



Towards a publicly available, map-based regional software tool to estimate unregulated daily streamflow at ungauged rivers

S. A. Archfield¹, P. A. Steeves¹, J. D. Guthrie², and K. G. Ries III³

¹New England Water Science Center, US Geological Survey, 10 Bearfoot Road, Northborough, MA 01532, USA

²Rocky Mountain Geographic Science Center, US Geological Survey, P.O. Box 25046 MS 516, Denver Federal Center, Denver, CO 80225, USA

³Office of Surface Water, US Geological Survey, 5522 Research Park Drive, Baltimore, MD 21228, USA

Correspondence to: S. A. Archfield (sarch@usgs.gov)

Received: 24 July 2012 – Published in Geosci. Model Dev. Discuss.: 31 August 2012

Revised: 17 December 2012 – Accepted: 18 December 2012 – Published: 28 January 2013

Abstract. Streamflow information is critical for addressing any number of hydrologic problems. Often, streamflow information is needed at locations that are ungauged and, therefore, have no observations on which to base water management decisions. Furthermore, there has been increasing need for daily streamflow time series to manage rivers for both human and ecological functions. To facilitate negotiation between human and ecological demands for water, this paper presents the first publicly available, map-based, regional software tool to estimate historical, unregulated, daily streamflow time series (streamflow not affected by human alteration such as dams or water withdrawals) at any user-selected ungauged river location. The map interface allows users to locate and click on a river location, which then links to a spreadsheet-based program that computes estimates of daily streamflow for the river location selected. For a demonstration region in the northeast United States, daily streamflow was, in general, shown to be reliably estimated by the software tool. Estimating the highest and lowest streamflows that occurred in the demonstration region over the period from 1960 through 2004 also was accomplished but with more difficulty and limitations. The software tool provides a general framework that can be applied to other regions for which daily streamflow estimates are needed.

1 Introduction

Streamflow information at ungauged rivers is needed for any number of hydrologic applications; this need is of such importance that an international research initiative known as Prediction in Ungauged Basins (PUB) had been underway for the past decade (2003–2012) (Sivapalan et al., 2003). Concurrently, there has been increasing emphasis on the need for daily streamflow time series to understand the complex response of ecology to river regulation and to develop streamflow prescriptions to restore and protect aquatic habitat (Poff et al., 1997, 2010). Basin-wide water allocation decisions that meet both human and ecological demands for water require daily streamflow time series at river locations 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 locations that have both constraints. Often, these locations are ungauged and no information is available to make informed decisions about water allocation.

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 that 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 transferred: (1) rainfall-runoff model parameters (see Zhang and Chiew, 2009 for a review) and

(2) gauged streamflows, or related streamflow properties. The first category assumes that rainfall-runoff models have been developed and calibrated at gauged locations within a region of interest. The rainfall-runoff model parameters are then either used to interpolate parameter values at an ungauged location (as examples see Abdulla and Lettenmaier, 1997; Seibert, 1999; Merz and Blöschl, 2004; Parajka et al., 2005; Oudin et al., 2008) or the calibrated parameter set is directly transferred from a gauged to an ungauged catchment using some measure of similarity between the gauged and ungauged locations (Merz and Blöschl, 2004; McIntyre et al., 2005; Parajka et al., 2005; Oudin et al., 2008, 2010; Zhang and Chiew, 2009; Reichl et al., 2009). Rainfall-runoff models are time and data intensive to develop and calibrate; furthermore, no consistently successful method has been introduced to reliably regionalize model parameters for ungauged locations (Merz and Blöschl, 2004; McIntyre et al., 2005; Parajka et al., 2005; Oudin et al., 2008, Zhang and Chiew, 2009; Oudin et al., 2010). The second category transfers information directly from a stream-gauge or streamgauges to an ungauged location. 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 method (Hirsch, 1979), and a non-linear spatial interpolation method, applied by Fennessey (1994), Hughes and Smakhtin (1996), Smakhtin (1999), Mohamoud (2008), and Archfield et al. (2010), which all transfer a scaled historical streamflow time series from a gauged to an ungauged location. These methods have the advantage of being relatively easy to apply but are limited by the availability of the historical data in the study region.

For the software tool presented in this paper, only the second category of approaches is utilized and a hybrid approach combining the drainage-area ratio and non-linear spatial interpolation methods is introduced to estimate unregulated daily streamflow time series. When streamflow information is presented in a freely available software tool, this information can provide a scientific framework for water-allocation negotiation amongst all stakeholders. Software tools to provide streamflow time series at ungauged locations have been previously published for predefined locations on a river; however few – if any – tools currently exist that provide daily streamflow time series at any stream location for which this information is needed. Smakhtin and Eriyagama (2008) and Holtschlag (2009) introduced software tools to provide monthly streamflows for ecological streamflow assessments at predefined river locations 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, yet, in most cases, at the

monthly – not daily – time step and, in all cases, for only predefined locations on a river that may not be coincident with a river location of interest. The U.S. Geological Survey (USGS) StreamStats tool (Ries et al., 2008) does provide 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 software tool presented here is one of the first such tools to provide unregulated, daily streamflow time series at ungauged locations in a regional framework for any user-desired location on a river. For this study, unregulated streamflow is considered to be streamflow that is not altered – or regulated – by human alteration within the contributing area to the river. This paper first briefly describes the methods used by the software tool. The software tool is then presented and its functionality is described. The software tool can be considered a general framework to provide daily streamflow time series at ungauged locations in other regions of the United States and possibly other areas. Lastly the utility of the software tool to provide reliable estimates of daily streamflow is demonstrated for a large (29 000 km²) basin in the northeast United States. For this region, the software tool utilizes the map-based user interface of the USGS StreamStats tool paired with a macro-based spreadsheet program that allows users to “point-and-click” on a river location of interest and obtain the historical daily streamflow time series.

2 Methods underlying the software tool

Streamflow in the study region is estimated by a multi-step regionalization approach, which starts with the delineation of the contributing area to the ungauged river location of interest and computation of related catchment characteristics (Fig. 1a). For the purposes of this text, catchment and basin are used interchangeably. The flow-duration curve (FDC) for the ungauged location is then obtained using these catchment characteristics (Sect. 2.1; Fig. 1b). The FDC can be considered analogous to the inverse of the empirical cumulative distribution of daily streamflow as it shows the probability of a particular observed streamflow being exceeded. Specific quantiles on the FDC are estimated at the ungauged location by first establishing a regression relation between those flow values observed at the stream-gauges in the study region and measurable catchment characteristics obtained for the contributing areas to those stream-gauges (Sect. 2.1; Fig. 1b). Interpolation is then used to obtain the FDC values for streamflows between the regression-estimated quantiles (Sect. 2.1; Fig. 1b). Lastly, the FDC at the ungauged location is transformed into a time series of streamflow by the selection (Sect. 2.2; Fig. 1c) and use (Sect. 2.3; Fig. 1d) of a donor streamgauge. To ensure that the estimated streamflow represents unregulated conditions, only streamgauges whose catchments have been unaffected

