A decision support tool to estimate unregulated, daily streamflow at ungauged sites in the Connecticut River Basin, northeast United States

S. A. Archfield[1], P. A. Steeves[2], J. D. Guthrie [3], and K. G. Ries III[4]

---

[1] S. A. Archfield, Massachusetts-Rhode Island Water Science Center, U.S. Geological Survey, 10 Bearfoot Road, Northborough, MA 01532, USA, sarch@usgs.gov, p: +1.508.490.5072, f: +1.508.490.5068 *(Corresponding Author)*

[2] P. A. Steeves, Massachusetts-Rhode Island Water Science Center, U.S. Geological Survey, 10 Bearfoot Road, Northborough, MA 01532, USA, psteeves@usgs.gov

[3] J. D. Guthrie, Rocky Mountain Geographic Science Center, U.S. Geological Survey, P.O. Box 25046 MS 516, Denver Federal Center, Denver, CO 80225, USA, jdguthrie@usgs.gov

[4] K. G. Ries III, Office of Surface Water, U.S. Geological Survey, 5522 Research Park Drive, Baltimore, Maryland 21228, USA, kries@usgs.gov

Abstract.

Streamflow information is needed for any number of hydrologic applications. Because most stream reaches are ungauged, this information is commonly needed for rivers that have no readily available measurements of streamflow. In the Connecticut River Basin, dam operation and its effects on the aquatic habitat are of particular interest. Here, daily streamflow is needed for use as input to dam simulation and optimization models as well as to develop ecological-flow prescriptions for rivers and streams. To provide a common scientific foundation for water allocation decisions, a freely available and easy-to-use software tool termed the Connecticut River UnImpacted Streamflow Estimation (CRUISE) tool was developed to estimate a daily streamflow time series at ungauged locations in the Connecticut River Basin. CRUISE is used in conjunction with the U.S. Geological Survey StreamStats web application. Through the coupling of CRUISE and StreamStats, users are able to point and click on a stream location of interest and obtain a delineated catchment as well as a daily time series of streamflow. Daily streamflow was shown to be reliably estimated by the CRUISE tool, with efficiency values between the observed and estimated streamflows ranging from 0.69 to 0.92 and ecologically-relevant streamflow statistics derived from the estimated daily streamflow to be generally within +/- 10 percent of the streamflow statistics computed from the observed daily streamflow values.

Keywords: decision support; ungauged; ungaged; streamflow; water availability; basin delineation; water resources

1. Introduction

Streamflow information at ungauged stream reaches is needed for any number of hydrologic applications. Furthermore, when streamflow information is presented in easy-to-use, freely-available software tools, this information can provide a scientific framework for water-allocation negotiation amongst stakeholders. Unfortunately, many rivers of the world are not gauged and streamflow information is not available where it is often needed to make informed water-allocation decisions. Additionally, there has been increasing emphasis on the need for daily streamflow time series to understand the response of ecology to river regulation and develop streamflow prescriptions to restore and protect aquatic habitat [*Poff et al.*, 1997]. For these reasons, a software tool was developed to estimate daily streamflow time series at ungauged streams in the Connecticut River Basin (CRB), located in the northeast United States. The Connecticut River UnImpacted Streamflow Estimation (CRUISE) software tool is based on a geographic information system (GIS) that allows users to point and click on an ungauged stream location of interest in the CRB. The tool then delineates a contributing area to the stream location and estimates a daily streamflow time series.

The CRB has thousands of dams along the mainstem and tributary rivers that are used for hydropower, flood control, and water supply just as the CRB is home to a number of important fish species that rely on the river for all or part of their life cycle. These competing interests for water led the Army Corps of Engineers and The Nature Conservancy to embark on a partnership to understand how dam management can be optimized to meet both human and ecological needs for water. To answer this question, daily streamflow time series are needed at locations in the CRB that have ecological constraints on water (locations where important or protected fish or ecological communities reside or rely on for life), human constraints on water (locations on the river that are dammed or otherwise managed), or have both constraints to consider. Often times, these locations are unmonitored.

Methods to estimate daily streamflow time series at ungauged locations can be broadly characterized under the topic of regionalization [*Blöschl and Sivapalan*, 1995], an approach which pools information about streamgauges in a region and transfers this information to an ungauged location. Generally there are two main categories of information that is pooled and trasnferred: 1) rainfall-runoff model parameters that are calibrated at gauged catchments and transferred in some way to an ungauged location [see *Zhang and Chiew*, 2009 for a review] and 2) gauged streamflows, or related streamflow properties, are directly transferred to ungauged locations. Examples of this type of regionalization approach include geostatistical methods such as top-kriging [*Skøien and Blöschl*, 2007] and more commonly used methods such as the drainage-area ratio method as described in *Archfield and Vogel* [2010], the MOVE methods [*Hirsch*, 1979], which are primarily used to patch and extend missing daily streamflow information in existing records but can be easily extended to ungauged methods, and a non-linear spatial interpolation method, applied by *Fennessey* [1994], *Hughes and Smakhtin* [1996], *Smakhtin* [1999], *Mohamoud* [2008], and *Archfield et al.* [2010]. For the software tool presented in this paper, a hybrid approach combining the drainage-area ratio and non-linear spatial interpolation methods is used to estimate daily streamflow time series.

Software tools to provide streamflow time series at ungauged locations have been previously published for fixed catchments. *Smakhtin and Eriyagama* [2008] and *Holtschlag* [2009] introduced software tools to provide monthly streamflows for ecological streamflow assessments around the globe and in the Great Lakes region of the United States, respectively. *Williamson et al.* [2009] developed The Water Availability Tool for Environmental Resources (WATER) to serve daily streamflow information at fixed stream locations in non-karst areas of Kentucky. These existing tools provide valuable streamflow information, and in most cases at the monthly – not daily – time step for fixed catchments; yet, often the locations of most interest on a river are not coincident with the pre-defined hydrologic units presented in these tools. There are few – if any – software tools that can provide daily streamflow for user-specified (unfixed) locations within a region. The U.S. Geological Survey StreamStats tool [*Ries and others*, 2008] provides the utility to delineate a contributing area to a user-selected location on a river; however, only streamflow statistics – not streamflow time series – are provided for the ungauged location.

The CRUISE tool is one of the first such tools to combine the utility of catchment delineation at any location along a stream with the estimation and serving of daily streamflow information. *Archfield et al.* [2010] developed a GIS-based software tool to estimate daily streamflow; however, this tool requires software and licensing not available to all users, and covers only the state of Massachusetts. This paper extends the work of *Archfield et al.* [2010] and presents the first such software tool to obtain daily streamflow time series at ungauged locations in a regional framework that requires only the use of an internet connection and a fairly ubiquitous spreadsheet program. This framework has the potential to be applied to other regions to provide daily streamflow information for ungauged locations.

This paper first describes the study area and the data required by the software tool. The underlying methods to estimate daily streamflow time series in the software tool are then presented and the software tool and functionality are described. Lastly the utility of the software tool to provide reliable estimates of daily streamflow is demonstrated.

2. Study area and data

The study area is located in the northeast United States and covers an area of approximately 29,000 km^2 (fig. 1). The region is characterized by a temperate climate with distinct seasons. Snowfall is common from December through March, with generally more snowfall falling in the northern portion of the CRB. The geology and hydrology of the study region is heavily affected by the growth and retreat of glaciers during the last ice age, which formed the present-day stream network and drainage patterns [*Armstrong et al.*, 2008]. The retreat of the glaciers filled the river valleys with outwash sands and gravel as well as fine- to coarse-grained lake deposits [*Armstrong et al.*, 2008], and these sand and gravel deposits have been found to be important controls on the magnitude and timing base flows in the southern portion of the study region [*Ries and Friesz*, 2000].

Data from streamgauges located within the CRB and surrounding area are used in the CRUISE tool to estimate daily streamflow time series at ungauged locations (fig. 1; table 1). The 63 study streamgauges have at least 20 years of daily streamflow record and have minimal regulation in the contributing catchment to the streamgauge [*Armstrong et al.*, 2008; *Falcone et al.*, 2010]. Previous work in the southern portion of the study area by *Archfield et al.* [2010] showed that the contributing area to the streamgauge, percent of the contributing area with surficial sand and gravel deposits, and mean annual precipitation values for the contributing area are important variables in modeling streamflows at ungauged locations. For this reason, these characteristics were summarized for the study streamgauges (fig. 2) and used in the streamflow estimation process. Contributing area to the study streamgauges ranges from 0.5 km^2 to 1,845 km^2 with a median value of 200 km^2 (fig. 2A). Mean annual precipitation ranges from 101 cm per year to 157 cm per year with a median value of 122 cm per year (fig. 2B).Percent of the contributing area with surficial sand and gravel ranges from 0 percent to 67 percent with a median value of 9.5 percent (fig. 2C).

3. Methods underlying the software tool

Streamflow in the CRUISE tool is estimated for a 44-year (16,071-day) period of record spanning October 1, 1960 through September 30, 2004 using information from an index streamgauge and catchment characteristics computed for the contributing area to the ungauged stream location of interest (fig. 3). Catchment characteristics and the selected index streamgauge are used to first estimate a continuous, daily flow-duration curve (FDC) for the 44-year simulation period (fig 3). The estimated FDC at the ungauged location is then transformed to a time series of streamflow values by the index streamgauge (fig. 3). The methods to estimate the FDC, select the index streamgauge, and transform the FDC to a time series of daily streamflow are explained in detail in the following sections.

