove.1.R). We appreciate insights offered by R. Ward Hickman (USGS) relating to the workings of the legacy code that provided the basis for the R script. Thanks are also extended to Will Farmer (USGS), who verified the equation used to estimate the percent standard error of estimate needed to calculate weights for the weighted average of low-flow statistics at index stations.

References Cited

Crawley, M.J., 2013, The R book: Chichester, UK, John Wiley and Sons, Ltd., 1076 p.

Gillespie, B.D., and Schopp, R.D, 1982, Low-flow characteristics and flow duration of New Jersey streams: U.S Geological Survey Open-File Report 81–1110, 164 p. [Also available at https://pubs.usgs.gov/of/1981/1110/report.pdf.]

Halfon, E., 1985, Regression method in ecotoxicology: a better formulation using the geometric mean functional regression: Environmental Science and Technology, v. 19, iss. 8, p. 747–749.

Hartung, J., Knapp, G., and Sinha, B.K., 2008, Statistical meta-analysis with applications: Hoboken, N.J., John Wiley and Sons, Ltd., 248 p.

Helsel, D.R, and Hirsch, R.M., 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 510 p.

Hirsch, R.M., 1982, A comparison of four streamflow regression techniques: Water Resources Research, v. 18, no. 4, p. 1081–1088.

Hirsch, R.M., and Gilroy, E.J., 1984, Methods of fitting a straight line to data: examples in water resources: Journal of the American Water Resources Association, v. 20, p. 705–711.

Hoffman, J.L., and Domber, S.E., 2013, History of passing flows in New Jersey, with contemporary and future applications: New Jersey Geological and Water Survey Open-File Report OFR 13–1, 62 p.

Journel, A.G., and Huijbregts, C.J., 1978, Mining geostatistics: San Diego, Calif., Academic Press, 600 p.

Kite, G.W., 1988, Frequency and risk analyses in hydrology: Littleton, Colo., Water Resources Publications, 257 p.

Kritskiy, S.N., and Menkel, M.F., 1968, Some statistical methods in the analysis of hydrologic data: Soviet Hydrology Selected Papers 1, p. 80–88.

Matalas, N.C., and Jacobs, B., 1964, A correlation procedure for augmenting hydrologic data: U.S. Geological Survey Professional Paper 434–E, 7 p.

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

Searcy, J.K., 1959, Flow-duration curves, Manual of hydrology—Part 2. Low-flow techniques: U.S. Geological Survey Water-Supply Paper 1542–A, 33 p.

State of New Jersey, Department of Environmental Protection, 2011, Surface water quality standards: New Jersey Administrative Code, Title 7, Chapter 9B, 113 p., accessed December 22, 2017, at http://www.nj.gov/dep/rules/rules/njac7_9b.pdf.

State of New Jersey, Department of Environmental Protection, Division of Water Quality, 2015, NJPDS Rules N.J.A.C. 7:14A, accessed December 22, 2017, at http://www.state.nj.us/dep/dwq/714a.htm.

Stedinger, J.R., and Thomas, W.O., Jr., 1985, Low-flow frequency estimation using base-flow measurements: U.S. Geological Survey Open-File Report 85–95, 22 p.

Telis, P.A., 1991, Low-flow and flow-duration characteristics of Mississippi streams: U.S Geological Survey Water-Resources Investigations Report 90–4087, 214 p.

Watson, K.M., Reiser, R.G., and Schopp, R.D., 2005, Streamflow characteristics and trends in New Jersey, water years 1897–2003: U. S. Geological Survey Scientific Investigations Report 2005–5105, 131 p.

For additional information, contact:
Director, New Jersey Water Science Center
U.S. Geological Survey
3450 Princeton Pike, Suite 110
Lawrenceville, NJ 08648

or visit our website at:
https://nj.usgs.gov/

Publishing support provided by the West Trenton Publishing Service Center

Colarullo and others—**Implementation of MOVE.1, Censored MOVE.1, and Piecewise MOVE.1 Low-Flow Regressions**—Open-File Report 2018–1089

ISSN 2331–1258 (online)
https://doi.org/10.3133/ofr20181089