A decision-support system to assess surface-water resources in Massachusetts

By Stacey A. Archfield[1] and Richard M. Vogel[2]

[1]Department of Civil and Environmental Engineering, Tufts University, Medford, MA, 01255. Phone: 508-490-5072, Email: stacey.archfield@tufts.edu

[2]Department of Civil and Environmental Engineering, Tufts University, Medford, MA, 02155. Phone: 617-627-4260, Email: richard.vogel@tufts.edu

ABSTRACT

Federal, State and local water supply, regulatory, and planning agencies require easy-to-use, technically-defensible, decision-support (DS) applications that can evaluate impacts of proposed water withdrawals, determine baseline streamflow conditions needed for sustainability of aquatic habitat, and estimate inflows to drinking-water-supply reservoirs for safe yield analyses at ungaged locations. An interactive, point-and-click DS application is developed in combination with a geographic-information system to address these needs. The DS application estimates unimpacted daily streamflow at any user-selected location – gaged or ungaged -- on a perennial stream in Massachusetts. A new method is proposed to estimate a daily flow-duration curve at an ungaged site by exploiting the strong structural relationship among streamflow quantiles. This method offers improvement -- particularly for low flows -- over traditional regression-based approaches that relate flows at selected flow quantiles to measurable basin characteristics. A time series of daily flows is then created by transferring the timing of the daily flows at an index gage to the ungaged site at equivalent exceedance probabilities. Estimated daily streamflows show remarkably good agreement with observed daily flows and are generally comparable to the agreement obtained from a calibrated rainfall-runoff model. A jack-knife cross-validation experiment indicates that the agreement between observed and estimated flow series at an ungaged site is also remarkably good.

INTRODUCTION

Whereas a variety of calibrated watershed models are available to stakeholders to assess water-availability and management, there is no analog for this type of assessment at ungaged sites where no models are available and/or where no data to calibrate such models are available. In Massachusetts, the legislation authorizes the Massachusetts Department of Environmental Protection (MassDEP) to assess permits of water withdrawals relative to the basin's sustainable yield, where sustainable yield is defined as the unimpacted streamflow at some location on a stream less some amount of water necessary to sustain the natural habitat. Estimates of unimpacted streamflow at any location on a stream – gaged or ungaged – is critical to the calculation of sustainable yield.

Previous work in Massachusetts to estimate streamflow time-series at ungaged sites has employed regional regression to relate watershed characteristics of an ungaged basin to selected flow-duration curve (FDC) statistics with the goal of estimating a daily, period-of-record FDC at the ungaged site. Ries and Friesz (2000) related physical characteristics of basins to selected exceedence probabilities associated with only low flows. Flows at the 50-, 60-, 70-, 75-, 80-, 85-, 90-, 95-, 98-, and 99-percent exceedence probabilities were regressed against basin characteristics such as drainage area and percent of sand and gravel deposits in the basin. Alternatively, Fennessey (1994) and Fennessey and Vogel (1990) assumed an underlying probability density function for daily streamflow and regressed the parameters of the assumed distribution against basin characteristics. Fennessey (1994) then developed a method to transform a FDC estimated for the ungaged site into a time series of streamflows though use of an index gage. Building on the work of both Ries and Friesz (2000) and Fennessey (1994), the Massachusetts Sustainable-Yield Estimator (SYE) is developed to provide estimates of continuous, daily streamflow at ungaged sites. The purpose of this paper is to describe the SYE and the methods used to estimate unimpacted streamflow at ungaged locations in Massachusetts.

MASSACHUSETTS SUSTAINABLE-YIELD ESTIMATOR DSS

The Massachusetts SYE is an interactive decision-support application which can estimate the unimpacted and impacted continuous, daily hydrograph from October 1, 1960 to September 30, 2004 at user-selected ungaged sites in Massachusetts. The Massachusetts SYE was designed as a desktop application that employs ArcGIS, a geographic information system (GIS) (fig. 1), and Microsoft Excel and Microsoft Access, commonly-used and widely-available spreadsheet and database programs, respectively (fig .2). Based on user-defined constraints such as existing water-use in the basin and instream-flow regimes necessary for sustainability of aquatic habitat, the SYE computes the sustainable yield of the basin, defined as the difference between the estimated hydrograph and the user-specified instream-flow regime. Users can quickly and easily compute the sustainable yield of the basin for a variety of water-management scenarios and instream-flow regimes. The calculation of sustainable yield is based not only on user-specified instream-flow targets (time-varying or constant flow targets) but also for a user-specified time period (drought year, wet year, or average year).