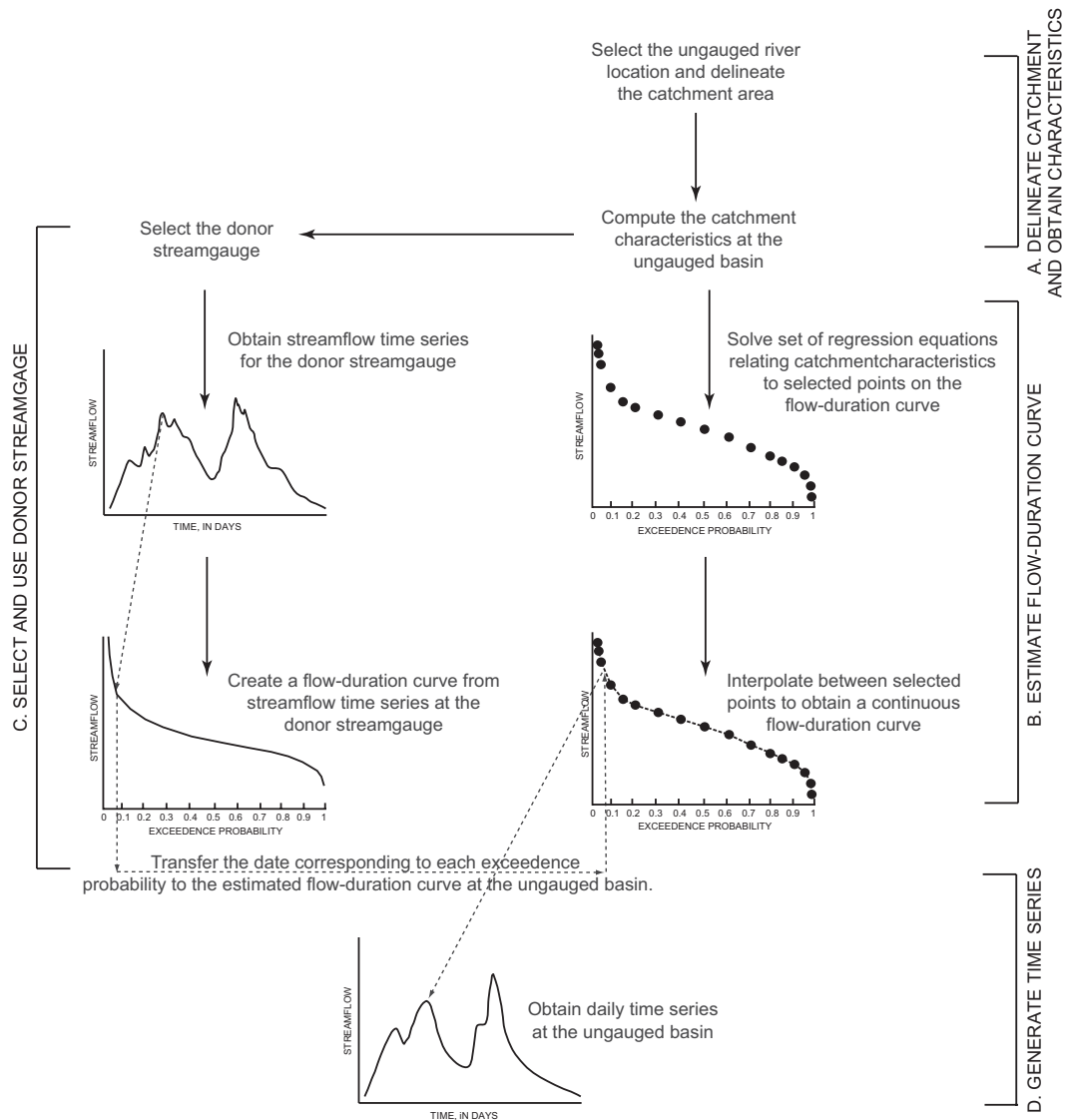


Fig. 1. Diagram of the process to estimate unregulated, daily streamflow at ungauged locations. An ungauged river location is selected, and the catchment characteristics are computed (A). The flow-duration curve is then estimated using regression relations between the catchment characteristics and selected points on the flow-duration curve (B). A donor streamgauge is then selected (C) and used to transfer the estimated flow-duration curve into a time series of daily streamflow at the ungauged location (D).

by anthropogenic influences are utilized to develop the regional regression equations and are considered as a potential donor streamgauge.

2.1 Estimation of the flow-duration curve for the ungauged location

Estimation of the daily FDC at an ungauged location remains an outstanding challenge in hydrology. Castellarin et al. (2004) provide a review of several methods to estimate FDCs at ungauged locations and found that no particular method was consistently better than another. For this study, an empirical, piece-wise approach to estimate the FDC is

used in the software tool (Fig. 2). This overall approach is similar to that used by Mohamoud (2008), Archfield et al. (2010), and Shu and Ourda (2012) 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. The selected quantiles were chosen to be evenly distributed across the FDC with additional quantiles added at the tails of the FDC to provide further resolution to the portions of the FDC that contain the extreme high- and low-streamflow values.

With the exception of streamflows having less than or equal to a 0.01 probability of being exceeded (streamflows

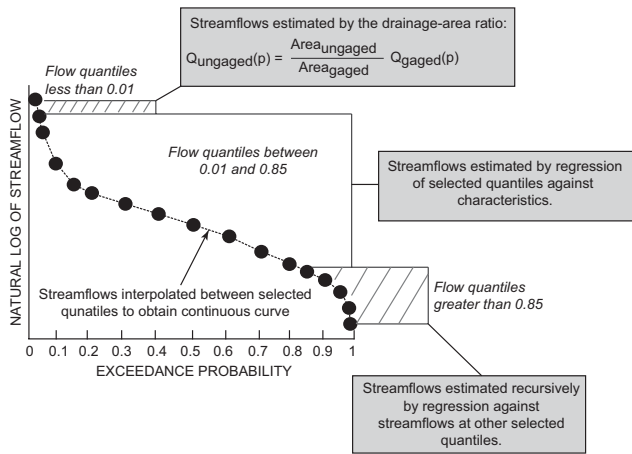


Fig. 2. Diagram showing the methods used to estimate a continuous, daily flow duration at an ungauged location.

with a probability of being exceeded less than 1 percent of the time), selected quantiles on the FDC are estimated from regional regression equations and a continuous FDC is log-linearly interpolated between these quantiles to obtain a continuous FDC (Fig. 2). 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 exceedance probabilities are estimated by independently regressing each streamflow quantile against catchment characteristics (Fig. 2). In this approach, catchment characteristics (the independent variables) are regressed against the streamflow quantiles (the dependent variable) to determine which catchment characteristics have a statistically significant relation with each streamflow quantile. The catchment characteristics tested for inclusion in the regression equations are based on the availability of the spatial data layers in the particular study area of interest and, therefore, vary from region to region. In practice, multiple linear regression is typically applied using the logarithms of the streamflow values and catchment characteristic values, with the form of the regression equation as

$$Y = a_0 + \sum_{i=1}^M a_i X_i + \varepsilon \quad (1)$$

where Y is a vector of the log-transformed values of the streamflow quantile across the study stream gauged, X_i s are the vectors of the log-transformed values of the observed catchment characteristics, a_0 is a constant term estimated by the regression, a_i s are the coefficients estimated by the regression, M is the total number of catchment characteristics and ε is the vector of the model residuals.

Mohamoud (2008) and Archfield et al. (2010) observed that when regressions with catchment characteristics are used across all quantiles on the FDC, there is increased potential for the estimated quantiles to violate the constraint that streamflows must decrease as the exceedance proba-

bility increases, because the uncertainty in the flow estimates is greatest at the lowest portion of the FDC. As confirmed by Archfield et al. (2010), when all streamflow quantiles were regressed against catchment characteristics, there was no constraint to ensure that estimated streamflows decreased with increasing exceedance probability and some estimated streamflow values were larger at higher exceedance probabilities than streamflows estimated at lower exceedance probabilities. Thus, the inherent structure of the data that ensures streamflow quantiles decrease with increasing exceedance probability was not preserved – a physical impossibility. To enforce physical consistency, relations between streamflow quantiles at the 0.9, 0.95, 0.98, 0.99 and 0.999938 exceedance probabilities were estimated by regressing streamflows at these quantiles against one another and using these relations to recursively estimate streamflows (Fig. 2). Regressing quantiles against one another ensures that this constraint is not violated. In this case, the form of the regression equation is equivalent to that of Eq. (1) for the case where i equals 1. This is an alternative approach to that used by Mohamoud (2008), who suggested discarding any estimated quantiles that violate the constraint that streamflows must decrease with increasing exceedance probability.