3.1 Estimation of the period-of-record flow-duration curve

Estimation of the period-of-record FDC at an ungauged location remains an outstanding challenge in hydrology. *Castellarin et al.* [2004] provides a review of several methods to estimate FDCs at ungauged locations and found that no particular method was consistently reliable. For this study, an empirical, piece-wise approach to estimate the period-of-record FDC is used in the CRUISE tool (fig. 4). This overall approach is similar to that used by *Mohamoud* [2008] and *Archfield et al* [2010] in that the FDC is estimated by first developing regional regressions relating catchment characteristics to selected FDC quantiles and then interpolating between those quantiles to obtain a continuous FDC.

Streamgauges having at least 20 years of daily streamflow record that also contain the drought of record for the study area were used for this portion of the CRUISE tool development. A total of 52 streamgauges fit these criteria (fig. 1). Streamflows at each of these streamgauges were ranked and corresponding exceedence probabilities were determined using the Wiebull plotting position [*Stedinger et al.*, 1993]. Selected streamflow quantiles were then determined by applying equation 2 presented in *Vogel and Fennessey* [1994].

With the exception of streamflows having less than or equal to a 0.01 probability of being exceeded (streamflows with a probability of being exceeded more than 1 percent of the time), selected quantiles on the FDC are estimated from explanatory variables (fig. 4) and a continuous FDC is log-linearly interpolated between these quantiles to obtain a continuous FDC (fig. 4). Relations between streamflow quantiles at the 0.02, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8 and 0.85 exceedence probabilities were estimated by independently regressing each streamflow quantile against catchment characteristics (table 2; fig. 4). Following the approach in *Archfield et al.* [2010], relations between streamflow quantiles at the 0.9, 0.95, 0.98, 0.99 and 0.999938 were estimated by regressing streamflows at these quantiles against one another and using these relations to recursively estimate streamflows (table 3; fig. 4). Recursively estimating low streamflows, as was done in *Archfield et al.* [2010], exploits the strong structural relation between the streamflow quantiles (fig. 4) and enforces the constraint that streamflows must decrease as the exceedence probability increases. *Mohamoud* [2008] and *Archfield et al.* [2010] observed that when regression is done against catchment characteristics, there is increased potential for the estimated quantiles to violate the constraint that streamflows must decrease as the exceedence probability increases because the uncertainty in the flow estimates is greatest at the lowest portion of the FDC. Regressing quantiles against one another ensures that this constraint is not violated. This is an alternative approach to that used by *Mohamoud* [2008], who suggested discarding any estimated quantiles that violate the constraint.

All regressions were fit using the natural logarithms of the dependent and independent variables. Retransformation was needed to obtain estimated streamflows in their original units of measure The Smearing adjustment [*Duan*, 1983] was used to eliminate the bias in the streamflow estimates caused by the retransformation. Following the regression-screening protocol described in *Archfield et al.* [2010], independent variables having coefficient that were statistically significant at the 0.05 level were included in the final equations. For regression equations with multiple significant independent variables, muticollinarity was evaluated using the variance inflation factor [*Stedinger et al.*, 1993]. Independent variables with a variance inflation factor higher than 2.5 were removed from the regression equations. Residuals were evaluated for normality and streamgauges that exerted a high degree of influence on the regression were omitted. Percent root-mean-square error and the Nash-Sutcliffe efficiency value [*Nash and Sutcliffe*, 1970] were computed from a leave-one-out cross validation of each site used in the development of the regression equations. Goodness-of-fit metrics, regression diagnostics and coefficients, as well as explanatory variables for each quantile, are shown in tables 2 and 3.

Arcfield et al. [2010] showed that estimated streamflows determined bylog-linear interpolation between exceedence probabilities of 0.01 or less do not match the shape of the FDC in this range and this interpolation method creates a bias in the estimated streamflows, which can substantially overestimate the peak streamflows. The shape of the FDC at the highest streamflows is so complex that, instead of using another interpolation method, the CRUISE tool uses scaled streamflows from an index streamgauge to estimate the highest streamflows at the ungauged location. The assumption here is that the shape of the left tail of the FDC is better approximated by the streamflow quantiles at an index streamgauge than by a curve fit. Therefore, for streamflows having less than or equal to a 0.01 probability of being exceeded, streamflows are scaled by a drainage-area approach (eqn. 1) in conjunction with the selected index streamgauge:

$$q_{p_u} = \frac{A_u}{A_g} q_{p_g} \qquad (1)$$

where q_{p_u} is the value of the streamflow quantile at the ungauged location for exceedence proability, p, A_u is the contributing drainage area to the ungauged location, A_g is the contributing drainage area to the index streamgauge, and q_{p_g} is the value of the streamflow quantile at the index streamgauge for exceedence proability, p.

3.2 Selection of the index streamgauge

As shown in figure 3, the index streamgauge is used for two purposes in the CRUISE tool: 1) to estimate streamflows that have less than a 1-percent chance of being exceeded, and 2) to transform the estimated FDC into a time series of streamflow at the ungauged location. The index streamgauge is selected by the map-correlation method [*Archfield and Vogel*, 2010]. The map-correlation method selects the index streamgauge estimated to have the highest cross-correlation between streamflow time series at the index streamgauge and the ungauged location. *Archfield and Vogel* [2010] showed that the selection of the index streamgauge using cross-correlation between streamflow time series outperformed the selection of the nearest index streamgauge when used with the drainage-area ratio method to estimate daily streamflow time series at ungauged locations. This finding supports the use of the map-correlation method in the CRUISE tool for two reasons: 1) the drainage-area ratio approach is also used in the CRUISE tool to estimate streamflows that have less than a 1-percent chance of being exceeded, and 2) because the streamflow time series in the CRUISE tool is constructed by transferring the timing of the streamflows at an index streamgauge to the ungauged location, it follows that one would seek to select the index streamgauge that maximizes the cross-correlation between the streamflows at the ungauged location and the index streamgauge.

Underlying the map-correlation method is a set of variogram models – one for each index streamgauge – that is fitted to the observed cross-correlations between the streamflows at the index streamgauge and each of the other index streamgauges in the study region. The map-correlation method uses these variogram models with ordinary kriging to estimate the cross-correlation between each index streamgauge and the ungauged location and, ultimately, selects the index streamgauge whose streamflows are estimated to be most correlated with the ungauged location of interest.

To develop the variogram models, observed cross-correlations between daily streamflows were computed from a long-term, 30-year common period of record from January 1, 1960 through December 31, 1982 that was available at 45 of the study streamgauges (fig. 1). This 30-year common period was selected to maximize the number of index streamgauges in the CRUISE tool while ensuring the observed cross-correlations were not affected by small or uneven sample sizes. The nonparametric, rank-based Kendall tau correlation measure was used the estimate cross-correlation. Kendall tau estimates the monotonic relation between two variables [*Helsel and Hirsch*, 2002] and, therefore, its application requires fewer assumptions than the Pearson *r* correlation coefficient, which measures only the linear correlation between two variables.

For each index streamgauge, the observed cross-correlation between that streamgauge and each of the other streamgauges is determined. Then, the differences between cross-correlation, formally expressed as the semi-variance [*Isaaks and* S*rivastava*, 1990], taken for each pair of streamgauges were plotted against the separation distance between each of pair of streamgauges. Therefore, for each of the 45 index streamgauges, this results in a plot of $\binom{44}{2}$ points. To discern a relation between the semi-variance and separation distance, the points were placed into 12 bins, with each bin having a length of 200,000 meters. A spherical variogram was fit to the binned values to provide a continuous relation between the semi-variance and separation distance exactly as described by *Archfield and Vogel* [2010] and using the geoR statistical package [*Ribeiro Jr. and Diggle*, 2001]. Variogram models were developed for each of the 45 index streamgauges (table 4) and a leave-one-out cross validation procedure was applied to evaluate the utility of the variogram model to estimate the cross-correlation between the removed site and the index streamgauge upon which the variogram model was fit. The root-mean-square error resulting from this cross validation at each index streamgauge is reported in table 4.

3.3 Generation of streamflow time series

With an index streamgauge and estimated daily FDC at the ungauged location, a time series of daily streamflow for the 44-year simulation period is then constructed by use of the QPPQ transform method [*Fennessey*, 1994; *Hughes and Smakhtin* [1996]; *Smakhtin*, 1999; *Mohamoud*, 2008; *Archfield et al.* 2010]. The term QPPQ-transform method was coined by *Fennessey* [1994]; however, this method has been by published *Smakhtin* [1999], *Mohamoud* [2008], and *Archfield et al.* [2010] under names including “non-linear spatial interpolation technique" [*Hughes and Smakhtin* [1996]; *Smakhtin*, 1999] and "reshuffling procedure" [*Mohamoud*, 2008]. The method assumes that the exceedence probability associated with a streamflow on a given day at the index streamgauge also occurred on the same day as the ungauged location. For example, if the streamflow on October, 1, 1974 was at the 0.9 exceedence probability at the index streamgauge, then it is assumed that the streamflow on that day at the ungauged location also was at the 0.9 exceedence probability . To implement the QPPQ-transform method in the CRUISE tool, a FDC is constructed for the observed streamflows at the index streamgauge, keeping track of the dates associated with each exceedence probability. The exceedence probabilities are then equated between the index streamgauge FDC and the estimated FDC at the ungauged location. The date associated with each exceedence probability at the index streamgauge is then transferred the estimated FDC at the ungauged location.

The QPPQ-transform method requires that each index streamgauge has streamflow values for each day of the simulation period. Recall that the 45 index streamgauges shared a common period of observed daily streamflow record for 30 years of the 44-year simulation period. However, for 20 of the 45 index streamgauges, the observed streamflow record does not cover the full 44-year simulation period (table 1). For these streamgauges, the MOVE3 [*Vogel and Stedinger*, 1985] record extension method was utilized to estimate a complete 44-year period of record. The software program Streamflow Record Extension Facilitator (SREF) [*Granato*, 2009] was used to extend the streamflow records. The QPPQ-transform method uses only the timing of the streamflows at the index streamgauge and not the magnitudes of the streamflows; therefore, the MOVE3-estimated streamflow values themselves are not used in the transfer process.

