Response to Anonymous Referee #2

The paper presents a first attempt to develop a software tool to generate daily stream flow time series at an arbitrary location on a stream network. In my opinion, such a tool will be an extremely useful addition to the "hydrology toolbox" currently used by scientists and water practitioners. The paper is generally well written and structured. However, I do see the need for a number of adjustments to improve its clarity and comprehensibility (not only to scientists but also to general water planners), before it is published, as set out below.

Thank you for your comments and review. We have revised the manuscript substantially to address these concerns, including adding new sections to the text, expanding existing sections and revising the figures to provide more detail and clarity. We address your general and specific comments related to these points below.

General Comments

1. As far as I understand, the tool, as it is currently applied in the Connecticut River Basin consists of two parts: (a) The StreamStats tool to delineate watershed boundary and basin characteristics, and (b) The spreadsheet tool which performs the rest of the estimation procedure. Currently only (a) works online (web-based) while (b) works offline (non- web-based), whereas the title of the paper suggests the existence of a fully "web-based" operational tool. It is not clear whether the intention of the authors is to develop the current tool into a fully web-based tool in the future. If so, this is not stated in the text. The tool in its present form appears to be a "work in progress" towards a fully web-based tool. Hence I suggest that the title of the paper is amended to read as "Developing a web-based software tool. . .. ." or "Towards a webbased software tool. . .. ."

We agree and revised the titled as, "Towards a publicly-available, map-based regional software tool to estimate unregulated daily streamflow at ungauged rivers."

2. The StreamStats tool contains information limited to US watersheds and only those in some states at that. The authors do not sufficiently explain how the tool may be practically implemented in any other part of the world, including the underlying data of the StreamStats tool, the components, applications and functionality expected in such a tool, and what data is output by it (See specific comment on "basin Characteristics" given below). The CRUISE worksheet tool too is specific to the Connecticut River Basin as far as I understand. What features should be included in this tool if it is to be implemented in any other region of the world? It would also be useful to know how practical it is to build the two separate components into one compact standalone software (whether web-based or not).

We agree this is not clear. We have added additional text to the methods section (Section 2) to describe how the methods can be applied to other regions as well as text as to how the tool can be viewed as a framework. Specifically, we add in Section 3:

"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. Furthermore, all data and methods underlying 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 is discussed in Section 4."

We also added a new Discussion section to the text with a paragraph to describe the underlying data and methods needed to develop a watershed delineation tool for other regions across the globe.

3. The paper is written assuming that the reader is familiar with all the methods mentioned in it. For example the "map-correlation" method is referred to in several places, but nowhere is it explained. The authors also state that the FDC at the ungauged site is estimated using regional regression equations based on basin characteristics, but do not elaborate further on what specific characteristics are considered or what the form and type of the regression equations are (Also see specific comments below). 4. The text does not sufficiently explain the information presented in figures and tables leaving it to the reader to figure them out, which makes the reader's life extremely difficult. All figures are too small and it is next to impossible to read some of them (especially Figures 3 and 4). Also see some specific comments below on figures.

We agree with this comment. We have substantially expanded the methods sections to provide equations and text for the regression equations and map correlation method used in the software tool. We have added a new section (now Section 3.2) to the text titled "Estimation of daily streamflow in the demonstration area." This section describes how the methods introduced in Section 2 were implemented for the demonstration area. Figure 3 has been broken into two figures (now figures 3 and 4) and figure 4 (now figure 6) has been reworked so that the hydrographs are larger and more visible. Figure 1 has also been completely revised to show additional detail on the software methods.

Specific Comments

1. The one before the last sentence in the abstract reads as "For the demonstration region. . .. .with efficiency values computed from observed and estimated streamflows ranging from 0.69 to 0.92". I suggest that the term "efficiency" here is qualified as "Nash-Sutcliff Efficiency" or the sentence is reworded in another way to indicate that the values presented are goodness of fit statistics, in order not to confuse readers by using the term "efficiency" which could mean any number of things. Also it is noted that the 0.69 to 0.92 values have been obtained using natural logarithms of generated streamflows while the same for untransformed streamflows is 0.04 to 0.92. Generally, goodness of fit statistics are evaluated against untransformed streamflows, and presenting the statistics for the transformed streamflows in the abstract might mislead *readers about how good the suggested methods are. I suggest that the actual (0.4 t0 0.92) values are reported in the abstract and that the authors try to identify the reasons behind this large variation (for example the method may work well for only a certain range in watershed area).*

We agree and changed this sentence in the abstract to read:

"For a demonstration region in the northeast United States, daily streamflow was, in general, shown to be reliably estimated by the software tool, with more difficulty estimating the highest and lowest streamflows that occurred over the historical period from 1960 through 2004."

2. Line 8-9 on page 2507 reads as ". . .. .first developing regional regressions relating catchment characteristics to selected FDC quantiles. . .. .", but does not elaborate on what type of catchment characteristics are considered here, leaving the reader guessing. Neither are they explained later, apart from within the section on the CRB where only three characteristics are discussed. The authors should present a broad range of possible characteristics which may be adopted in any other part of the world if the tool is to be reproduced.