To estimate unimpacted flows, an FDC at the ungaged basin is estimated by relating watershed characteristics of the ungaged basin to properties of the flow-duration curve. The QPPQ-transform method developed by Fennessey (1994) is then used to construct a time series of flows from the estimated flow-duration curve at the ungaged site. The QPPQ-transform method transfers the timing of flow at a nearby index gage to the ungaged site. Although not the focus of this paper, ground- and surface-water withdrawals, ground-water discharges, and return flows are used to estimate the time series of impacted streamflow. This study only focuses on the component of the SYE DSS which estimates the time series of streamflow at an ungaged site, ignoring the components which deal with impacts of water withdrawals and instream flow requirements.

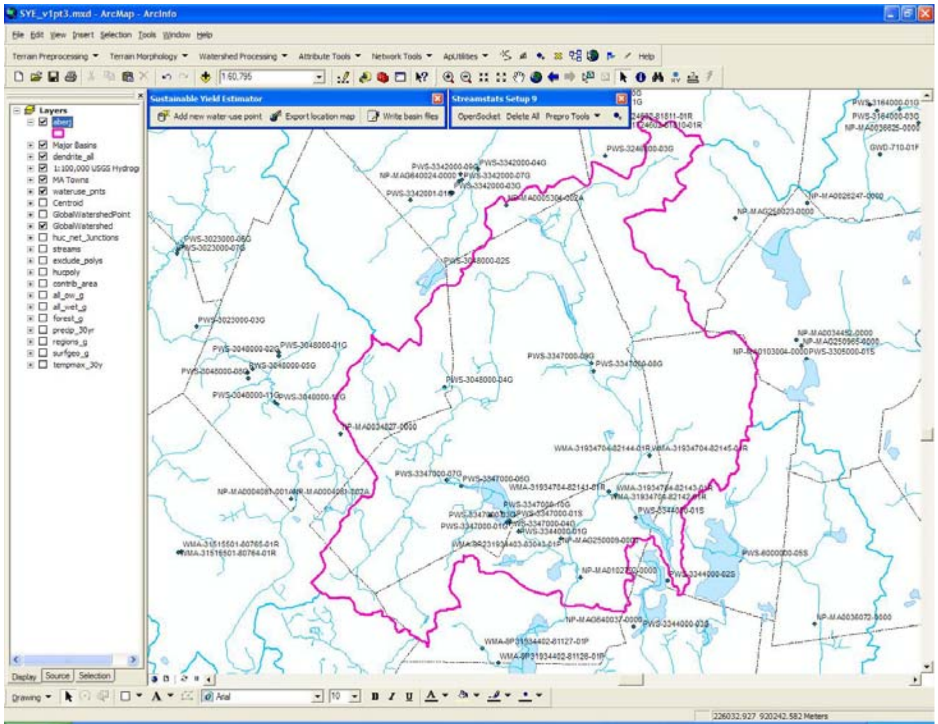

Figure 1. ArcGIS user-interface and related toolbars for the Massachusetts Sustainable-Yield Estimator. Delineated watershed shown above is located in approximately 15 kilometers north of Boston, Massachusetts.