4. The Connecticut River UnImpacted Streamflow Estimator (CRUISE) software tool

The Connecticut River UnImpacted Streamflow Estimator (CRUISE) tool is freely available for download at http://webdmamrl.er.usgs.gov/s1/sarch/ctrtool/index.html. The CRUISE tool website contains additional information about the software required, a user manual, history of updates, file sizes, and contact information. To use the CRUISE tool, users must have an internet connection, a web browser program, and Microsoft Excel version 2003 or higher. The U.S. Geological Survey StreamStats tool [*Ries and others*, 2008] is first used to delineate the contributing area to the ungauged location and compute the catchment characteristics needed to estimate the FDC, and then the CRUISE tool, which is a customized Microsoft Excel spreadsheet with Visual Basic macros is used to select the index streamgauge and compute the unimpacted daily streamflow time series for the ungauged location.

The StreamStats tool operates within a web browser, and is accessible at http://streamstats.usgs.gov. The StreamStats home page provides a general description of the application. A gray box on the left side of the page contains a series of links to pages that document how to use the application, define terminology, and so forth. Selecting the *Access User Interface* link will cause a new browser window to appear in which the StreamStats user interface will display a map of the United States. Selecting *Region* from the *Zoom To* pull-down list above the map will cause a small window to appear. Selecting *Connecticut River SYE* from the pull-down list in that window will cause the map in the StreamStats user interface to display the extent of the CRB and also will cause an introductory page to appear that explains the StreamStats functionality that is available for this area, provides citations to relevant reports, and identifies other organizations that contributed to the application development.

The map navigation tools provided in the StreamStats user interface should be used to locate a point along the stream of interest. With the map zoomed into a scale of at least 1:24,000, pressing on the *Watershed Delineation* button, and then on the map at location of interest will cause the catchment boundary for the selected location to be delineated and displayed on the map (fig. 5A). Once the catchment is delineated, pressing on the *Basin Characteristics* button will result in the appearance of a new browser window that contains a table of the catchment characteristics for the selected location (fig. 5B). StreamStats uses the processes described by ESRI, Inc. (2009) for catchment delineation and computation of catchment characteristics.

StreamStats provides a *Download* tool to export a shapefile of the contributing catchment (fig. 5A) for use in other mapping applications. If catchment characteristics are determined before the shapefile is created, then the catchment characteristics will be saved as attributes with the shapefile.

The CRUISE tool consists of a Microsoft Excel spreadsheet with five worksheets. The features of the CRUISE tool are shown in figure 5. The spreadsheet opens on the *MainMenu* worksheet, which provides additional instruction, a report citation, and support contact information (fig. 5C). The user enters the catchment characteristics summarized by StreamStats into the *BasinCharacteristics* worksheet (fig. 5D) and then presses the command button to compute the unregulated daily streamflows. The program then follows the process outlined in figure 3 by calculating the FDC, selecting the index streamgauge using the map-correlation method and transferring the timing of the streamflows at the index streamgauge to the ungauged location by the QPPQ-transform method. The CRUISE-estimated streamflows are, in part, computed from regional regression equations that were developed using the catchment characteristics discussed in Section 3. Streamflows estimated for ungauged catchments having characteristics outside the range of values used to develop the regression equations are highly uncertain because these values were not used to fit the regression equations. Therefore, the

CRUISE tool includes a message in the *BasinCharacteristics* worksheet next to each characteristic that is outside the respective ranges shown in figure 2 and detailed in Section 2.

The *ReferenceGaugeSelection* worksheet (fig. 5E) displays information about the ungauged catchment and the selected index streamgauge, including the percent difference between catchment characteristics at the ungauged and index streamgauge, the distance between the between catchment characteristics at the ungauged location and index streamgauge, and the estimated cross-correlation resulting from the map-correlation method. Whereas the CRUISE tool automatically selects the index streamgauge estimated to be most correlated with the ungauged location, the CRUISE tool also reports the five index streamgauges estimated to be most correlated with the ungauged location (fig. 5E). The CRUISE tool also allows users to choose from any of the potential index streamgauges in the study area if for some reason they would like to select another index streamgauge, either from the five most-correlated index streamgauges or another index streamgauge (fig. 5E). Users select a new index streamgauge from a pull-down list and choose the update button (fig. 5E). The FDCs at the ungauged location and the index streamgauge as well as the daily streamflow time series at the ungauged location and at the index streamgauge are reported in cubic feet per second and cubic feet per second per mile. The *ContinuousFlowDuration* worksheet (fig. 5F) displays the estimated continuous exceedence probabilities, and the *ContinuousDailyFlow* worksheet (fig. 5G) displays the estimated daily time series for the ungauged site. Both worksheets provide the estimated streamflows in units of cubic feet per second and in cubic feet per second per square mile.

5. Performance of streamflows estimated by the CRUISE model

To evaluate the utility of the CRUISE tool to estimate unregulated, daily streamflow at ungauged locations in the CRB, a leave-one-out cross validation for a subset of 31 study streamgauges (fig. 1) was applied. These study streamgauges were selected as validation streamgauges because they contained a complete period of observed record for the period estimated by the CRUISE tool. Each site was removed completely from the streamflow estimation process – one by one – and the parameters of the regression and variogram models were re-estimated. The re-estimated parameters were then used to generate streamflow at the removed site and estimated streamflows were compared with the observed streamflows.

Goodness of fit between observed and estimated streamflows for the entire simulation period was evaluated using the Nash-Sutcliffe efficiency value [*Nash and Sutcliffe*, 1970], which was computed from both the observed and estimated streamflows as well as the natural logarithms of the observed and estimated streamflows (fig. 6). The natural logarithm of the observed and estimated streamflows was taken to scale the daily streamflow values so that the high and low streamflow values were more equally weighted in the calculation of the efficiency metric. Efficiency values were mapped to determine if there was any spatial bias in the model performance (fig. 7). Selected hydrographs were also plotted to visualize the interpretation of the efficiency values (fig 7).

Percent errors in selected ecologically-relevant high- and low-flow statistics were also compared (fig. 8). These streamflow statistics, as defined by *Hendrickson et al.* [2006], include measures of the magnitude of streamflows (the median monthly streamflows), duration of streamflows (the 30-, 60-, and 90-day minimum and 90-day maximum streamflow), and timing of streamflows (the Julian days of the 1-day minimum and maximum streamflows) (fig. 8).

The efficiency values in figure 6 show that the streamflows estimated by the CRUISE tool generally have good agreement with the observed streamflows at the 31 validation streamgauges. The minimum efficiency computed from the transformed daily streamflows is 0.69 and the maximum value is 0.92 (fig. 6), with an efficiency value equal to 1 indicting perfect agreement between the observed and estimated streamflows. The efficiency values for the untransformed observed and estimated streamflows range from 0.04 to 0.92 (fig. 6). This decrease in efficiency between the transformed and untransformed observed and estimate streamflows suggest that the fit between the observed and estimated streamflows from the CRUISE tool at high streamflow values is more of a challenge than the fit at the other streamflow values. Despite this, the CRUISE model appears to result in high efficiency values across all validation sites (fig. 7). Streamgauges in the northern portion of the CRB have lower efficiency values than streamgauges in the middle and southern portions of the CRB; however, it should be noted from the hydrographs in figure 7 that the CRUISE tool is able to represent the daily features of the hydrographs at the validation streamgauges even though the efficiency values are relatively lower in the northern portion of the study area. The selected streamflow statistics estimated by the CRUISE tool also provide a reasonable match to the observed streamflow statistics at the validation sites (fig. 8). The percent error for the majority of the streamflow statistics is between +/- 10 percent (fig. 8), with the exception of low-flow statistics. Percent error tends to be inflated for these streamflow statistics because the streamflows are already low values and, when divided by the difference between the observed and estimated values, the percentages can appear high even though the absolute differences between the observed and estimated streamflows values are low. The efficiency values, hydrograph comparisons and flow statistics derived from the observed and estimated streamflows demonstrate that the CRUISE tool can provide a reasonable representation of natural streamflow time series at ungauged catchments in the CRB.

6. Summary and conclusions

This paper presents the Connecticut River UnImpacted Streamflow Estimation (CRUISE) tool, which estimates daily, unregulated streamflow at ungauged locations in the Connecticut River Basin (CRB). The CRUISE tool is freely-available and requires only an internet connection and Microsoft Excel version 2003 or higher. The StreamStats web application must be used to select the location of the ungauged site, delineate the catchment boundary, and determine its catchment characteristics before the CRUISE tool can be used. CRUISE estimates daily streamflow time series for a 44-year period of record from October 1, 1960 through September 30, 2004. Daily streamflow is estimated by a three-part process: 1) estimation of the daily, period-of-record flow-duration curve at the ungauged location, 2) selection of an index streamgauge, and 3) use of the index streamgauge to transfer the flow-duration curve to a time series of daily streamflow. The CRUISE tool provided reliable estimates of observed daily streamflows at 31 validation streamgauges across the CRB. The coupling of the StreamStats and CRUISE tools presents a modeling and software framework that can be used to develop point-and-click, GIS-based, daily-streamflow estimates needed for water management decisions at ungauged stream locations for other regions.

Acknowledgements.

The authors would like to acknowledge Kimberly Lutz and Colin Apse of The Nature Conservancy, Richard Palmer, Austin Polebitski and Scott Steinschneider of the University of Massachusetts Amherst and Christopher Hatfield, Woodrow Fields, and John Hickey of the U.S.