Using the regression equations to solve for the selected quantiles, the continuous, daily FDC is then determined by log-linear interpolation between the quantiles and ensuring that the interpolation passes through each quantile estimated by regression. Archfield et al. (2010) showed that estimated streamflows determined by log-linear interpolation for exceedance probabilities of 0.01 or less do not match the shape of the FDC 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 curved such that an alternative interpolation scheme such as parabolic or cubic splines is not capable of capturing the shape. Instead of using another interpolation method, streamflows from a donor streamgauge are scaled by catchment area to estimate the highest streamflows at the ungauged location (Fig. 2). This is predicated on the assumption that the shape of the left tail of the FDC is better approximated by the observed streamflow at a donor 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 ratio approach (Eq. 2) in conjunction with the selected donor streamgauge:

$$q_{p_u} = \frac{A_u}{A_g} q_{p_g} \quad (2)$$

where q_{p_u} is the value of the streamflow quantile at the ungauged location for exceedance probability, p , A_u is the contributing drainage area to the ungauged location, A_g is the contributing drainage area to the donor streamgauge, and q_{p_g} is the value of the streamflow quantile at the donor streamgauge for exceedance probability, p . Whereas this piecewise

interpolation of the FDC – particularly at the tails – seems admittedly untidy, it is important to note that previous studies choose to ignore the estimation of the tails of the FDC because of the substantial challenges associated with their estimation (Mohamoud, 2008; Shu and Ourda, 2012).

2.2 Selection of the donor streamgauge

The donor streamgauge is used for two purposes in the streamflow estimation approach: (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. For the direct transfer of streamflow time series from a gauged to an ungauged location, several methods have been used to select the donor catchment. The most common method is the selection of the nearest donor catchment (Mohamoud, 2008; Patil and Stieglitz, 2012; Shu and Ourda, 2012). Also recently, Archfield and Vogel (2010) hypothesized that the cross-correlation between concurrent streamflow time series could be an alternative metric to select the donor streamgauge. For one streamflow transfer method – the drainage area ratio – Archfield and Vogel (2010) showed that the selection of the donor streamgauge with the highest cross-correlation results in a substantial improvement to the estimated streamflows at the ungauged location. Using this result, Archfield and Vogel (2010) introduced a new method – the map correlation method – to estimate the cross-correlation between an ungauged location and a donor streamgauge.

Based on the findings of Archfield and Vogel (2010), the donor streamgauge is selected by the map-correlation method; however, the software tool provides information on the similarity of the selected donor streamgauge to the ungauged location in terms of both distance and similarity in catchment characteristics should the user prefer to use another selection method. Through the use of geostatistics, the map-correlation method selects the donor streamgauge estimated to have the highest cross-correlation between concurrent streamflow time series at the donor streamgauge and the ungauged location. For a given donor streamgauge, the cross-correlations between daily streamflow at the donor streamgauge and the other study streamgauges in the region are computed. Ordinary kriging (Isaaks and Srivastava, 1989) is used to create a relational model – termed the variogram model – for the separation distances between the study streamgauges and the differences in observed cross-correlation. There are several commonly used variogram model forms (Isaaks and Srivastava, 1989); Archfield and Vogel (2010) use a spherical variogram model because of its relatively simple formulation and its visual agreement with the majority of the sample variograms. The spherical variogram, here represented as the covariance function and as presented in Ribeiro Jr. and Diggle (2001), has the form

$$C(h) = \begin{cases} \sigma^2 \left(1 - 1.5 \frac{h}{a} + 0.5 \left(\frac{h}{a} \right)^3 \right), & \text{if } h < 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $C(h)$ is the covariance function variogram model (also referred to as the correlation function), h is the separation distance between streamgauges, σ^2 is the partial sill, and a is the range parameter. Following from traditional geostatistics techniques for ordinary kriging as presented in Isaaks and Srivastava (1989) and as applied by Archfield and Vogel (2010), the variogram model is then used to map the cross-correlation between the donor streamgauge and any location within the study region, including an ungauged location of interest. This mapping is repeated for each possible donor streamgauge in the study region so that estimates of the cross-correlation between the ungauged location and all possible donor streamgauges can be obtained. The software tool then selects the donor streamgauge resulting in the highest estimated cross-correlation with the ungauged location. Additional details on the map correlation method are described in Archfield and Vogel (2010).

2.3 Generation of streamflow time series

With a donor streamgauge selected and estimated daily FDC at the ungauged location, a time series of daily streamflow for the 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; Shu and Ourda, 2012). The term QPPQ-transform method was coined by Fennessey (1994); however, this method has been published by 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 exceedance probability associated with a streamflow value on a given day at the donor streamgauge also occurred on the same day at the ungauged location. For example, if the streamflow on 1 October 1974 was at the 0.9 exceedance probability at the donor streamgauge, then it is assumed that the streamflow on that day at the ungauged location also was at the 0.9 exceedance probability. To implement the QPPQ-transform method, a FDC is assembled from the observed streamflows at the donor streamgauge (Fig. 1c). The exceedance probabilities at the donor and ungauged FDC are then equated (Fig. 1d) and the date that each exceedance probability occurred at the donor streamgauge is transferred to the ungauged catchment (Fig. 1d).

3 Software tool

The software tool can be considered a general framework to provide daily streamflow time series at ungauged locations

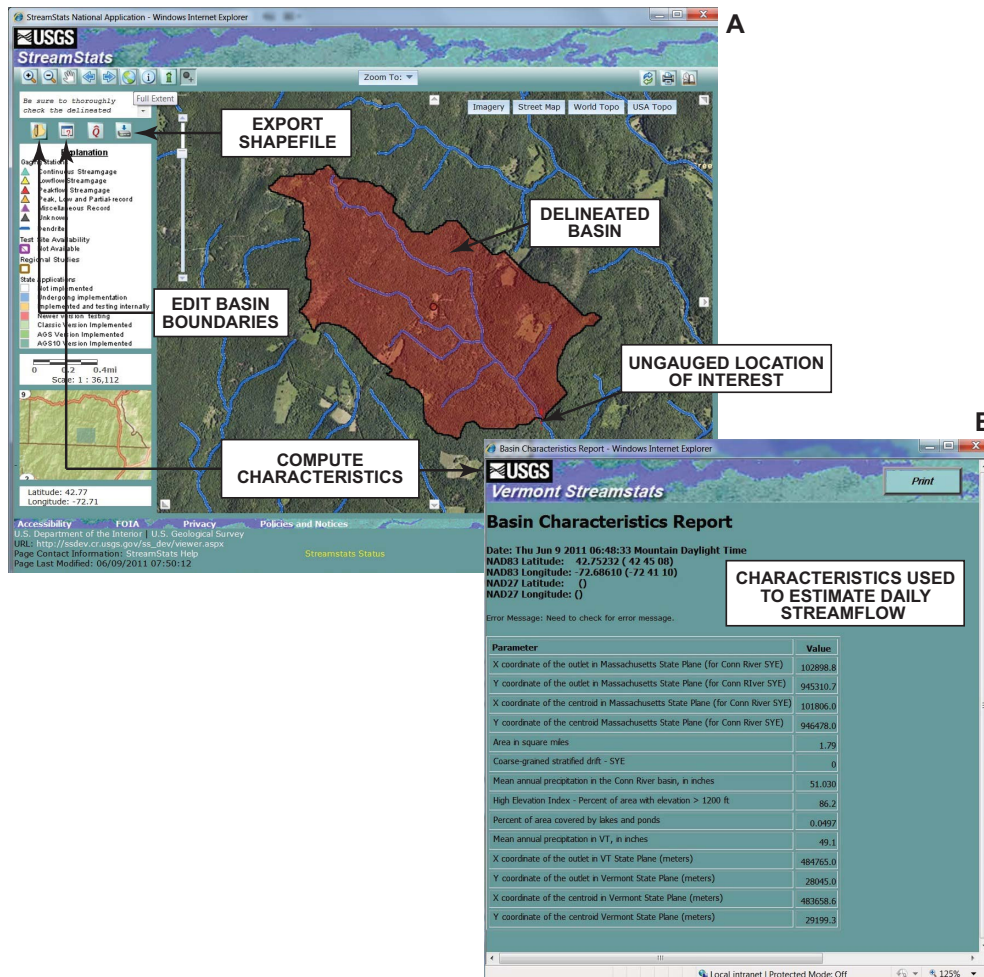


Fig. 3. Screen captures showing the map portion of the software tool used to estimate daily, unregulated time series. The program delineates a catchment (or basin, as named in the tool) for the ungauged location selected by the user (A) and summarizes the catchment characteristics (B). The user also has the option to export the shapefile of the delineated catchment or edit the catchment boundaries (A).

in other regions of the United States and possibly other areas. Furthermore, all data and methods underlying the tool are freely available. Whereas the tool is a general framework for providing a map-based, “point-and-click” approach to estimate daily streamflow at an ungauged river location of interest, the underlying data, including the river network and catchment characteristics, are specific to the region of interest. Much like other modeling frameworks, the software tool must be calibrated based on the data available in the region of interest. Details of the functionality of the regional tool presented in this study follow. Additional details on the customization of the catchment delineation for application to other regions are discussed in Sect. 4.