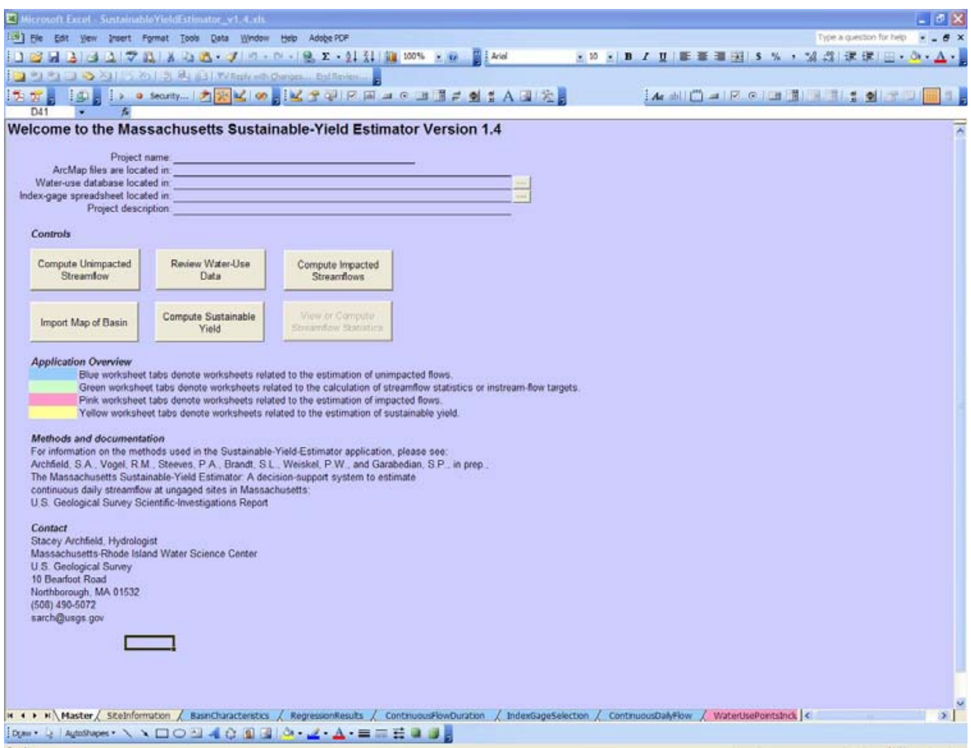

Figure 2. Microsoft Excel user-interface for the Massachusetts Sustainable-Yield Estimator.

Estimation of unimpacted flow-duration curves at ungaged sites

To estimate the unimpacted flow-duration curve (FDC) at an ungaged location, selected FDC quantiles were computed at 66 minimally-altered gages in southern New England (fig. 3) and were then regressed against measurable basin characteristics at these gages. Previously (Archfield and others, 2007), two methods to estimate the unimpacted FDC – resulting in two sets of regression equations - were tested and compared: (1) regression of moments of an assumed distribution of daily streamflow against watershed characteristics, and (2) regression of selected FDC quantiles against watershed characteristics and interpolation of quantiles between the regression-estimated points. Regression equations developed from the 66 gages were jack-knifed to compare differences in estimated streamflows across the methods and across all gages. Archfield and others (2007) showed that method (2) consistently outperformed method (1), particularly at low streamflows (streamflows with exceedance probabilities greater than 0.8); however, the errors in both methods increased with increasing exceedance probability.

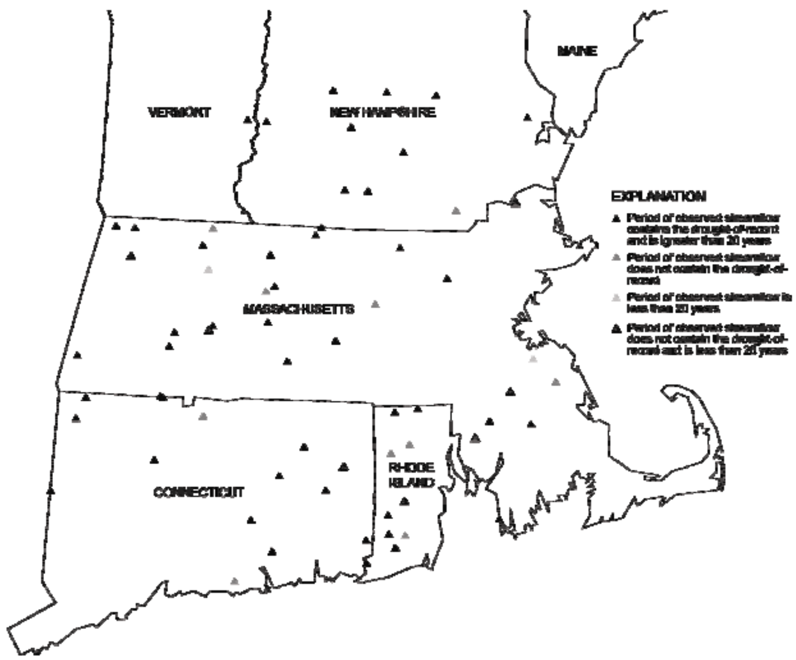

Figure 3. Locations of 66 minimally-impacted streamflow gages in southern New England.

Method (2) was selected for this study to estimate the unimpacted FDC at ungaged sites. Nineteen regression equations were developed for the following quantiles: 0.00062-, 0.01-, 0.05-, 0.10-, 0.15-, 0.20-, 0.30-, 0.40-, 0.50-, 0.60-, 0.70-, 0.80-, 0.85-, 0.9-, 0.95-, 0.99-, and 0.999938-exceedance probabilities. Exceedance probabilities of 0.00062 and 0.999938 are the exceedance probabilities for the largest and smallest streamflows calculated using the Weibull plotting position (Stedinger et al. 1993) for a record containing 16,071 streamflow observations (the number of daily streamflow observations from October 1, 1960 to September 30, 2004). Streamflow was log-linearly interpolated between the 19 regression-estimated streamflows.

Because regression equations were developed independently for each of the quantiles at the 19 exceedance probabilities, there was no constraint that ensured estimated streamflows decreased with increasing exceedance probability. When unimpacted FDCs were constructed from the jack-knifed regression equations, approximately half of the estimated FDCs did not have decreasing streamflow with increasing exceedance probability; this behavior was particularly pronounced at the highest and lowest streamflows, where the uncertainty in the estimated streamflows was greater.