Army Corps of Engineers, for their technical input. The authors would like to also acknowledge John Magee of the New Hampshire Fish and Game Department, who provided valuable review comments on the software tool, Tomas Smieszek of the U.S. Geological Survey for his help with the software website, Jamie Kendall and Melissa Weil for their help in assembling and preparing the data for the CRUISE tool, and Scott Olsen of the U.S. Geological Survey, who provided a list of unimpaired streamgauges located in northern portion of the CRB. This work was funded by the New England Association of Fish and Wildlife Agencies Northeast Regional Conservation Needs Grant number 2007-06 with matching funds from The Nature Conservancy.

References

Archfield, S. A., and R. M. Vogel, 2010. Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments, Water Resour. Res., 46, W10513, doi:10.1029/2009WR008481.

Archfield, S., R. Vogel, P. Steeves, S. Brandt, P. Weiskel, and S. Garabedian, 2010. The Massachusetts Sustainable-Yield Estimator: A decision-support tool to assess water availability at ungaged sites in Massachusetts, U.S. Geological Survey Scientific Investigations Report 2009-5227, 41 p. plus CD-ROM.

Armstrong, D. S., G. W. Parker, and T. A. Richards, 2008. Characteristics and classification of least altered streamflows in Massachusetts, U.S. Geological Survey Scientific Investigations Report, 20075291, 113 p. plus CD–ROM.

Duan, N., 1983) Smearing estimate—a nonparametric retransformation method, J. Am. Stat. Assoc., 78, 383, pp. 605-610.

Castellarin, A., G. Galeati, L. Brandimarte, A. Montanari, and A. Brath, 2004. Regional flow-duration curves: reliability for ungauged basins, Adv.Water Resour., 27, 10, 953-965

ESRI, Inc.,, 2009. Arc-Hydro Tools - Tutorial, Version 1.3 - January 2009, ESRI, Inc., Redlands, CA, available at http://andersonruhoff.googlepages.com/ArcHydro_Tutorial.pdf.

Falcone, J. A., D. M. Carlisle, D. M. Wolock, and M. R. Meador, 2010. GAGES: A stream gage database for evaluating natural and altered flow conditions in the coterminous United States, Ecology, 91, 612.

Fennessey, N. M., 1994. A hydro-climatological model of daily streamflow for the northeast United States, Ph.D. dissertation, Tufts University, Department of Civil and Environmental Engineering.

Granato, G.E., 2009. Computer programs for obtaining and analyzing daily mean streamflow data from the U.S. Geological Survey National Water Information System Web Site, U.S. Geological Survey Open-File Report 2008–1362, 123 p.

Helsel, D., and R. Hirsch, 2002. Statistical Methods in Water Resources Techniques of Water Resources Investigations, Book 4, Chapter A3, U.S. Geological Survey.

Henriksen, J.A., J. Heasley, J.G. Kennen, and S. Nieswand, 2006. User's manual for the Hydroecological Integrity Assessment Process software, U.S. Geological Survey Open-File Report 2006–1093, 71 p.

Hirsch, R., 1979. Evaluation of some record reconstruction techniques, Water Resour. Res., 15, 6, 1781-1790, ISSN 0043-1397.

Holtschlag, D.J., 2009. Application guide for AFINCH, analysis of flows in networks of channels) described by NHDPlus, U.S. Geological Survey Scientific Investigations Report 2009-5188, 106 p.

Hughes, D.A., and V.U. Smakhtin, 1996. Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves: Hydrolog. Sci. J., 41, 6, 851–871.

Isaaks, E. H., and R. M. Srivastava, 1989. An Introduction to Applied Geostatistics, first ed., Oxford University Press, New York.

Mahoamoud, Y. M., 2008. Prediction of daily flow duration curves and streamflow for ungauged cacthments using regional flow duration curves, Hydrolog. Sci. J., 53, 4, 706-724.

Nash, J. E., and J. V. Sutcliffe, 1970. River flow forecasting through conceptual models part I - a discussion of principles, J. of Hydrol., 10, 3, 282–290.

Poff, N.L., J.D. Allen, M.B. Bain, J.R. Karr, K.L. Prestagaard, B.D. Richter, R.E. Sparks, and J.C. Stromberg, 1997. The natural-flow regime—A paradigm for river conservation and restoration: Bioscience, 47, 769–784.

Ribeiro Jr., P., and P. Diggle, 2001. `geoR`: A package for geostatistical analysis, R-News, 1, 2.

Ries, K. G., III; Guthrie, J. G.; Rea, A. H.; Steeves, P. A.; Stewart, D. W., 2008. StreamStats: A Water Resources Web Application, U.S. Geological Survey Fact Sheet 2008-3067, 6 p., available on line at http://pubs.usgs.gov/fs/2008/3067/.

Ries, K.G., III, and P.J. Friesz, 2000. Methods for estimating low-flow statistics for Massachusetts Streams, U.S. Geological Survey Water-Resources Investigations Report 2000-4135, 81 p., available on line at http://pubs.usgs.gov/wri/wri004135/.

Skøien, J. O., and G. Blöschl, 2007. Spatiotemporal topological kriging of runoff time series, Water Resour. Res., 43, 9, doi:10.1029/2006WR005760.

Smakhtin, V. U., 1999. Generation of natural daily flow time-series in regulated rivers using a non-linear spatial interpolation technique, Regul. Rivers: Res. Mgmt, 15, 311-323.

Smakhtin V.U. and N. Eriyagama, 2008. Developing a software package for global desktop assessment of environmental flows. Environ. Model. Softw. 23, 12, December 2008. 1396-1406. doi:10.1016/j.envsoft.2008.04.002.

Stedinger, J.R., R.M. Vogel, and E. Foufoula-Georgiou, 1993. Frequency analysis of extreme events, Chapter 18, in Maidment, D.R., ed., Handbook of Hydrology: McGraw-Hill Book Company, New York.

Vogel, R.M., and N.M. Fennessey, 1994. Flow duration curves I: A new interpretation and confidence intervals, J. Water Res. Pl.-ASCE, 120,4, p. 485-504.

Vogel, R.M., and J.R. Stedinger, 1985. Minimum variance streamflow record augmentation procedures, Water Resour. Res., 21,5, p. 715–723.

Williamson, T.N., K.R. Odom, J.K. Newson, A.C. Downs, H.L. Nelson Jr., P.J. Cinotto and M.A. Ayers, 2009. The Water Availability Tool for Environmental Resources, WATER)—A water-budget modeling approach for managing water-supply resources in Kentucky—Phase I—Data processing, model development, and application to non-karst areas: U.S. Geological Survey Scientific Investigations Report 2009–5248, 34 p.

Zhang, Y., and F. H. S. Chiew, 2009. Relative merits of different methods for runoff predictions in ungauged catchments, Water Resour. Res., 45, W07412, doi:10.1029/2008WR007504.

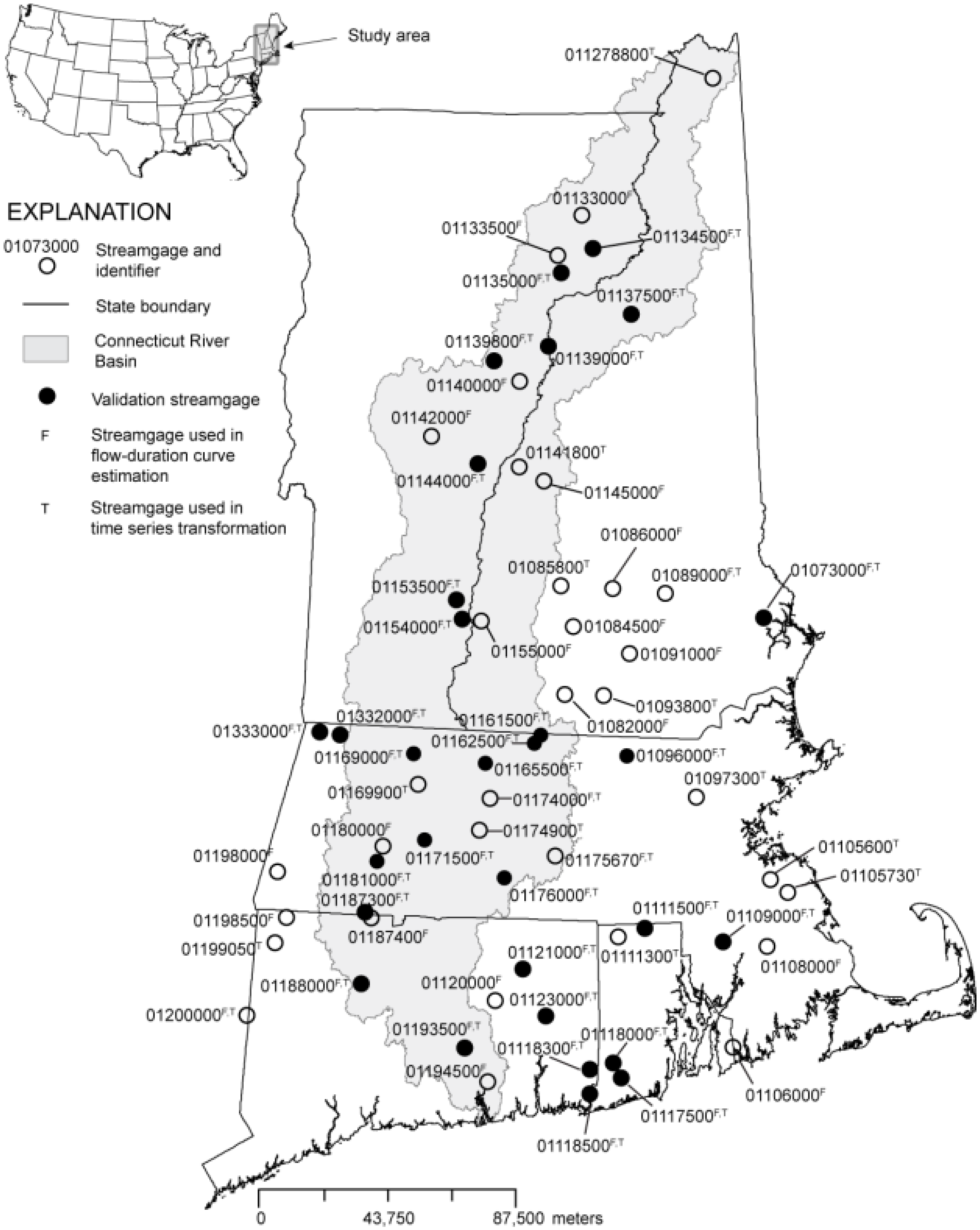


Figure 1. Map showing the locations of the streamgauges used to estimate unregulated, daily streamflow at ungauged locations in the Connecticut River Basin, northeast United States.