The software tool initially interfaces with the USGS StreamStats tool (Ries et al., 2008 or <http://streamstats.usgs.gov>) to delineate a catchment area for any user-selected location on a river and to compute the catchment characteristics needed to estimate the FDC at the ungauged location (Fig. 1).

The selection of the donor streamgauge, the computation of the FDC and the estimate of the time series of daily streamflow are then executed by a Microsoft Excel spreadsheet program with Visual Basic for Applications (VBA) coding language. The spreadsheet itself, which contains the VBA source code, can be used independently of the StreamStats interface and is, therefore, able to be customized to interface with other watershed delineation tools or with any study area for which the methods in Sect. 2 have been applied. Additionally, any macro-enabled spreadsheet program could be used in place of the Microsoft Excel spreadsheet program.

The catchment delineation portion of the software tool is handled by the USGS StreamStats tool, which operates within a web browser, and is accessible at <http://streamstats.usgs.gov>. The StreamStats tool implements a watershed delineation process described in Ries et al. (2008) and contains basin-wide spatial data layers of the catchment characteristics needed to solve the regional regression

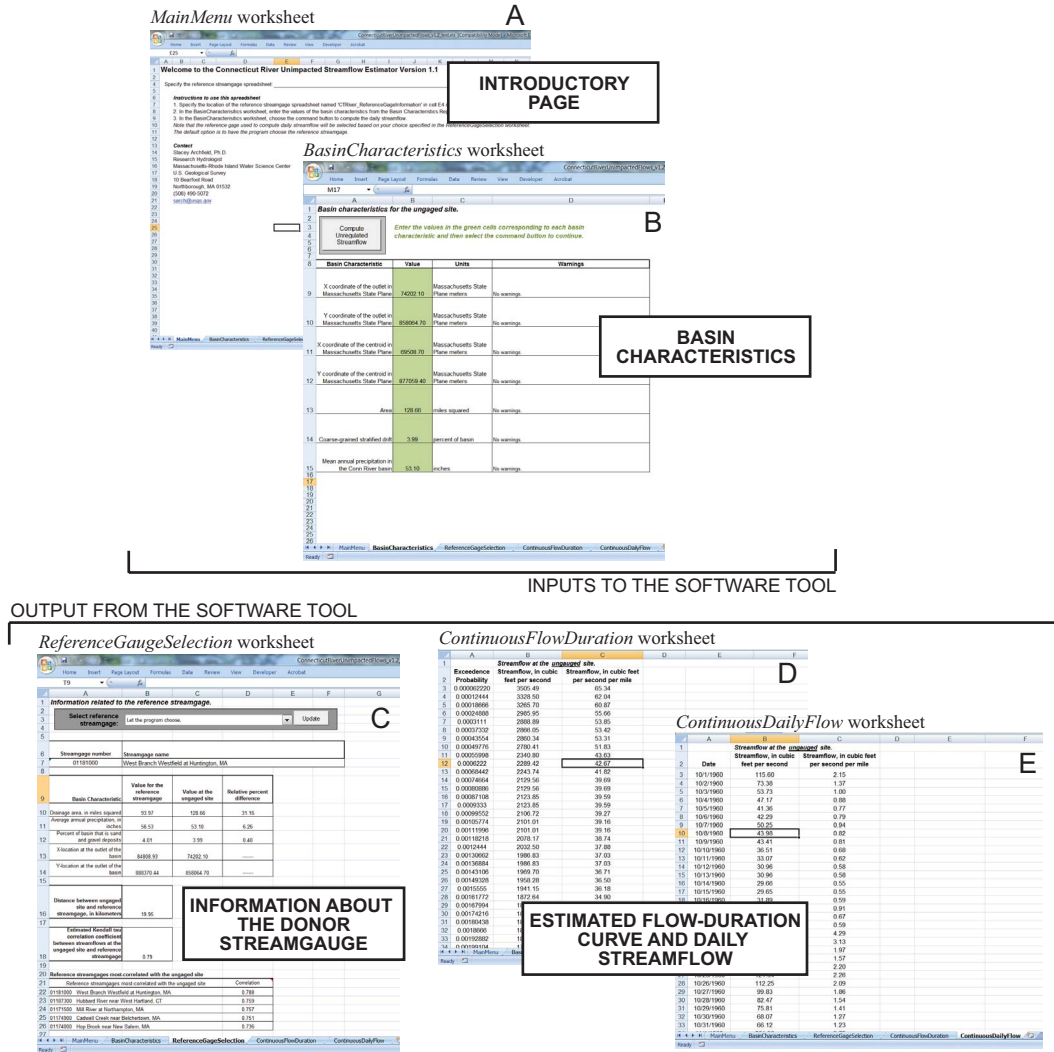


Fig. 4. Screen captures showing the spreadsheet portion of the software tool used to estimate daily, unregulated time series. After reading the introductory page (A), the user inputs the catchment characteristics (or basin characteristics, as named in the tool) into the BasinCharacteristics worksheet (B). The spreadsheet program then selects the donor streamgauge (C) and generates the flow-duration curve (D) and the daily streamflow time series (E).

equations described in Sects. 2.2 and 3.2. The map navigation tools provided in the StreamStats user interface are used to locate a point along the stream of interest. In addition to the stream network, users can view satellite imagery, topographic maps, and street maps to find the river location of interest. This background information can then be used to locate the ungauged river location of interest (Fig. 3a). Users simply click on the river location of interest and the catchment boundary will be delineated and displayed on the map (Fig. 3a). Once the catchment is delineated, pressing a command button will open a new browser window that shows a table of the catchment characteristics for the selected location (Fig. 3b). StreamStats uses the processes described by ESRI, Inc. (2009) for catchment delineation and computation of catchment characteristics. StreamStats also provides

a command button to export a shapefile of the contributing catchment (Fig. 3a) for use in other mapping applications.