We agree. To address this comment, we have added more detail to this section on the regression approach, including how catchment characteristics are selected as well as present the general form of the regression equation. We also added the statement:

"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."

We also now added a new section (now Section 3.2), which details the specific methods used to develop the regression equations for the demonstration area. We discuss in this section the rationale for the specific basin characteristics used in the demonstration area and the details of the regression fits.

3. Line 13 page 2507: ". . .. . .selected quantiles on the FDC are estimated from regional regression equations and a continuous. . .. .". What is the form of these regression equations and how are the catchment characteristics related to FDC quantiles in these equations? Without this knowledge, the tool cannot be reproduced anywhere else. Although these regression equations are mentioned even later in the text at several places, nowhere are they presented. Merely referring to another paper where the method has been applied is not sufficient for a reader of this paper to understand the procedures presented here.
4. Tables 2 and 3 do present information purported to be on these regression equations, however, they are not at all helpful since (a) the equations themselves are not explained in the text, and (b) the tables are utterly confusing leaving the reader guessing as to what most of these columns stand for. The tool should be understandable to

any interested party who wants to reproduce it for water management purposes.

We completely agree and address these two comments together. We have substantially expanded the text in Section 2.1 to include the general form of the regression equations and additional text to describe the regressions, included in this section is text such as:

"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 the new section 3.2, we describe the reasons for why certain catchment characteristics were selected for the demonstration region:

"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."

We also broke out the lower portion of Figure 2 into a new figure (now Figure 5), which shows the relations between the streamflow quantiles for the high exceedence probabilities and combines table 3 and Figure 2. We simplified Table 2 and spelled out abbreviations in the column headings to make the table easier to read.

5. Line 1 in Page 2507 first mentions the "map-correlation" method, but does not explain how the cross correlation takes place between the ungauged site and the index stream gauge. For example, what specific characteristics are correlated, and what equations are used?

We agree. We have substantially expanded this section and the description of the method, adding equations to the text and further justification for its use in the software tool.

6. Line 23 in page 2513 refers to "leave-one-out" cross validation, but does not explain the rationale behind it. I suggest that it is explained at least in broad terms, since this is not a standard term that one comes across every day.

We agree. We have added a new paragraph to describe this validation approach:

"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 Sections 2 and 3.2. These 31 study streamgauges 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 Sections 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."

7. Line 14, page 2514 says ". . .. . .from the Cruise tool at high streamflow values is more of a challeng. . .. ." I am not sure how the difference between goodness of fit values for the transformed and untransformed streamflows explains that only high values (or both high and low values for that matter) are a challenge. Might not this difference be caused by discrepancies in mid-range values too?

We agree that it is not clear why this is the case. We removed this statement from the manuscript.

8. In Figure 2, text on the top graph which reads as "Flow quantiles greater than 0.01" should read as "Flow quantiles less than 0.01" if I understand the text correctly. Figure 3 is extremely small and none of the screen shots are clearly visible. I think it is better to break this figure into 2 and expand the size of each screen-shot to have more clarity. All numbers and lettering in Figure 4 is too small to read, while the comparisons between observed and generated streamflows (graphs) are not at all visible to the naked eye. However, I think the figure itself represents a neat way of summarizing the goodness of fit information, if it can be made larger and the signs for different efficiency ranges are made distinct from each other.

We agree with this comment and thank the reviewer for their suggestions. We made the suggested changes to the text on Figure 2. We also broke Figure 3 into two figures (now figures 3 and 4). We have now broken Figure 2 into two figures (Figures 2 and 5) – the first is a more general figure used in the methods section to describe how the flow-duration curve is estimated. The second figure (now figure 5) is referenced in the new section (Section 3.2) titled "Estimation of daily streamflow in the demonstration area," which uses the data from the demonstration area to show the relation between the streamflow quantiles for the demonstration area.

9. Line 15 on page 2510 refers to a "Microsoft Excel" spreadsheet. However, as far as I understand the spreadsheet doesn't necessarily have to be a "Microsoft Excel" one. Any spreadsheet program with capability to run macros, or perhaps a standalone code to perform the underlying procedures may be used instead. Perhaps the authors need to qualify that they have currently used a "Microsoft Excel" spreadsheet (If this journal is okay with the use of brand names), but the same functionality may be obtained by other means.

We agree and added text to clarify that any macro-enabled spreadsheet program will suffice (see Section 3, end of paragraph 2):

"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 Section 2 have been applied. Additionally, any macro-enabled spreadsheet program could be used in place of the Microsoft Excel spreadsheet program."

We have also added the standard U.S. Government disclaimer in the acknowledgements sections:

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

10. Other comments of minor nature are:
(a) Use the word "often" instead of "often times" (b) Line 19, page 2506: use "characteristics of" instead of "characteristics computed for" (c) Word "recursively" is spelled wrong in Fig. 2 (d) Line 21, page 2509: typo "by published Smakhtin(1999)" to be corrected as "published by" (e) Line 25 page 2509: typo "on the same day as" to be corrected as "on the same day at" (f) Line 15, page 2512: "Fig.1" should perhaps be "Fig.4"? (g) Line 10, page 2514: the word "indicating" is spelled wrong.