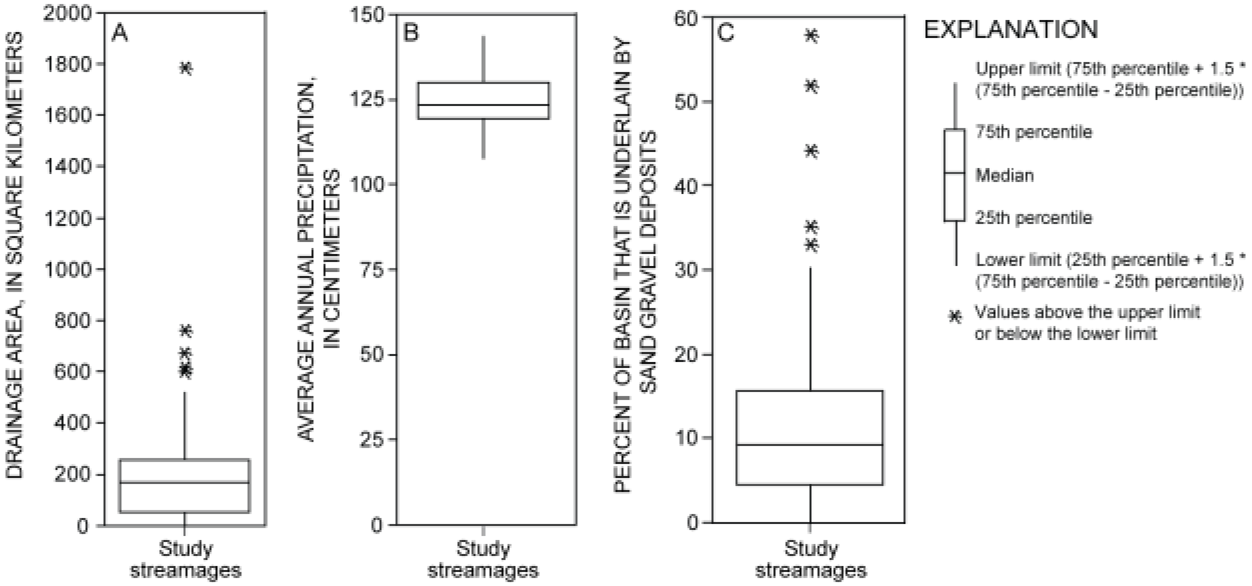


Figure 2. Boxplots of the catchment characteristics used to estimate unregulated, daily streamflow at ungauged locations in the Connecticut River Basin, northeast United States and the subset of catchment characteristics used to validate the streamflow estimates.

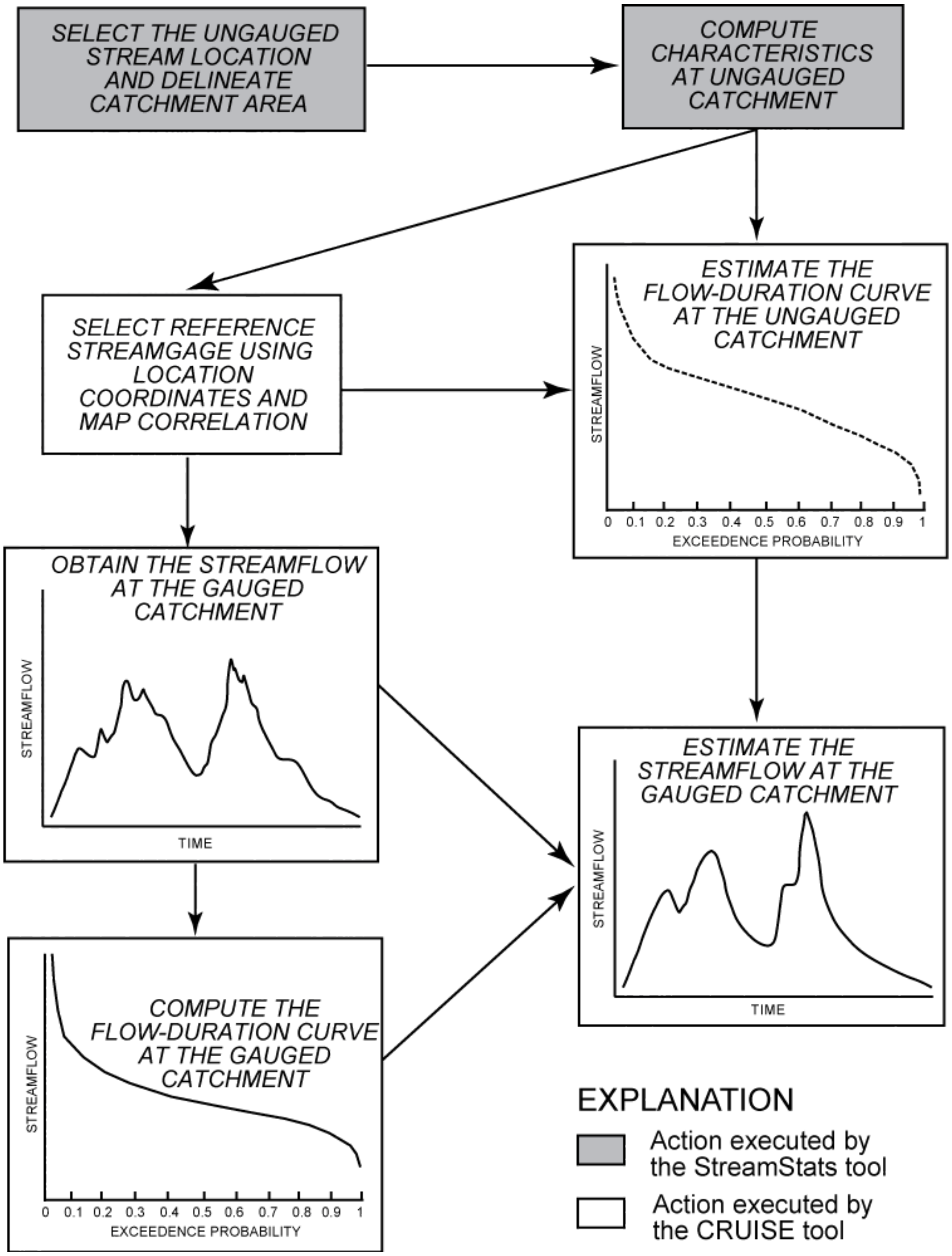


Figure 3. Diagram of the process to estimate unregulated, daily streamflow at ungauged locations.