Once the catchment characteristics are determined for the ungauged location of interest, the user opens the spreadsheet program and inputs the catchment characteristics into the spreadsheet program to compute the daily streamflow (Fig. 4); the spreadsheet program contains five worksheets (Figs. 4a–e). The spreadsheet opens on the MainMenu worksheet, which provides additional instruction and support contact information (Fig. 4a). The user enters the catchment characteristics summarized by StreamStats (Fig. 4b) into the BasinCharacteristics worksheet (Fig. 4b) and then presses the command button to compute the unregulated daily streamflows. The program then follows the process outlined in Fig. 1b to 1d and Sect. 2. The estimated streamflows

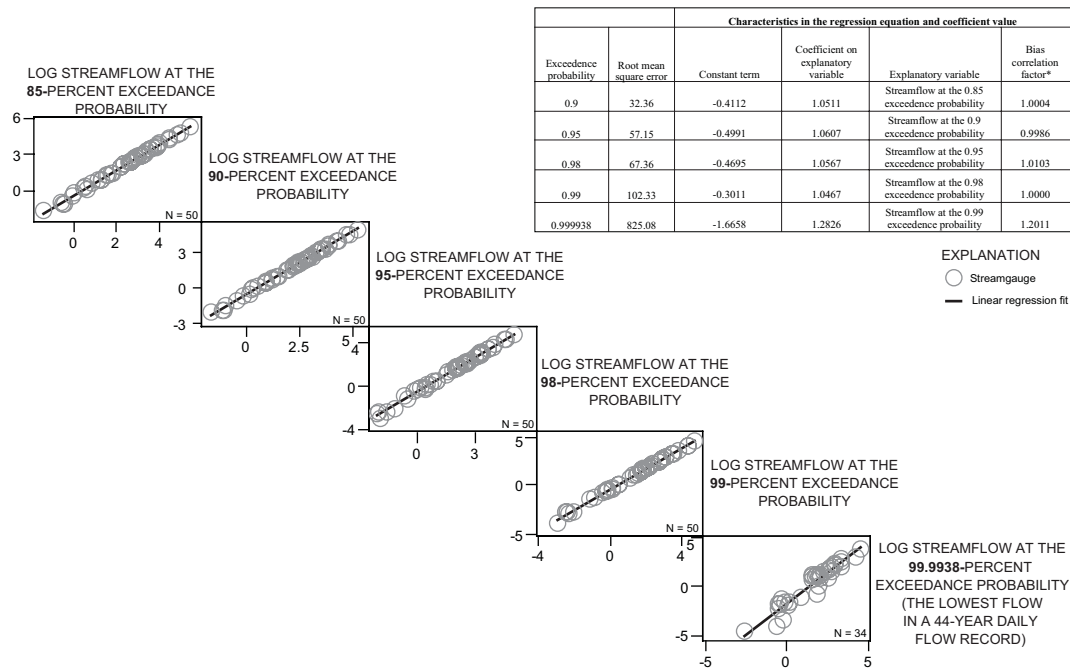


Fig. 5. Relations between streamflows at the 0.9, 0.95, 0.98, 0.99 and 0.999938 exceedance probabilities and the corresponding goodness of fit values resulting from a least-squares linear regression to estimate streamflows recursively from other streamflow quantiles. (*, Bias correction factor computed from Duan (1983).)

are, in part, computed from regional regression equations that were developed using the catchment characteristics from the approach discussed in Sect. 2.1. 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 software tool includes a message in the *BasinCharacteristics* worksheet (Fig. 4b) next to each characteristic that is outside the respective ranges of those characteristics used to solve the regression equations.

The *ReferenceGaugeSelection* worksheet (Fig. 4c) displays information about the ungauged catchment and donor streamgauge that was selected by the map correlation method described in Sect. 2.2; however, additional measures of similarity between the donor and ungauged location are also provided, including the percent difference between catchment characteristics at the ungauged location and the donor streamgauge as well as the distance between the ungauged location and donor streamgauge (Fig. 4c). The estimated cross-correlation resulting from the map-correlation method is also reported (Fig. 4c). If a user selects a new donor streamgauge, they then press the update button (Fig. 4c) and daily streamflows will be recomputed using the newly selected donor streamgauge. The *ContinuousFlowDuration* worksheet (Fig. 4d) displays the estimated FDC, and the *ContinuousDailyFlow* worksheet (Fig. 4e) displays the estimated daily time series for the ungauged site.

3.1 Demonstration area

The methods described in Sect. 2 were applied to the Connecticut River basin (CRB), located in the northeast United States, and incorporated into a basin-specific tool termed the Connecticut River UnImpacted Streamflow Estimator (CRUISE) tool. The CRUISE tool is freely available for download at <http://webdmamrl.er.usgs.gov/s1/sarch/ctrtool/index.html>. The CRB is located in the northeast United States and covers an area of approximately 29 000 km². The region is characterized by a temperate climate with distinct seasons. Snowfall is common from December through March, with generally more snow falling in the northern portion of the CRB than in the south. The geology and hydrology of the study region are 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 of base flows in the southern portion of the study region (Ries and Friesz, 2000). The CRB has thousands of dams along the main stem 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. To understand how dam management can be optimized

to meet both human and ecological needs for water, unregulated daily streamflows are needed to provide inflow time series to dams that can be routed through operation and optimization models being developed in the CRB.

3.2 Estimation of daily streamflow in the demonstration area

Data from streamgauges located within the CRB and surrounding area are used in the CRUISE tool to estimate unregulated daily streamflow time series at ungauged locations (Table 1). The study streamgauges have at least 20 yr of daily streamflow record and have minimal regulation in the contributing catchments to the streamgauges (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, from a larger set of 22 catchment characteristics, 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 and used in the streamflow estimation process. Contributing area to the study streamgauges ranges from 0.5 km² to 1 845 km² with a median value of 200 km². Mean annual precipitation ranges from 101 cm per year to 157 cm per year with a median value of 122 cm per year. 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. Streamflow in the CRUISE tool is estimated for a 44-yr (16 071-d) period spanning 1 October 1960 through 30 September 2004 using the methods described in Sect. 2.

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 exceedance probabilities were determined from the observed streamflow time series and regressed against 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 using the conventions described in Archfield et al. (2010). Regression equations were developed using weighted, least-squares multiple linear regression. Regression weights were applied to the dependent variables and computed as a function of the number of days of observed streamflow on which the estimated streamflow statistic was based. Natural-log transformations of the dependent variables (streamflow quantiles at selected exceedance probabilities) and independent variables (catchment characteristics) were made to effectively linearize the relations between the variables. Bias correction factors were estimated using the smearing estimator (Duan, 1983) to remove bias in the regression estimates of the streamflow quantiles when transferred out of logarithmic space. All non-zero regression coefficients in the regression equations (Table 2) were significantly different from zero at the 0.05 significance level.

Residuals (observed minus regression-estimated streamflow values) (plotted in log space) were generally homoscedastic and normally distributed. Variables in the final equations had variance-inflation factors of less than 2.5, meaning the correlations between the independent variables are minimal. Regression-coefficient values and goodness of fit values are shown in Table 2.

To enforce physical consistency as described in Sect. 2.1, streamflow quantiles at the 0.9, 0.95, 0.98, 0.99 and 0.999938 exceedance probabilities were recursively regressed against one another (Fig. 5). This approach also exploits the strong structural relation of the observed quantiles, as observed in Fig. 5. Linear regression equations were fit between the observed quantiles to establish a relation between the quantiles (Fig. 5); this relation was then carried recursively through the estimation of the FDC. For example, streamflow at the 85-percent exceedance probability is obtained by solving the multiple-linear regression equation that is a function of basin characteristics. However, streamflow at the 90-percent exceedance probability is obtained by the relation fit between the streamflows at the 85- and 90-percent exceedance probabilities (Fig. 5). Only the estimated streamflow at the 85-percent exceedance probability is needed to estimate the streamflow at the 90-percent exceedance probability. Subsequent streamflow quantiles are estimated from the relation between one quantile and another (Fig. 5). The remainder of the FDC curve was then estimated as described in Sect. 2.1.

Mapping of the cross-correlation for each of the study streamgauges was applied using the general approach described in Sect. 2.3 and in Archfield and Vogel (2010). Archfield and Vogel (2010) use the Pearson *r* correlation coefficient to model the cross-correlation across their study region. In this study, the Spearman rho cross-correlation metric is utilized. The Spearman rho cross-correlation metric is a non-parametric measure of cross-correlation that uses the ranks of the data; therefore, it is resistant to outliers and has fewer assumptions than the more commonly used Pearson *r* correlation coefficient (Helsel and Hirsch, 2002). As described by Archfield and Vogel (2010), spherical variogram models were fit for each study streamgauge. Variogram model (Eq. 3) parameters and root-mean-square errors between observed cross-correlations and cross-correlations estimated by the variogram model are shown in Table 3. The donor streamgauge and estimated FDC were then used to obtain continuous daily streamflow at the ungauged location, as described in Sect. 2.3.