Thank for your thorough review and catching these errors. We have made all recommended changes.

REVISED MANUSCRIPT

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

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

----

[1]Massachusetts-Rhode Island Water Science Center, U.S. Geological Survey, 10 Bearfoot Road, Northborough, MA 01532, USA

[2]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

[3]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 critical for addressing any number of hydrologic problems. Often, streamflow information is needed at locations which 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, with more difficulty estimating the highest and lowest streamflows that occurred over the historical period from 1960 through 2004. The software tool provides a general framework that can be applied to other regions for which daily streamflow estimates are needed.

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

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; *Poff et al.*, 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 unmonitored 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 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 transferred to an ungauged: 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 location [*Merz and Blöschl*, 2004; *McIntyre et al.*, 2005; *Parajka et al.*, 2005; *Oudin et al.*, 2008, *Zhang and Chiew*, 2009, *Reichl et al.*, 2009; *Oudin et al.*, 2010]. 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 streamgauge 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 and others*, 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 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 (Section 2.1; fig. 1B). The FDC can be considered analogous 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 streamgauges in the study region and measurable catchment characteristics obtained for the contributing areas to those streamgauges (Section 2.1; fig. 1B). Interpolation is then used to obtain the FDC values for streamflows between the regression-estimated quantiles (Section 2.1; fig. 1B). Lastly, the FDC at the ungauged location is transformed into a time series of streamflow by the selection (Section 2.2; fig. 1C) and use (Section 2.3; fig. 1D) of a donor streamgauge. To ensure that the estimated streamflow represents unregulated conditions, only streamgauges whose catchments have been unaffected 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] provides 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 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 streamgauged, 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 probability 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 exceedence probability and some estimated streamflow values were larger at higher exceedence probabilities than streamflows estimated at lower exceedence probabilities. Thus, the inherent structure of the data that ensures streamflow quantiles decrease with increasing exceedence 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 exceedence 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 equation (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 exceedence 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. *Arcfield 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 (eqn. 2) in conjunction with the selected donor streamgauge:

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

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 and *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

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

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 by 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 October, 1, 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 exceedence probabilities at the donor and ungauged FDC are then equated (fig. 1D) and the date that each exceedence 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 in other regions of the United States and possibly other areas. Furthermore, all data and methods underlying 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 is discussed in Section 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 is 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 Section 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 equations described in Sections 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 figures 1B to 1D and Section 2. The estimated streamflows are, in part, computed from regional regression equations that were developed using the catchment characteristics from the approach discussed in Section 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 Section 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 Sections 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 km^2. 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 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. 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 years 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 km^2 to 1,845 km^2 with a median value of 200 km^2. 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 October 1, 1960 through September 30, 2004 using the methods described in Section 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 exceedence 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 Section 2.1, streamflow quantiles at the 0.9, 0.95, 0.98, 0.99 and 0.999938 exceedence probabilities were recursively regressed against one another (fig. 5). This approach also exploits the strong structural relation of the observed quantiles, as observed in figure 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 exceedence probability is obtained by solving the multiple-linear regression equation that is a function of basin characteristics. However, streamflow at the 90-percent exceedence probability is obtained by the relation fit between the streamflows at the 85- and 90-percent exceedence probabilities (fig 5). Only the estimated streamflow at the 85-percent exceedence probability is needed to estimate the streamflow at the 90-percent exceedence 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 Section 2.1.

Mapping of the cross-correlation for each of the study streamgauges was applied using the general approach described in Section 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 (eqn. 3) parameters and root-mean-square errors between observed cross-correlations and cross-correlations estimated by the variogram model are shown in table 2. The donor streamgauge and estimated FDC were then used to obtain continuous daily streamflow at the ungauged location, as described in Section 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 Sections 2 and 3.2. These 31 study streamgauges 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 Sections 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 figure 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 figure 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 is 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 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 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 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 streamgauges in the methods 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 streamgauges 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 large 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://webdmamrl.er.usgs.gov/s1/sarch/ctrtool/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 Apse 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 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. The authors also thank two anonymous reviewers and David Bjerklie of the U.S. Geological Survey whose comments improved the manuscript substantially. 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. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Abdulla, F. A., and D. P. Lettenmaier (1997), Development of regional parameter estimation equations for a macroscale hydrologic model, *J. Hydrol.*, 197(1-4), 230-257, ISSN 0022-1694, doi: 10.1016/S0022-1694(96)03262-3.

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.

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.

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

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.

McIntyre, N., H. Lee, H.S. Wheater, A. Young and T. Wagener (2005), Ensemble prediction of runoff in ungauged watersheds Water Resourc. Res., 41, W12434, doi: 10.1029/2005WR004289.

Merz, R. and G. Bloschl (2004), Regionalisation of catchment model parameters, J. Hydrol., 287(1-4), 95-123, ISSN 0022-1694, doi: 10.1016/j.jhydrol.2003.09.028.

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.

Patil, S. and Stieglitz, M.: Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment, Hydrol. Earth Syst. Sci., 16, 551-562, doi:10.5194/hess-16-551-2012, 2012.

Oudin, L., V. Andréassian, C. Perrin, C. Michel, and N. Le Moine (2008), 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.