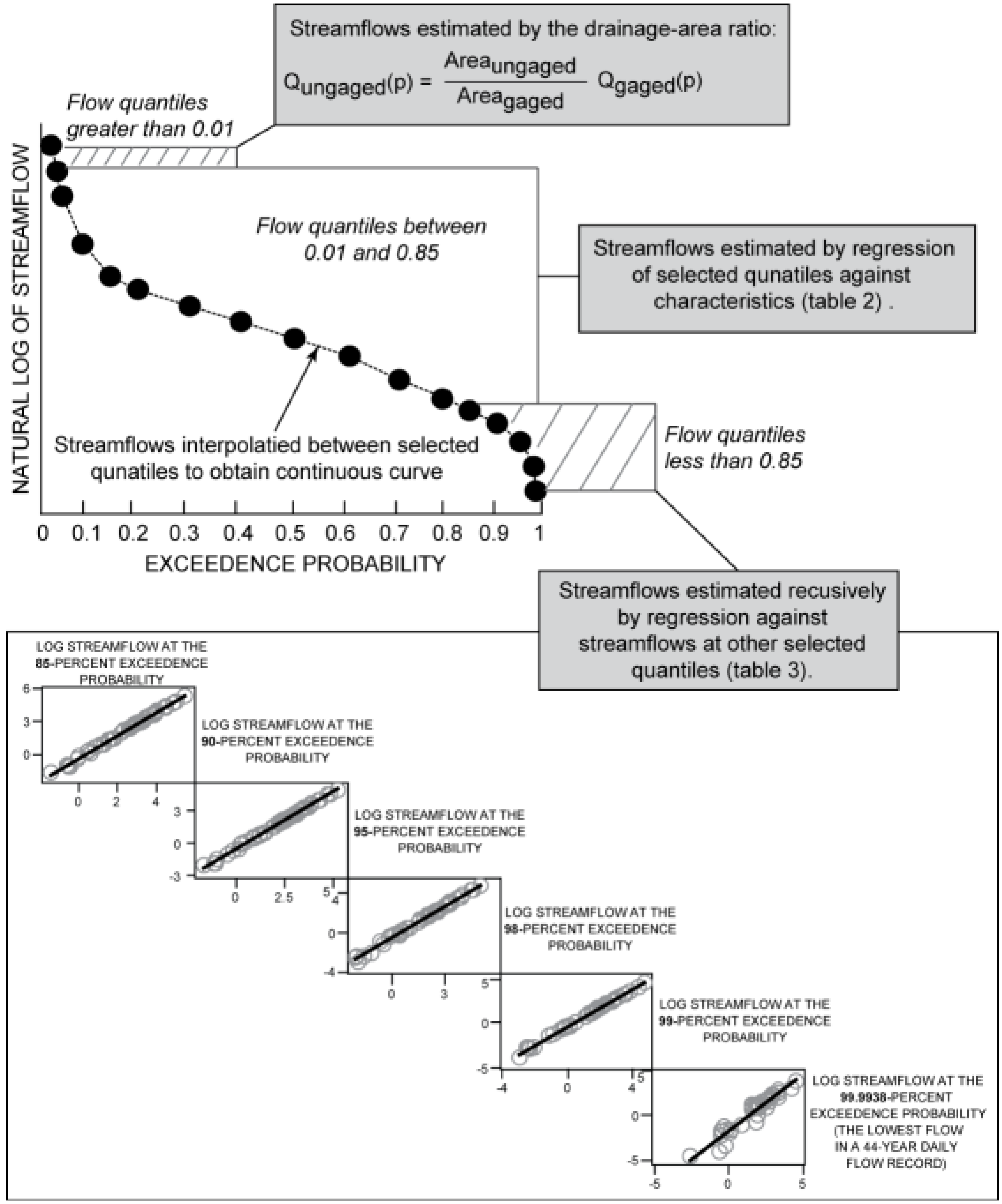
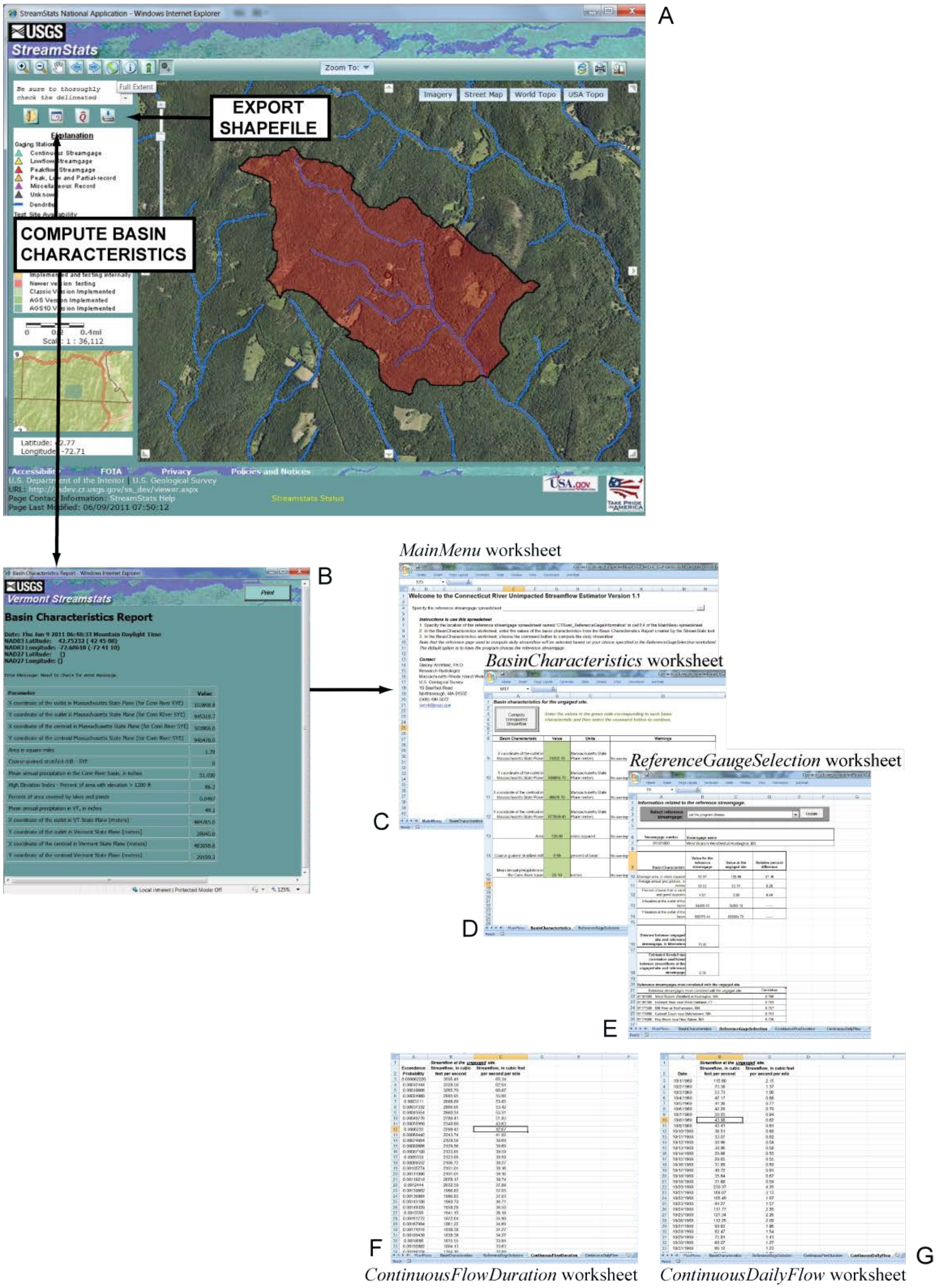


Figure 4. Diagram showing the methods used to estimate a continuous, daily flow duration at an ungaued location.



Basin Characteristics Report - Windows Internet Explorer

Vermont Streamstats

Basin Characteristics Report

Date: Thu Jun 9 2011 06:48:33 Mountain Daylight Time
NAD83 Latitude: 42.75232 (42 45 08)
NAD83 Longitude: -72.68610 (-72 41 10)
NAD27 Latitude: ()
NAD27 Longitude: ()

Error Message: Need to check for error message.

Parameter	Value
X coordinate of the outlet in Massachusetts State Plane (for Conn River SYE)	102898.8
Y coordinate of the outlet in Massachusetts State Plane (for Conn River SYE)	945310.7
X coordinate of the centroid in Massachusetts State Plane (for Conn River SYE)	103806.0
Y coordinate of the centroid Massachusetts State Plane (for Conn River SYE)	940478.0
Area in square miles	1.79
Coarse-grained stratified drift -- SYE	0
Mean annual precipitation in the Conn River basin, in inches	51.030
High Elevation Index - Percent of area with elevation > 1200 ft	86.2
Percent of area covered by lakes and ponds	0.0497
Mean annual precipitation in VT, in inches	49.1
X coordinate of the outlet in VT State Plane (meters)	484765.0
Y coordinate of the outlet in Vermont State Plane (meters)	28045.0
X coordinate of the centroid in Vermont State Plane (meters)	483658.8
Y coordinate of the centroid Vermont State Plane (meters)	29199.3

Local intranet | Protected Mode: Off

Figure 5. Screen captures showing the decision-support tool used to estimate daily, unregulated time series in the Connecticut River Basin. The program delineates a catchment for the ungauged location selected by the user (A) and summarizes the catchment characteristics (B). The user then inputs these characteristics into a spreadsheet program (C-E) that generates the daily, period of record flow-duration curve (F) and the daily streamflow time series (G).

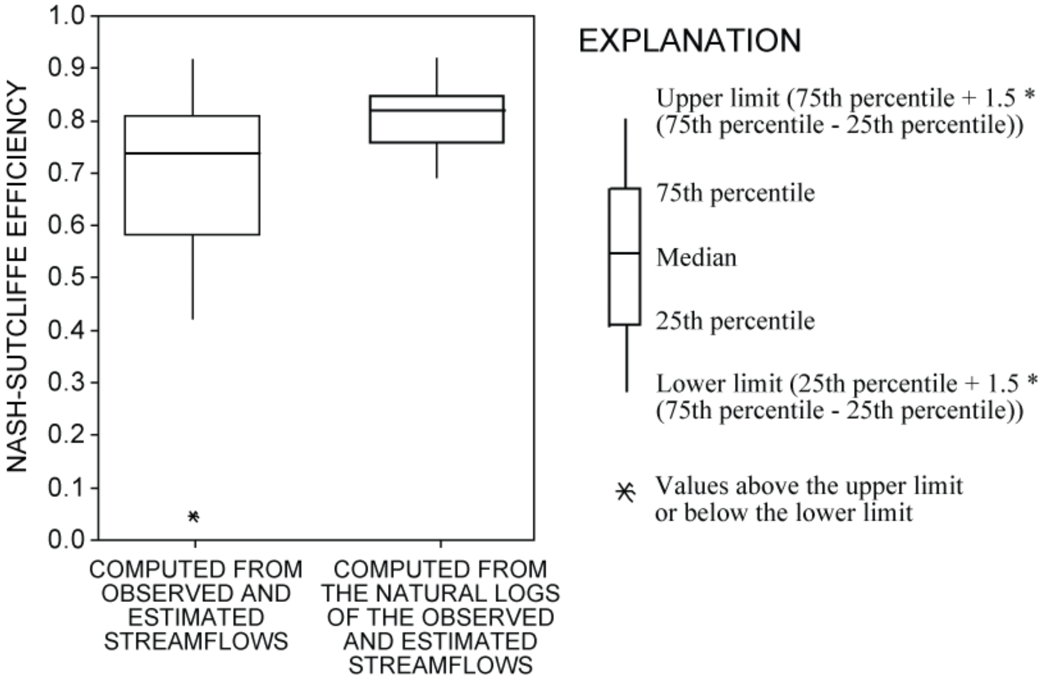


Figure 6. Range of efficiency values computed between the observed and estimated streamflows at the 31 validation streamgauges.