3.3 Performance of estimated streamflows

To evaluate the utility of the underlying methods to estimate unregulated, daily streamflow at ungauged locations, a leave-one-out cross validation for 31 study streamgauges (Fig. 6) was applied in conjunction with the methods described in Sects. 2 and 3.2. These 31 study 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	15 December 1934 – 31 December 2004
01082000	Contoocook River at Peterborough, NH	7 July 1945 – 30 September 1977
01084500	Beard Brook near Hillsboro, NH	1 October 1945 – 30 September 1970
01085800	West Branch Warner River near Bradford, NH	22 May 1962 – 30 September 2004
01086000	Warner River at Davisville, NH	1 October 1939 – 30 September 1978
01089000	Soucook River near Concord, NH	1 October 1951 – 30 September 1987
01091000	South Branch Piscataquog River near Goffstown, NH	27 July 1940 – 30 September 1978
01093800	Stony Brook tributary near Temple, NH	1 May 1963 – 30 September 2004
01096000	Squannacook River near West Groton, MA	1 October 1949 – 31 December 2004
01097300	Nashoba Brook near Acton, MA	26 July 1963 – 31 December 2004
01105600	Old Swamp River near South Weymouth, MA	20 May 1966 – 24 July 2006
01105730	Indian Head River at Hanover, MA	8 July 1966 – 24 July 2006
01106000	Adamsville Brook at Adamsville, RI	1 October 1940 – 30 September 1978
01108000	Taunton River near Bridgewater, MA	1 October 1929 – 23 April 1976
01109000	Wading River near Norton, MA	1 June 1925 – 31 December 2004
01111300	Nipmuc River near Harrisville, RI	1 March 1964 – 30 September 1991
01111500	Branch River at Forestdale, RI	24 January 1940 – 31 December 2004
01117500	Pawcatuck River at Wood River Junction, RI	7 December 1940 – 31 December 2004
01118000	Wood River Hope Valley, RI	12 March 1941 – 31 December 2004
01118300	Pendleton Hill Brook near Clarks Falls, CT	1 October 1958 – 31 December 2004
01118500	Pawtucket River at Westerly, RI	27 November 1940 – 31 December 2004
01120000	Hop Brook near Columbia, CT	1 October 1932 – 6 October 1971
01121000	Mount Hope River near Warrenville, CT	1 October 1940 – 31 December 2004
01123000	Little River near Hanover, CT	1 October 1951 – 31 December 2004
01127880	Big Brook near Pittsburg, NH	1 December 1963 – 1 January 1984
01133000	East Branch Passumpsic River near East Haven, VT	1 October 1948 – 1 September 1979
01133500	Passumpsic River near St. Johnsbury, VT	1 May 1909 – 1 July 1919
01134500	Moose River at Victory, VT	1 January 1947 – 12 May 2010
01135000	Moose River at St. Johnsbury, VT	1 August 1928 – 1 September 1983
01137500	Ammonoosuc River at Bethlehem Junction, NH	1 August 1939 – 12 May 2010
01139000	Wells River at Wells River, VT	1 August 1940 – 12 May 2010
01139800	East Orange Branch at East Orange, VT	1 June 1958 – 12 May 2010
01140000	South Branch Waits River near Bradford, VT	1 April 1940 – 1 September 1951
01141800	Mink Brook near Etna, NH	1 August 1962 – 1 September 1998
01142000	White River near Bethel, VT	1 June 1931 – 1 September 1955
01144000	White River at West Hartford, VT	1 October 1951 – 12 May 2010
01145000	Mascoma River at West Canaan, NH	1 July 1939 – 1 September 1978
01153500	Williams River near Rockingham, VT	1 June 1940 – 1 September 1984
01154000	Saxtons River at Saxtons River, VT	20 June 1940 – 30 September 1982
01155000	Cold River at Drewsville, NH	23 June 1940 – 30 September 1978
01161500	Tarbell Brook near Winchendon, MA	29 May 1916 – 6 September 1983
01162500	Priest Brook near Winchendon, MA	1 October 1936 – 31 December 2004
01165500	Moss Brook at Wendell Depot, MA	1 June 1916 – 30 September 1982
01169000	North River at Shattuckville, MA	13 December 1939 – 31 December 2004
01169900	South River near Conway, MA	1 January 1967 – 31 December 2004
01171500	Mill River at Northampton, MA	18 November 1938 – 31 December 2004
01174000	Hop Brook near New Salem, MA	19 November 1947 – 30 September 1982
01174900	Cadwell Creek near Belchertown, MA	13 July 1961 – 30 September 1997
01175670	Seven Mile River near Spencer, MA	1 December 1960 – 31 December 2004
01176000	Quaboag River at West Brimfield, MA	19 August 1912 – 31 December 2004
01180000	Sykes Brook at Knightville, MA	20 June 1945 – 18 July 1974
01181000	West Branch Westfield at Huntington, MA	1 September 1935 – 31 December 2004
01187300	Hubbard River near West Hartland, CT	4 August 1959 – 31 December 2004
01187400	Valley Brook near West Hartland, CT	1 October 1940 – 30 September 1972
01188000	Burlington Brook near Burlington, CT	1 October 1931 – 31 December 2004
01193500	Salmon River near East Hampton, CT	1 October 1928 – 31 December 2004
01194500	East Branch Eightmile River near North Lyme, CT	1 October 1937 – 6 October 1981
01198000	Green River near Great Barrington, MA	1 October 1951 – 30 September 1971
01198500	Blackberry River at Canaan, CT	1 October 1949 – 20 October 1971
01199050	Salmon Creek at Lime Rock, CT	1 October 1961 – 31 December 2004
01200000	Ten Mile River, CT	1 October 1930 – 4 April 1988
01332000	North Branch Hoosic River at North Adams, MA	22 June 1931 – 30 September 1990
01333000	Green River at Williamstown, MA	20 September 1949 – 31 December 2004

Table 2. Number of streamgauges, goodness of fit values, explanatory variables, and estimated regression parameters for streamflow quantiles estimated from catchment characteristics using multiple least squares regression (* Bias correction factor computed from Duan, 1983).

Exceedance probability	General regression information			Estimated regression coefficients				
	Number of streamgauges used to develop regression equation	Percent root-mean-square error	Constant term	Drainage area	Average annual precipitation.	Percent of basin that is underlain by sand and gravel deposits	Y-location of the basin centroid	Bias correlation factor*
0.02	51	1.49	-26.5758	0.9590	2.3262	0	1.4462	1.0103
0.05	51	0.62	-19.3148	0.9775	1.7521	0	1.0457	1.0023
0.1	51	0.73	-2.1224	0.9982	0.9106	0	0	1.0015
0.15	51	0.60	-2.9777	1.0050	1.0589	0	0	0.9972
0.2	51	0.86	-3.6935	1.0037	1.1920	0	0	0.9957
0.25	51	1.32	-4.6684	1.0110	1.3890	0	0	0.9950
0.3	51	1.86	-5.5394	1.0137	1.5688	0	0	0.9950
0.4	51	3.00	-6.7591	1.0206	1.8000	0	0	0.9960
0.5	51	3.86	-7.6803	1.0269	1.9577	0	0	0.9982
0.6	50	4.40	-8.3466	1.0184	2.0123	0.0804	0	1.0184
0.7	50	6.61	-8.4500	1.0480	1.9072	0.0949	0	1.0278
0.75	50	9.24	-8.7450	1.0655	1.9073	0.1040	0	1.0243
0.8	50	13.58	-9.1085	1.0951	1.9008	0.1251	0	1.0379
0.85	50	21.20	-9.3154	1.1239	1.8480	0.1515	0	1.0565

were selected, because they have observed streamflow covering the entire 44-yr historical period of streamflow estimated by the CRUISE tool. In the leave-one-out cross validation, each of the 31 study streamgauges was assumed to be ungauged and removed from the methods described in Sect. 2 and 3.2. The methods were then reapplied without inclusion of the removed site. Using the catchment characteristics of the removed site, daily streamflow was determined and compared to the observed streamflow data at the removed streamgauge. This cross-validation procedure ensured that the comparison of observed and estimated streamflow at each of the study streamgauges represented the truly ungauged case, because the streamgauge was not used in any part of the methods development. This procedure was repeated for each of the 31 validation streamgauges to obtain 31 estimated and observed streamflow time series from which to assess the performance of the study methods.