Oudin, L., A. Kay, V. Andréassian, and C. Perrin (2010), Are seemingly physically similar catchments truly hydrologically similar?, Water Resour. Res., 46, W11558, doi:10.1029/2009WR008887.

Parajka, J., R. Merz, and G. Blöschl (2005), A comparison of regionalisation methods for catchment model parameters. Hydrol. Earth Syst. Sc., 9(3), 157-171.

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.

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., Warner A., 2010, The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards, Freshwater Biology 55: 147–170.

Reichl J, Western A., McIntyre N and Chiew F. (2009), Optimisation of a similarity measure for estimating ungauged streamflow, Water Resour. Res., doi:10.1029/2008WR007248.

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/.

Seibert, J. (1999) Regionalization of parameters for a conceptual rainfall–runoff model, Agric. For. Met., 98, 279–293.

Shu, C., and T. B. M. J. Ouarda, 2012, Improved methods for daily streamflow estimates at ungauged sites, *Water Resour. Res.*, 48, W02523, doi:10.1029/2011WR011501.

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.

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.

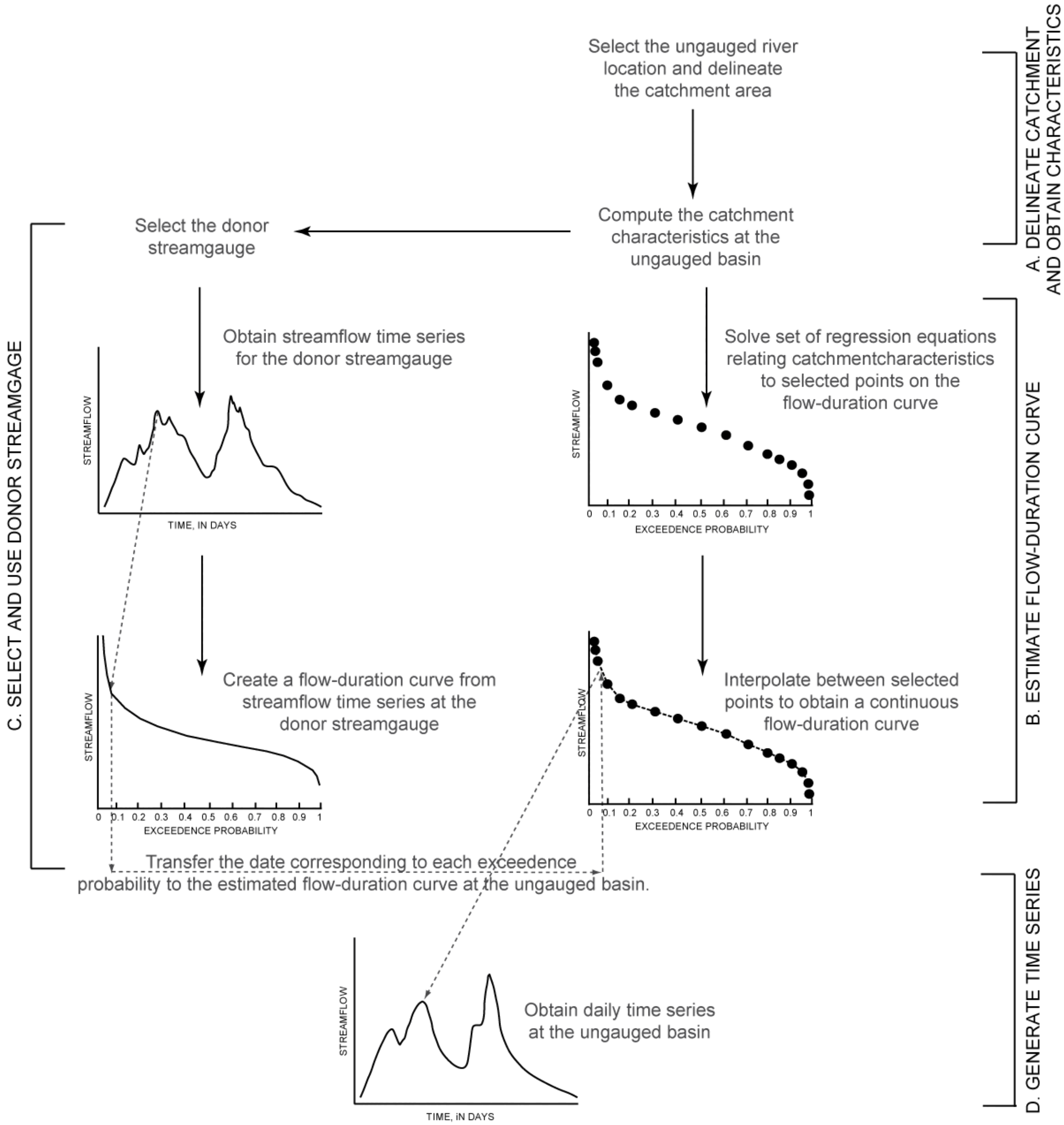

Figure 1. Diagram of the process to estimate unregulated, daily streamflow at ungauged locations. An ungauged river location is selected and the catchment characteristics are computed (A). The flow-duration curve is then estimated using regression relations between the catchment characteristics and selected 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).