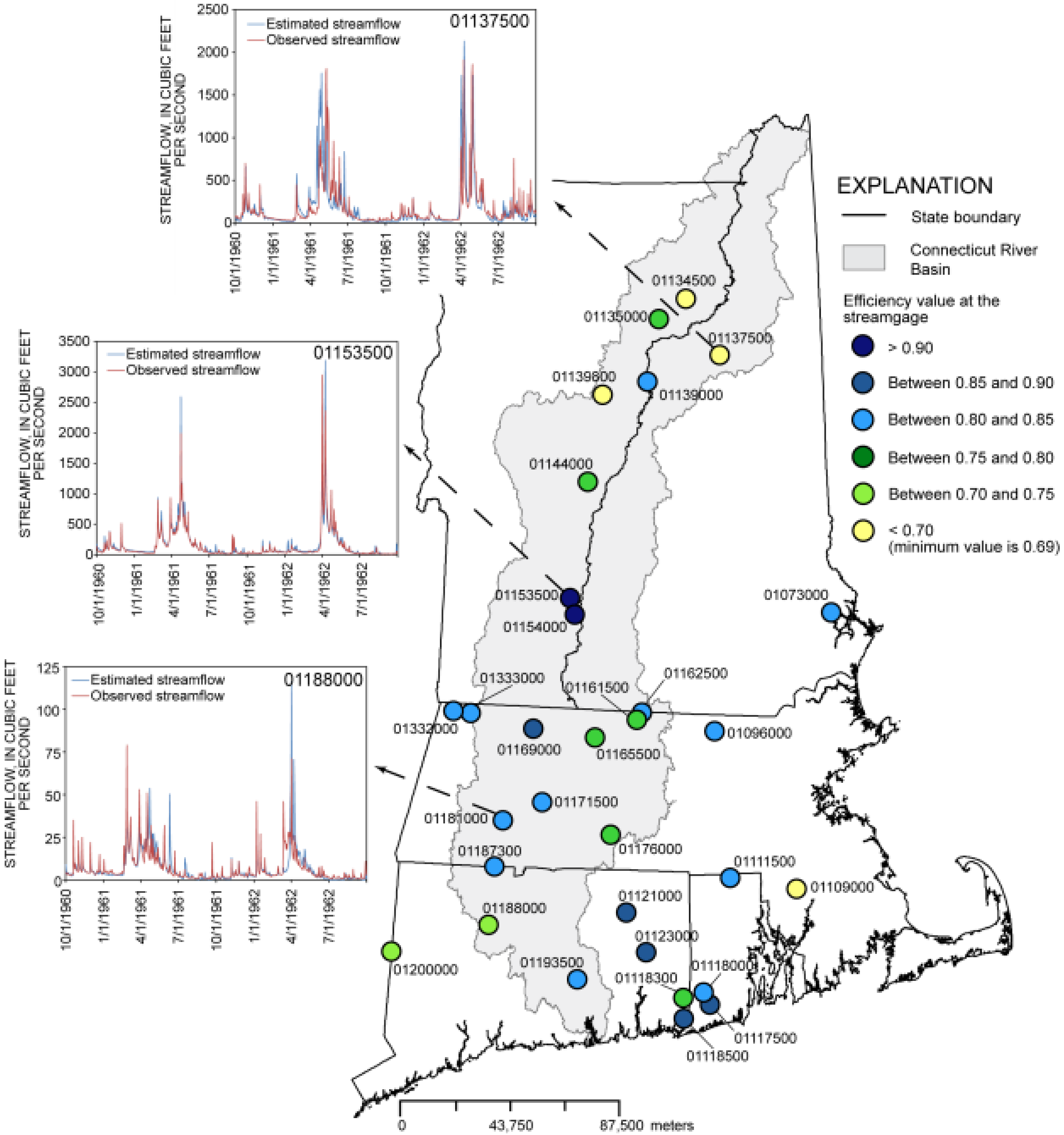


Figure 7. Spatial distribution of efficiency values resulting from log-transformed observed and estimated daily streamflow at 31 validation streamgages and selected hydrographs of observed and estimated streamflow for the period from October 1, 1960 through September 30, 1962.

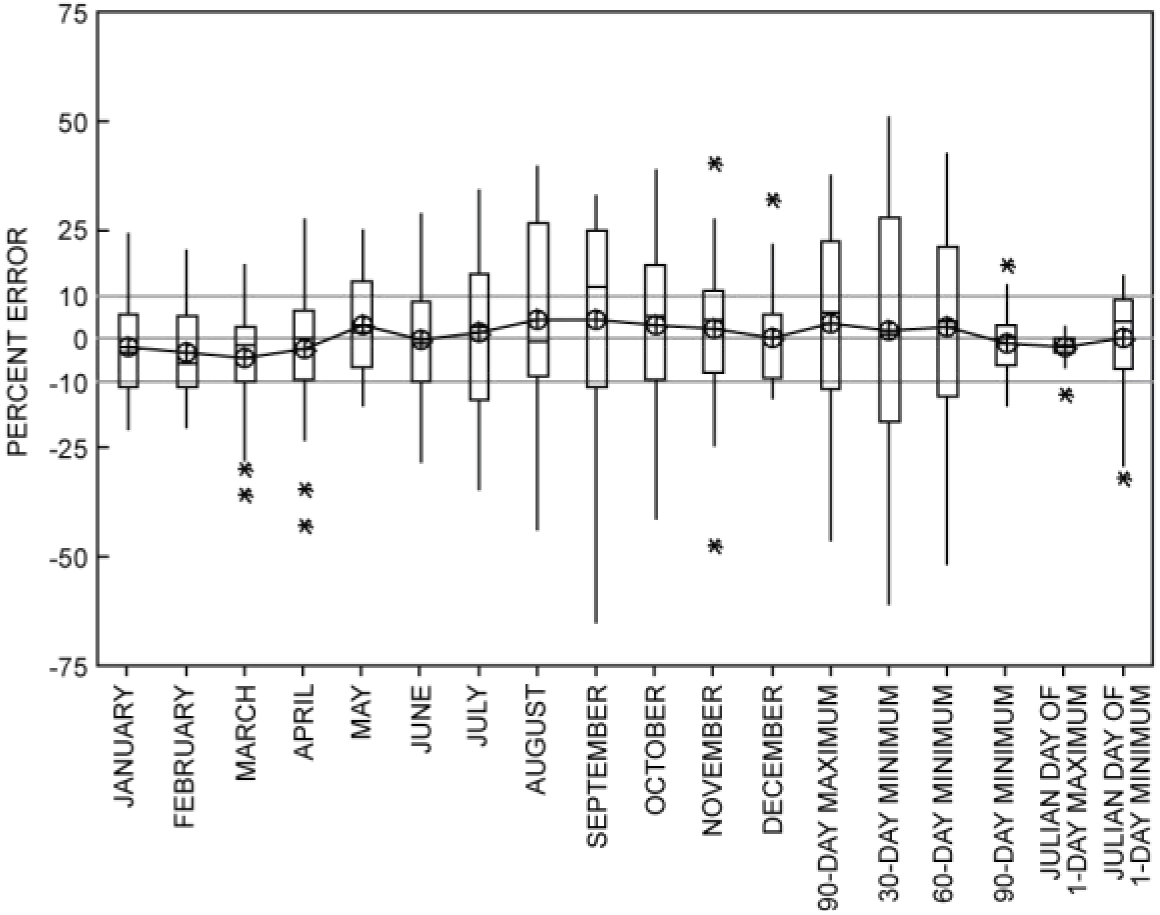


Figure 8. Percent error between observed and estimated monthly mean flows and other ecologically-relevant streamflow statistics at the 31 validation streamgauges.

Table 1. List of streamgauges used to estimate unregulated, daily streamflow at ungauged locations in the Connecticut River Basin.