Goodness of fit between observed and estimated streamflows 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. 6a). The natural logarithms of the observed and estimated streamflows were 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. 6b). Selected hydrographs were also plotted to visualize the interpretation of the efficiency values (Figs. 6c–e).

The values in Fig. 4 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. 6a), with an efficiency value equal to 1 indicating 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. 6a). Despite this, the CRUISE tool appears to result in high efficiency values across all validation sites (Fig. 6). Streamgauges in the northern portion of the basin have lower efficiency values than streamgauges in the middle and southern portions of the basin; however, it should be noted from the hydrographs in Fig. 4 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 efficiency values and hydrograph comparisons demonstrate that the CRUISE tool can provide a reasonable representation of natural streamflow time series at ungauged catchments in the basin.

4 Discussion

As described, the software tool can be viewed as a general framework to provide estimates of daily streamflow in a publicly available, map-based manner. Whereas, the StreamStats user interface was developed specifically for the CRB, the watershed delineation and catchment characteristic algorithms underlying StreamStats are universally available across the globe through the ArcHydro platform (ESRI, Inc., 2009). To utilize the ArcHydro platform, a properly networked stream data layer is needed, which uniquely identifies each stream reach and provides such information as flow direction (Reis et al., 2008). Such a network is freely available for the United States and is termed the National Hydrography Dataset (NHD) (available at: <http://nhd.usgs.gov/>). It is likely that other regions around the globe already have

Table 3. 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

such a dataset developed. In addition to the stream network, region-wide spatial data layers of catchment characteristics are needed so that these characteristics can be computed at the ungauged location and used to solve the regression equations. If the stream network and spatial data layers of catchment characteristics are readily available, this software framework can be easily applied towards a map-based tool to

provide estimates of daily streamflow. The underlying data in the macro-enabled spreadsheet can then be customized to the catchment characteristics, fitted regression equations, and fitted variogram models to link with the catchment delineation.

There are several limitations to the methods described in the software tool. Notably, the software tool assumes that the topographic surface water divides of the catchment are

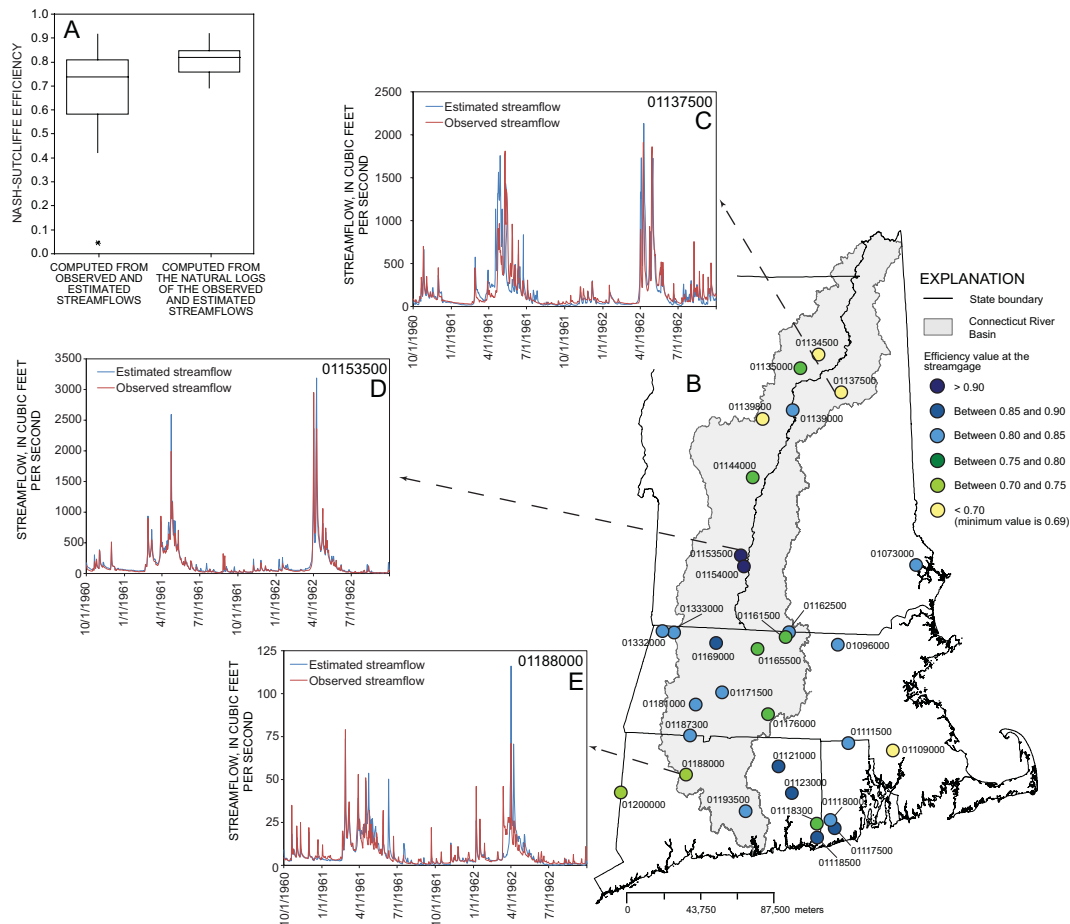


Fig. 6. Range of efficiency values computed between the observed and estimated streamflows at the 31 validation streamgages (A), spatial distribution of efficiency values resulting from log-transformed observed and estimated daily streamflow at 31 validation streamgages (B) and selected hydrographs of observed and estimated streamflow for the period from 1 October 1960 through 30 September 1962 (C–E). The boxplot (A) shows the median, interquartile ranges and the upper and lower limits (defined as 75th percentile $1.5 \times$ (75th percentile–25th percentile)). Values outside of the upper and lower limits are shown as an asterisk.

coincident with the underlying groundwater divides. Therefore, the tool assumes that water draining to the stream location of interest is contained entirely within the topographic catchment divides. For regions dominated by groundwater flow, this assumption may not be valid. The methods underlying the tool also currently do not account for routing, which is an important consideration for large catchment areas whose response to precipitation events may exceed more than a few days. Lastly, the purpose of the software tool is to provide reliable estimates of historical streamflow time series for an ungauged location, and non-stationarity is not explicitly considered in the underlying methods. By excluding streamgages in the method development that may have been affected by human use such as dams or water withdrawals, the effects of non-stationarity are seemingly minimized; however, no attempt was made to explicitly remove study streamgages affected by climate non-stationarity in the daily streamflow signal.

5 Summary and conclusions

This paper presents one of the first publicly available, map-based software tools to provide unregulated daily streamflow time series (streamflow not affected by human regulation such as dams or water withdrawals) for any user-selected river location in a particular study region. In this study, the software tool was developed and presented for the Connecticut River basin – a 29 000 km² river basin located in the northeast United States. For other regions, this study presents an overall framework, which can be applied toward development of a region-specific tool to estimate daily streamflow at any user-selected river location. The software tool is available at <http://webdamr1.er.usgs.gov/s1/sarch/ctrltool/index.html> and requires only an internet connection, a web browser program, and a macro-based spreadsheet program. Furthermore, the underlying data used to develop the tool and the source code are freely available and adaptable to other

regions. Daily streamflow is estimated by a four-part process: (1) delineation of the drainage area and computation of the basin characteristics for the ungauged location, (2) selection of a donor streamgauge, (3) estimation of the daily flow-duration curve at the ungauged location, and (4) use of the donor streamgauge to transfer the flow-duration curve to a time series of daily streamflow. The software tool, when applied to the Connecticut River basin, provided reliable estimates of observed daily streamflows at 31 validation streamgauges across the basin. This software framework and underlying methods can be used to develop map-based, daily-streamflow estimates needed for water management decisions at ungauged stream locations for this and potentially other regions.