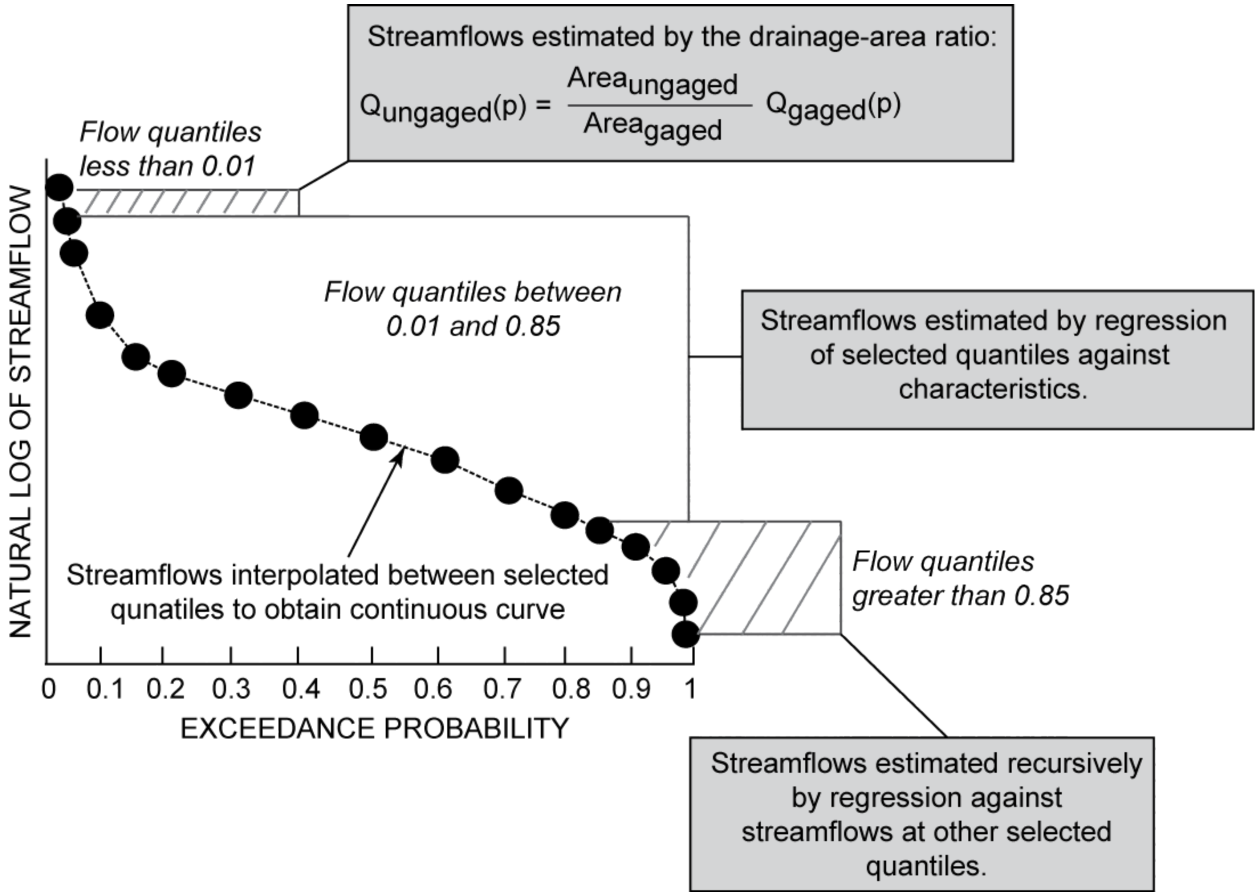

Figure 2. Diagram showing the methods used to estimate a continuous, daily flow duration at an ungauged location.