Station Number	Station name	Period of record
01073000	Oyster River near Durham, NH	December 15, 1934 - December 31, 2004
01082000	Contocook River at Peterborough, NH	July 7, 1945 - September 30, 1977
01084500	Beard Brook near Hillsboro, NH	October 1, 1945 - September 30, 1970
01085800	West Branch Warner River near Bradford, NH	May 22, 1962 - September 30, 2004
01086000	Warner River at Davisville, NH	October 1, 1939 - September 30, 1978
01089000	Soucook River near Concord, NH	October 1, 1951 - September 30, 1987
01091000	South Branch Piscataquog River near Goffstown, NH	July 27, 1940 - September 30, 1978
01093800	Stony Brook tributary near Temple, NH	May 1, 1963 - September 30, 2004
01096000	Squannacook River near West Groton, MA	October 1, 1949 - December 31, 2004
01097300	Nashoba Brook near Acton, MA	July 26, 1963 - December 31, 2004
01105600	Old Swamp River near South Weymouth, MA	May 20, 1966 - July 24, 2006
01105730	Indian Head River at Hanover, MA	July 8, 1966 - July 24, 2006
01106000	Adamsville Brook at Adamsville, RI	October 1, 1940 - September 30, 1978
01108000	Taunton River near Bridgewater, MA	October 1, 1929 - April 23, 1976
01109000	Wading River near Norton, MA	June 1, 1925 - December 31, 2004
01111300	Nipmuc River near Harrisville, RI	March 1, 1964 - September 30, 1991
01111500	Branch Riverb at Forestdale, RI	January 24, 1940 - December 31, 2004
01117500	Pawcatuck River at Wood River Junction, RI	December 7, 1940 - December 31, 2004
01118000	Wood River Hope Valley, RI	March 12, 1941 - December 31, 2004
01118300	Pendleton Hill Brook near Clarks Falls, CT	October 1, 1958 - December 31, 2004
01118500	Pawtucket River at Westerly, RI	November 27, 1940 - December 31, 2004
01120000	Hop Brook near Columbia, CT	October 1, 1932 - October 6, 1971
01121000	Mount Hope River near Warrenville, CT	October 1, 1940 - December 31, 2004
01123000	Little River near Hanover, CT	October 1, 1951 - December 31, 2004
01127880	Big Brook Near Pittsburg Nh	December 1, 1963 - January 1, 1984
01133000	East Branch Passumpsic River near East Haven, VT	October 1, 1948 - September 1, 1979
01133500	Passumpsic River near St. Johnsbury, VT	May 1, 1909 - July 1, 1919
01134500	Moose River at Victory, VT	January 1, 1947 - May 12, 2010
01135000	Moose River at St. Johnsbury, VT	August 1, 1928 - September 1, 1983
01137500	Ammonoosuc River at Bethlehem Junction, NH	August 1, 1939 - May 12, 2010
01139000	Wells River at Wells River, VT	August 1, 1940 - May 12, 2010
01139800	East Orange Branch at East Orange, VT	June 1, 1958 - May 12, 2010
01140000	South Branch Waits River near Bradford, VT	April 1, 1940 - September 1, 1951
01141800	Mink Brook near Etna, NH	August 1, 1962 - September 1, 1998
01142000	White River near Bethel, VT	June 1, 1931 - September 1, 1955
01144000	White River at West Hartford, VT	October 1, 1951 - May 12, 2010
01145000	Mascoma River at West Canaan, NH	July 1, 1939 - September 1, 1978
01153500	Williams River near Rockingham, VT	June 1, 1940 - September 1, 1984
01154000	Saxtons River at Saxtons River, VT	June 20, 1940 - September 30, 1982
01155000	Cold River at Drewsville, NH	June 23, 1940 - September 30, 1978
01161500	Tarbell Brook near Winchendon, MA	May 29, 1916 - September 6, 1983
01162500	Priest Brook near Winchendeon, MA	October 1, 1936 - December 31, 2004
01165500	Moss Brook at Wendell Depot, MA	June 1, 1916 - September 30, 1982
01169000	North River at Shattuckville, MA	December 13, 1939 - December 31, 2004
01169900	South River near Conway, MA	January 1, 1967 - December 31, 2004
01171500	Mill River at Northampton, MA	November 18, 1938 - December 31, 2004
01174000	Hop Brook near New Salem, MA	November 19, 1947 - September 30, 1982
01174900	Cadwell Creek near Belchertown, MA	July 13, 1961 - September 30, 1997
01175670	Sevenmile River near Spencer, MA	December 1, 1960 - December 31, 2004
01176000	Quaboag River at West Brimfield, MA	August 19, 1912 - December 31, 2004
01180000	Sykes Brook at Knightville, MA	June 20, 1945 - July 18, 1974
01181000	West Branch Westfield at Huntington, MA	September 1, 1935 - December 31, 2004

01187300	Hubbard River near West Hartland, CT	August 4, 1959 - December 31, 2004
01187400	Valley Brook near West Hartland, CT	October 1, 1940 - September 30, 1972
01188000	Burlington Brook near Burlington, CT	October 1, 1931 - December 31, 2004
01193500	Salmon River near East Hampton, CT	October 1, 1928 - December 31, 2004
01194500	East Branch Eightmile River near North Lyme, CT	October 1, 1937 - October 6, 1981
01198000	Green River near Great Barrington, MA	October 1, 1951 - September 30, 1971
01198500	Blackberry River at Canaan, CT	October 1, 1949 - October 20, 1971
01199050	Salmon Creek at Lime Rock, CT	October 1, 1961 - December 31, 2004
01200000	Ten Mile River, CT	October 1, 1930 - April 4, 1988
01332000	North Branch Hoosic River at North Adams, MA	June 22, 1931 - September 30, 1990
01333000	Green River at Williamstown, MA	September 20, 1949 - December 31, 2004

Table 2. Number of streamgauges, goodness of fit values, explanatory variables, and estimated regression parameters for streamflows estimated from catchment characteristics.

[%RMSE, Percent root-mean square error; **, parameters not included in regression equation; †, Bias correction factor computed from Duan (1983)]

	General regression information			**Characteristics in the regression equation and coefficient value**						
Exceedence probability	Number of streamgages used to develop regression equation	%RMSE	Efficiency value	Constant term	Drainage area	Average annual precipitation	Percent of basin that is underlain by sand and gravel deposits	Y-location of the basin centroid	X-location of the basin centroid	Bias correlation factor†
0.02	51	1.49	0.99	-26.57576	0.95898668	2.32615049	**	1.44624683	**	1.01034136
0.05	51	0.62	1.00	-19.31477	0.97753613	1.75206249	**	1.04571137	**	1.00231622
0.1	51	0.73	0.99	-2.122368	0.99821922	0.91063827	**	**	**	1.00149376
0.15	51	0.60	1.00	-2.977718	1.0050255	1.05886106	**	**	**	0.99718242
0.2	51	0.86	0.99	-3.693535	1.00370636	1.1919636	**	**	**	0.99567544
0.25	51	1.32	0.98	-4.668431	1.01096501	1.38901861	**	**	**	0.99499049
0.3	51	1.86	0.98	-5.539372	1.01366908	1.56884624	**	**	**	0.99497319
0.4	51	3.00	0.96	-6.759127	1.0206135	1.79999061	**	**	**	0.99601846
0.5	51	3.86	0.95	-7.680269	1.02689366	1.95768173	**	**	**	0.99821139
0.6	50	4.40	0.96	-8.346613	1.0184328	2.01229612	0.080378539	**	**	1.01839085
0.7	50	6.61	0.94	-8.449954	1.04799744	1.90718911	0.094903615	**	**	1.02784995
0.75	50	9.24	0.93	-8.745009	1.06545026	1.90731506	0.10398442	**	**	1.02430492
0.8	50	13.58	0.92	-9.108501	1.09514097	1.9007584	0.125122097	**	**	1.0379095
0.85	50	21.20	0.90	-9.315441	1.12388596	1.8479745	0.151546518	**	**	1.05647159

Table 3. Number of streamgauges, goodness of fit values, explanatory variables, and estimated regression parameters for streamflows estimated from other streamflow quantiles.

[%RMSE, Percent root-mean square error; †, Bias correction factor computed from Duan (1983)]

	General regression information			**Characteristics in the regression equation and coefficient value**			
Exceedence probability	Number of streamgages used to develop regression equation	%RMSE	Efficiency value	Constant term	Coefficient on explanatory variable	Explanatory variable	Bias correlation factor†
0.9	50	32.36	0.89	-0.4112	1.0511	Streamflow at the 0.85 exceedence probaility	1.0004
0.95	50	57.15	0.85	-0.4991	1.0607	Streamflow at the 0.9 exceedence probaility	0.9986
0.98	50	67.36	0.79	-0.4695	1.0567	Streamflow at the 0.95 exceedence probaility	1.0103
0.99	50	102.33	0.71	-0.3011	1.0467	Streamflow at the 0.98 exceedence probaility	1.0000
0.999938	34	825.08	-1.30	-1.6658	1.2826	Streamflow at the 0.99 exceedence probaility	1.2011

Table 4. Variogram model parameters and root-mean-square error value resulting from a leave-one-out cross validation of the variogram models.

Station Number	Variance parameter	Range parameter	Root-mean-square error
01073000	0.0411	697945.4362	0.0399
01085800	0.0115	267272.8077	0.0388
01089000	0.0112	269793.6063	0.0462
01093800	0.0147	267272.7273	0.0416
01096000	0.0389	607472.9297	0.0469
01097300	0.0261	374218.0554	0.0488
01105600	0.0621	557922.7912	0.0488
01105730	0.0677	547625.3299	0.0447
01109000	0.0588	489036.3840	0.0487
01111300	0.0444	435141.4397	0.0470
01111500	0.0649	664951.4696	0.0452
01117500	0.0964	846131.5260	0.0548
01118000	0.0680	547336.8809	0.0456
01118300	0.0541	478962.6030	0.0421
01118500	0.1548	1255724.6703	0.0469
01121000	0.0440	467562.3777	0.0442
01123000	0.0487	476803.1943	0.0457
01127880	0.0475	451474.0307	0.0241
01134500	0.0585	593052.1148	0.0491
01135000	0.0828	885228.5293	0.0574
01137500	0.0421	469510.7730	0.0194
01139000	0.0354	483627.8140	0.0309
01139800	0.0224	369057.2000	0.0255
01141800	0.0116	267272.7273	0.0264
01144000	0.0155	302281.0433	0.0328
01153500	0.0135	267272.7081	0.0409
01154000	0.0129	213818.1818	0.0470
01161500	0.0187	337256.6753	0.0447
01162500	0.0176	291135.1932	0.0436
01165500	0.0291	445510.0450	0.0417
01169000	0.0190	317944.4643	0.0402
01169900	0.0245	398758.9250	0.0442
01171500	0.0310	393869.0688	0.0454
01174000	0.0249	330495.4703	0.0443
01174900	0.0321	412573.1453	0.0430
01175670	0.0366	486730.2368	0.0463
01176000	0.0357	526274.7021	0.0498
01181000	0.0333	502453.4839	0.0426
01187300	0.0566	846080.6046	0.0422
01188000	0.0313	454196.0564	0.0427
01193500	0.0412	435477.5668	0.0445
01199050	0.0212	368184.1116	0.0414
01200000	0.0401	538909.4325	0.0444
01332000	0.0114	175180.2029	0.0370
01333000	0.0148	267272.7273	0.0341