Acknowledgements. The authors would like to acknowledge Kimberly Lutz and Colin Ape of The Nature Conservancy, Richard Palmer, Casey Brown and Scott Steinschneider of the University of Massachusetts, Amherst, Austin Polebitski of the University of Wisconsin, Platteville, and Christopher Hatfield, Woodrow Fields, and John Hickey of the US 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 US 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 US Geological Survey, who provided a list of gauges on streams not affected by dams or water withdrawals streamgauges located in northern portion of the CRB. The authors also thank two anonymous reviewers and David Bjerklie of the US Geological Survey whose comments improved the manuscript substantially. This work was made possible by the New England Association of Fish and Wildlife Agencies Northeast Regional Conservation Needs Grant number 2007-06 and The Nature Conservancy. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

Edited by: D. Lawrence

References

- Abdulla, F. A. and Lettenmaier, D. P.: Development of regional parameter estimation equations for a macroscale hydrologic model, *J. Hydrol.*, 197, 230–257, ISSN 0022-1694, doi:10.1016/S0022-1694(96)03262-3, 1997.
- Archfield, S. A. and Vogel, R. M.: Map correlation method: Selection of a reference streamgauge to estimate daily streamflow at ungauged catchments, *Water Resour. Res.*, 46, W10513, doi:10.1029/2009WR008481, 2010.
- Archfield, S., Vogel, R., Steeves, P., Brandt, S., Weiskel, P., and Garabedian, S.: The Massachusetts Sustainable-Yield Estimator: A decision-support tool to assess water availability at ungauged sites in Massachusetts, U.S. Geological Survey Scientific Investigations Report 2009-5227, 2010.
- Armstrong, D. S., Parker, G. W., and Richards, T. A.: Characteristics and classification of least altered streamflows in Massachusetts, U.S. Geological Survey Scientific Investigations Report, 20075291, 2008.
- Blöschl, G. and Sivapalan, M.: Scale issues in hydrological modeling – a review, *Hydrolog. Process.*, 9, 251–290, 1995.
- Castellari, A., Galeati, G., Brandimarte, L., Montanari, A., and Brath, A.: Regional flow-duration curves: reliability for ungauged basins, *Adv. Water Resour.*, 27, 10, 953–965, 2004.
- Duan, N.: Smearing estimate – a nonparametric retransformation method, *J. Am. Stat. Assoc.*, 78, 383, 605–610, 1983.
- ESRI, Inc.: Arc-Hydro Tools – Tutorial, Version 1.3 – January 2009, ESRI, Inc., Redlands, CA, <http://andersonruhoff.googlepages.com/ArcHydro-Tutorial.pdf>, 2009.
- Falcone, J. A., Carlisle, D. M., Wolock, D. M., and Meador, M. R.: GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States, *Ecology*, 91, 621, 2010.
- Fennessey, N. M.: A hydro-climatological model of daily streamflow for the northeast United States, Ph.D. dissertation, Tufts University, Department of Civil and Environmental Engineering, 1994.
- Helsel, D. and Hirsch, R.: Statistical Methods in Water Resources Techniques of Water Resources Investigations, Book 4, Chapt. A3, US Geological Survey, 2002.
- Hirsch, R.: Evaluation of some record reconstruction techniques, *Water Resour. Res.*, 15, 6, 1781–1790, ISSN 0043-1397, 1979.
- Holtschlag, D. J.: Application guide for AFINCH, analysis of flows in networks of channels) described by NHDPlus, U.S. Geological Survey Scientific Investigations Report 2009-5188, 2009.
- Hughes, D. A. and Smakhtin, V. U.: Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves, *Hydrolog. Sci. J.*, 41, 6, 851–871, 1996.
- Isaaks, E. H. and Srivastava, R. M.: An Introduction to Applied Geostatistics, first ed., Oxford University Press, New York, 1989.
- Mohamoud, Y. M.: Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves, *Hydrolog. Sci. J.*, 53, 706–724, 2008.
- McIntyre, N., Lee, H., Wheeler, H. S., Young, A., and Wagener, T.: Ensemble prediction of runoff in ungauged watersheds *Water Resour. Res.*, 41, W12434, doi:10.1029/2005WR004289, 2005.
- Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, *J. Hydrol.*, 287, 95–123, ISSN 0022-1694, doi:10.1016/j.jhydrol.2003.09.028, 2004.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – a discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, 44, W03413, doi:10.1029/2007WR006240, 2008.
- Oudin, L., Kay, A., Andréassian, V., and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, *Water Resour. Res.*, 46, W11558, doi:10.1029/2009WR008887, 2010.
- Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, *Hydrol. Earth Syst. Sci.*, 9, 157–171, doi:10.5194/hess-9-157-2005, 2005.
- Patil, S. and Stieglitz, M.: Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catch-

- ment, *Hydrol. Earth Syst. Sci.*, 16, 551–562, doi:10.5194/hess-16-551-2012, 2012.
- Poff, N. L., Allen, J. D., Bain, M. B., Karr, J. R., Prestagard, K. L., Richter, B. D., Sparks, R. E., and Stromberg, J. C.: The natural-flow regime – A paradigm for river conservation and restoration: *Bioscience*, 47, 769–784, 1997.
- Poff, N. L., Richter, B., Arthington, A., Bunn, S. E., Naiman, R. J., Kendy, E., Acreman, M., Apse, C., Bledsoe, B. P., Freeman, M., Henriksen, J., Jacobsen, R. B., Kennen, J., Merritt, D. M., O’Keefe, J., Olden, J., Rogers, K., Tharme, R. E., and Warner, A.: The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards, *Freshwater Biology*, 55, 147–170, 2010.
- Reichl, J., Western, A., McIntyre, N., and Chiew, F.: Optimisation of a similarity measure for estimating ungauged streamflow, *Water Resour. Res.*, doi:10.1029/2008WR007248, 2009.
- Ries, K. G. and Friesz, P. J.: Methods for Estimating Low-Flow Statistics for Massachusetts Streams: US Geological Survey Water-Resources Investigations Report 2000-4135, 2000.
- Ries, K. G., III, Guthrie, J. G., Rea, A. H., Steeves, P. A., and Stewart, D. W.: StreamStats: A Water Resources Web Application, U.S. Geological Survey Fact Sheet 2008-3067, <http://pubs.usgs.gov/fs/2008/3067/>, 2008.
- Ribeiro Jr., P. J. and Diggle, P. J.: geoR: A package for geostatistical analysis: *R-NEWS* vol. 1, no. 2, ISSN 1609-3631, 2001.
- Seibert, J.: Regionalization of parameters for a conceptual rainfall – runoff model, *Agr. Forest Meteorol.*, 98, 279–293, 1999.
- Shu, C. and Ouarda, T. B. M. J.: Improved methods for daily streamflow estimates at ungauged sites, *Water Resour. Res.*, 48, W02523, doi:10.1029/2011WR011501, 2012.
- Sivapalan, M. et al.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrol.Sci.J.*, 48, 857–880, 2003.
- Skøien, J. O. and Blöschl, G.: Spatiotemporal topological kriging of runoff time series, *Water Resour. Res.*, 43, 9, doi:10.1029/2006WR005760, 2007.
- Smakhtin, V. U.: Generation of natural daily flow time-series in regulated rivers using a non-linear spatial interpolation technique, *Regul. Rivers: Res. Mgmt*, 15, 311–323, 1999.
- Smakhtin, V. U. and Eriyagama, N.: 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, 2008.
- Williamson, T. N., Odom, K. R., Newson, J. K., Downs, A. C., Nelson H. L., Jr., Cinotto, P. J., and Ayers, M. A.: 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, 2009.
- Zhang, Y. and Chiew, F. H. S.: Relative merits of different methods for runoff predictions in ungauged catchments, *Water Resour. Res.*, 45, W07412, doi:10.1029/2008WR007504, 2009.