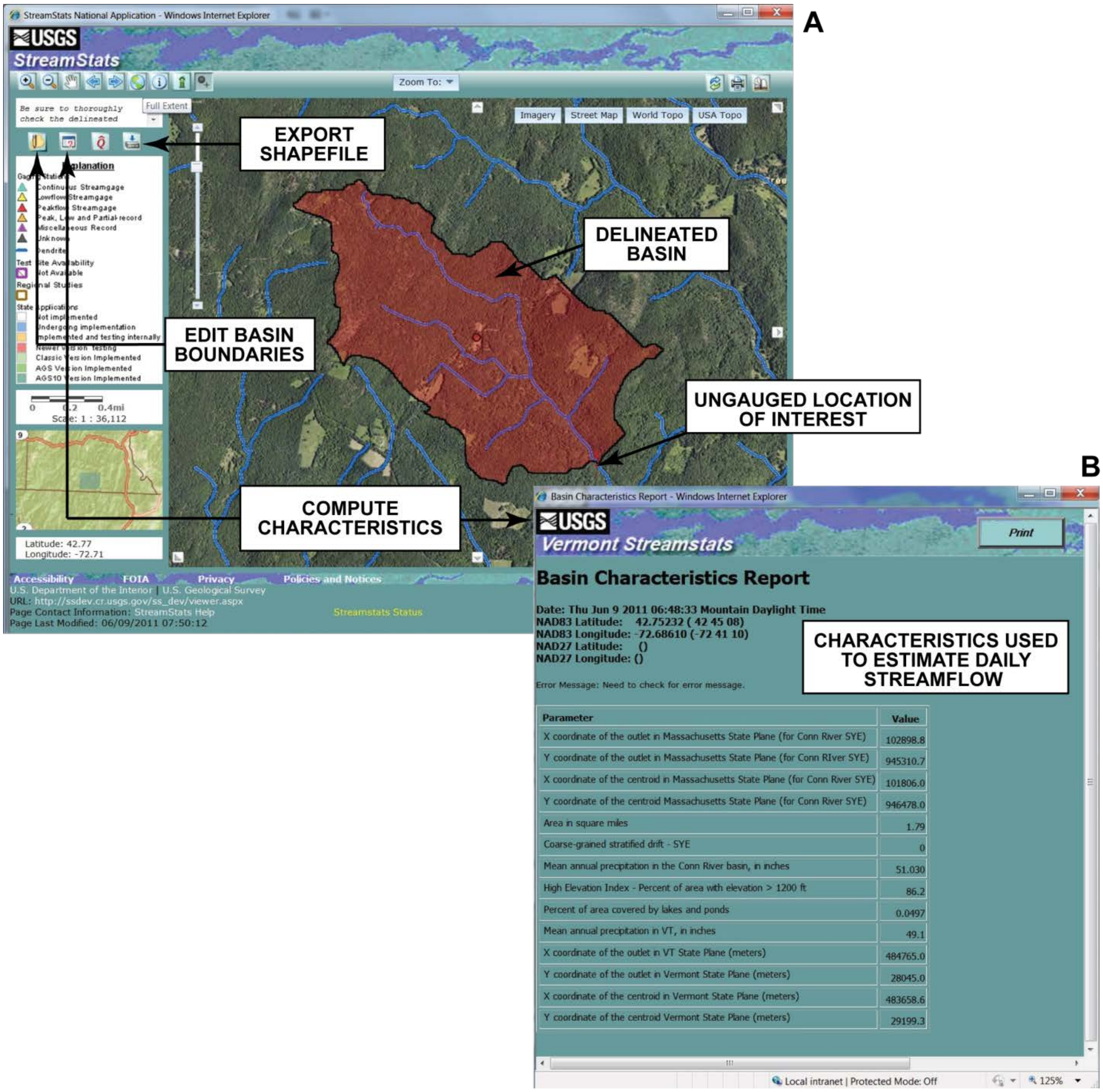

Figure 3. Screen captures showing the map portion of the software tool used to estimate daily, unregulated time series. The program delineates a catchment (or basin, as named in the tool) for the ungauged location selected by the user (A) and summarizes the catchment characteristics (B). The user also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

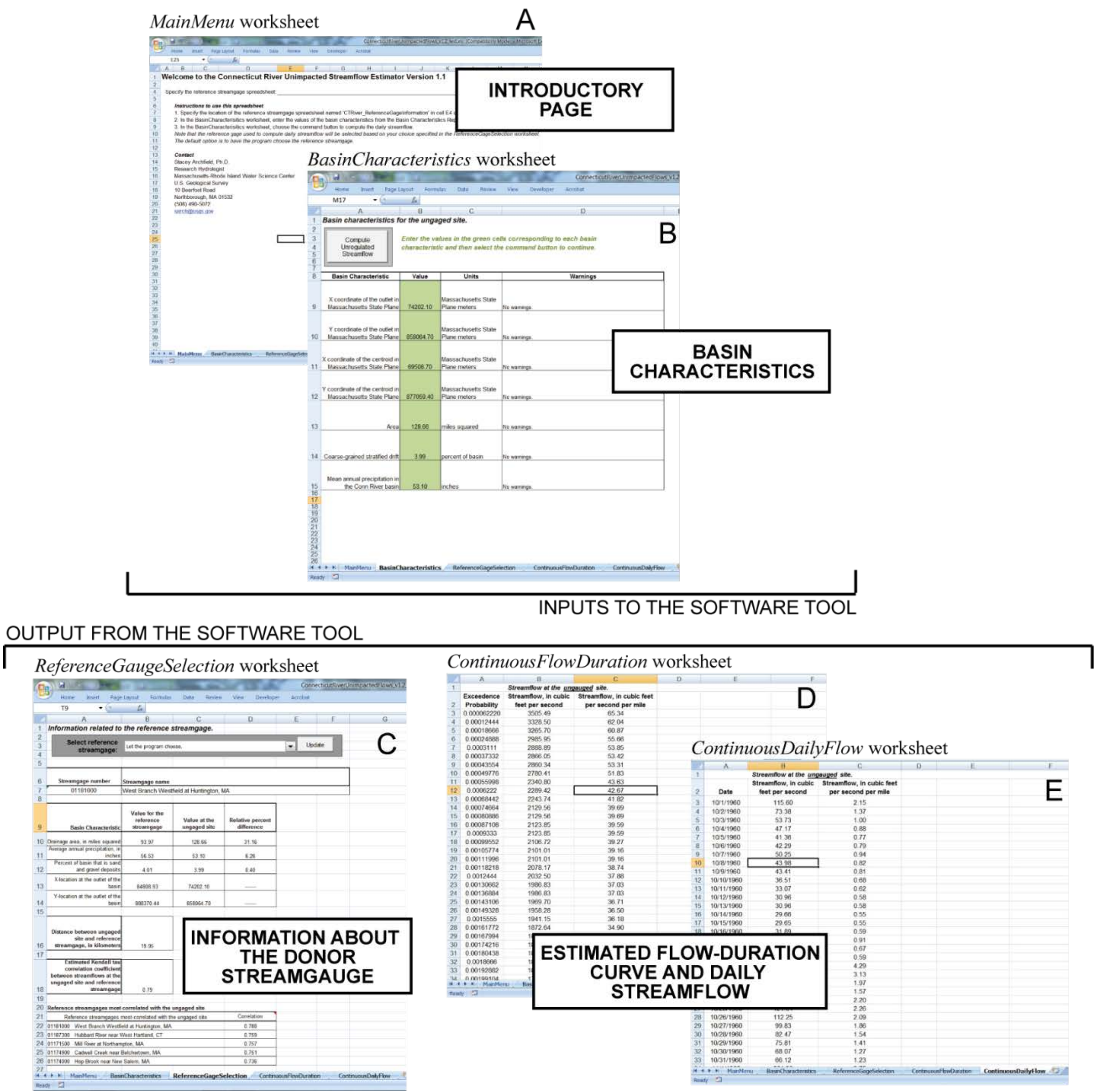

Figure 4. Screen captures showing the spreadsheet portion of the software tool used to estimate daily, unregulated time series. After reading the introductory page (A), the user inputs the catchment characteristics (or basin characteristics, as named in the tool) into the BasinCharacteristics worksheet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

		Characteristics in the regression equation and coefficient value			
Exceedence probability	Root mean square error	Constant term	Coefficient on explanatory variable	Explanatory variable	Bias correlation factor†
0.9	32.36	-0.4112	1.0511	Streamflow at the 0.85 exceedence probability	1.0004
0.95	57.15	-0.4991	1.0607	Streamflow at the 0.9 exceedence probability	0.9986
0.98	67.36	-0.4695	1.0567	Streamflow at the 0.95 exceedence probability	1.0103
0.99	102.33	-0.3011	1.0467	Streamflow at the 0.98 exceedence probability	1.0000
0.999938	825.08	-1.6658	1.2826	Streamflow at the 0.99 exceedence probaility	1.2011

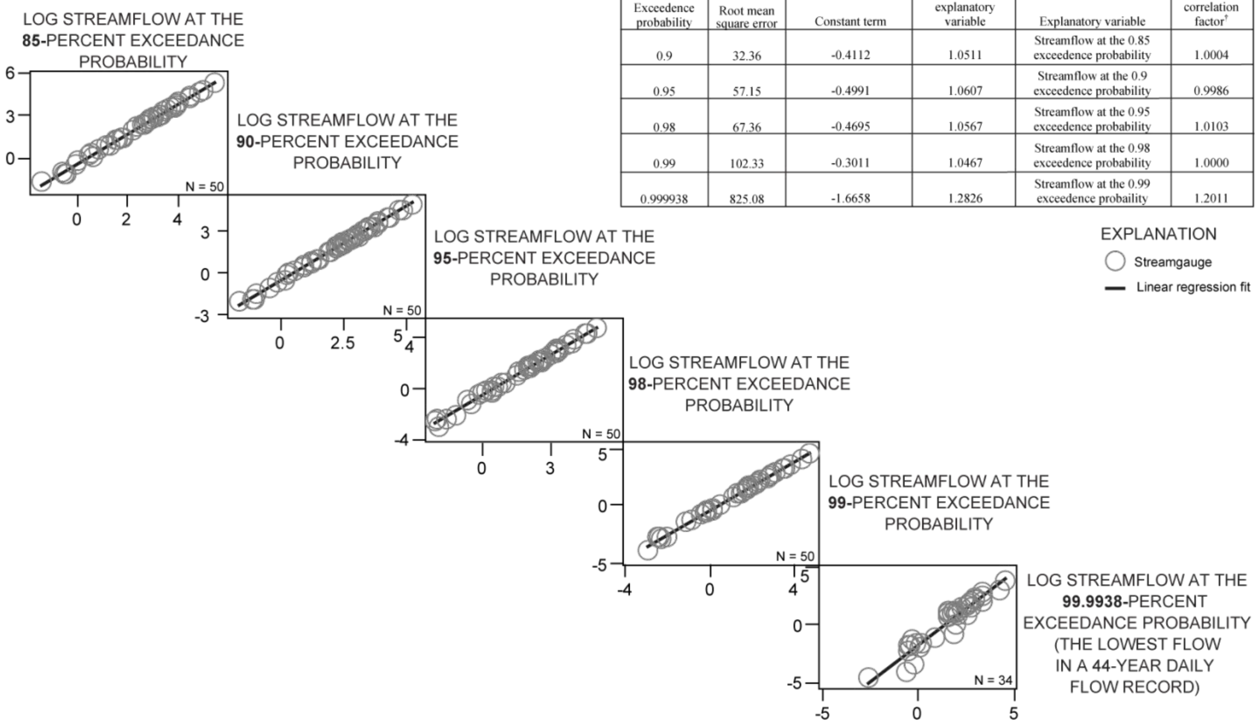

Figure 5. Relations between streamflows at the 0.9, 0.95, 0.98, 0.99 and 0.999938 exceedence probabilities and the corresponding goodness of fit values resulting from a least-squares linear regression to estimate streamflows recursively from other streamflow quantiles. (†, Bias correction factor computed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

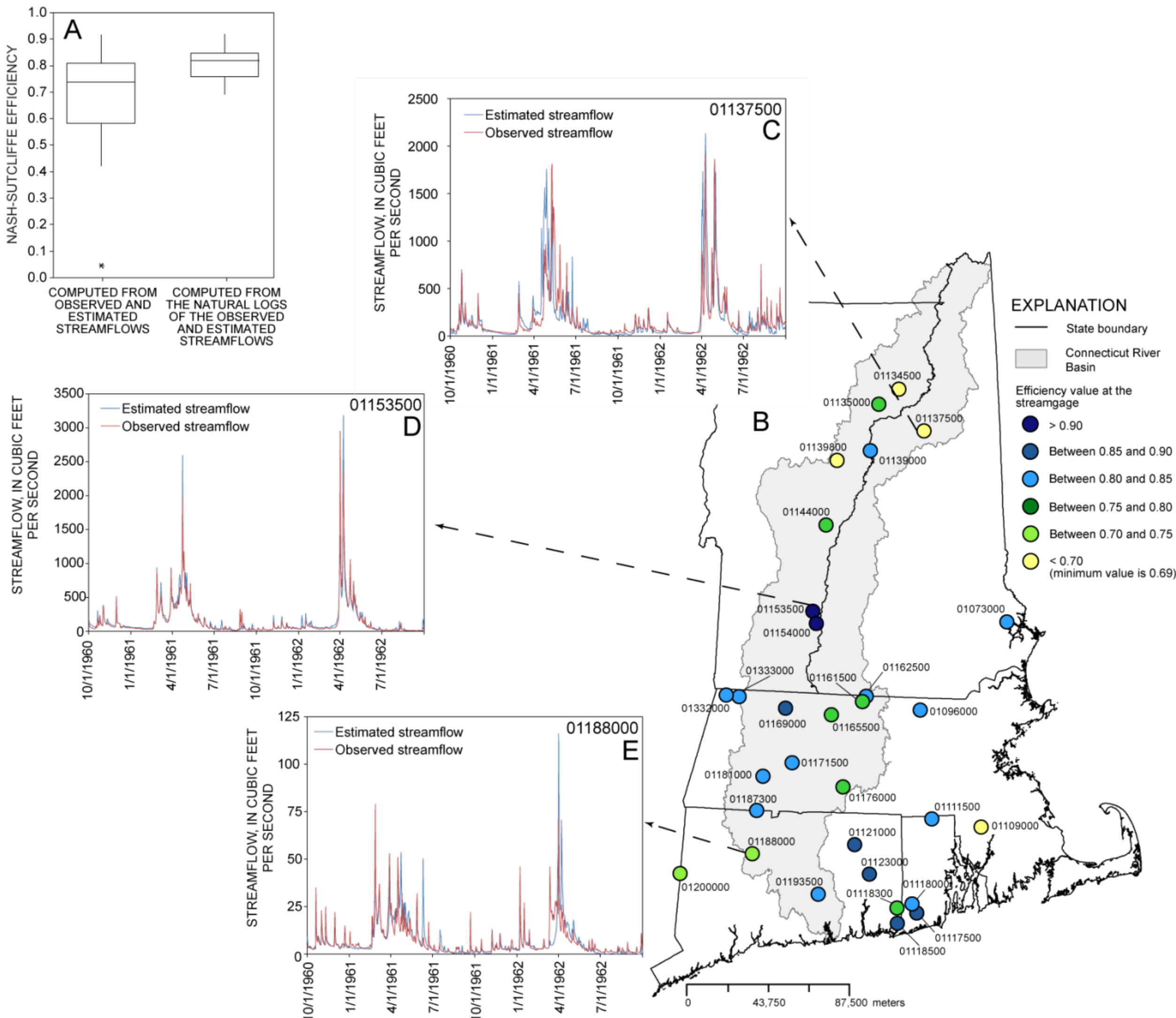

Figure 6. Range of efficiency values computed between the observed and estimated streamflows at the 31 validation streamgauges (A), spatial distribution of efficiency values resulting from log-transformed observed and estimated daily streamflow at 31 validation streamgauges (B) and selected hydrographs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

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 streamflow quantiles estimated from catchment characteristics using multiple least squares regression.

[†, Bias correction factor computed from *Duan* (1983)]

	General regression information		**Estimated regression coefficients**					
Exceedence probability	Number of stream-gauges used to develop regression equation	Percent root mean square error	Constant term	Drainage area	Average annual precip-itation.	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

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