It was observed that, when the natural log of the quantiles were plotted against one another, one observes a remarkable structure to the ordered quantiles (fig. 4) that is not necessarily preserved in the 19 independent regression equations. To enforce this structure among the quantiles, regression equations with explanatory basin characteristics were replaced with regression equations that use streamflows at other exceedance probabilities as explanatory variables at a subset of the 19 exceedance probabilities. Various combinations of streamflows regressed against basin characteristics and streamflows regressed against other streamflows were compared before the selecting the final set of explanatory variables, shown in table 1.

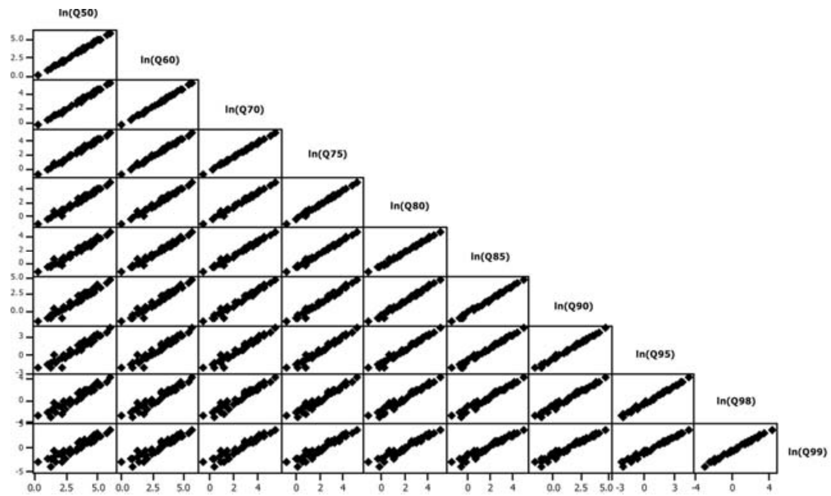

Figure 4. Relation between the natural log of streamflows at selected exceedance probabilities for 66-minimally-altered gages in southern New England. The label ln(QX) represents the natural log of streamflow at the X-divided-by-100 exceedance probability. Flows at selected exceedance probabilities less than 0.05 show the same structure.

Using the regression equations shown in table 1, the estimated FDC at all 66 gages resulted in estimated quantiles that decrease with increasing exceedance probability. The set of regression equations developed using the explanatory variables shown in table 1 and the set of regression equations developed using only basin characteristics were compared to ensure the new equations led to improvements in the estimated streamflows   Nash-Sutcliffe efficiency values (Nash and Sutcliffe, 1970) based on the 66 sites are compared in figure 5 using the two methods (1) independent regressions for each quantile in table 1 against basin characteristics and (2) regressions which relate quantiles to both other quantiles and basin characteristics. Figure 5 illustrates that using explanatory variables which exploit the structure of the ordered quantiles not only preserves the structure of the FDC but also improves upon our estimates of the FDC, particularly for low streamflows (large exceedance probabilities).

Table 1. Explanatory variables used to estimate streamflows at 19 exceedance probabilities.

Exceedance probability	Explanatory variable(s) for final regression equations
0.000062	Drainage area, Mean basin elevation
0.01	Drainage area, Average annual precipitation, Percent of basin that is wetlands, Y-location of gage
0.05	Drainage Area, Average annual precipitation, X-location of gage, Average maximum temperature
0.10	Estimated streamflow at the 0.15-exceedance probability
0.15	Estimated streamflow at the 0.20-exceedance probability
0.20	Drainage area, Average annual precipitation, Y-location of gage, X-location of gage
0.30	Estimated streamflow at the 0.40-exceedance probability
0.40	Drainage Area, Average annual precipitation, Percent of basin that is sand and gravel, Percent of basin that is open water, Y-location of gage, X-location of gage
0.50	Estimated streamflow at the 0.40-exceedance probability
0.60	Estimated streamflow at the 0.50-exceedance probability
0.70	Estimated streamflow at the 0.60-exceedance probability
0.80	Drainage area, Y-location of gage, Average maximum temperature, Percent of basin that is sand and gravel
0.85	Estimated streamflow at the 0.80-exceedance probability
0.90	Estimated streamflow at the 0.85-exceedance probability
0.95	Estimated streamflow at the 0.90-exceedance probability
0.99	Estimated streamflow at the 0.95-exceedance probability
0.999938	Estimated streamflow at the 0.99-exceedance probability

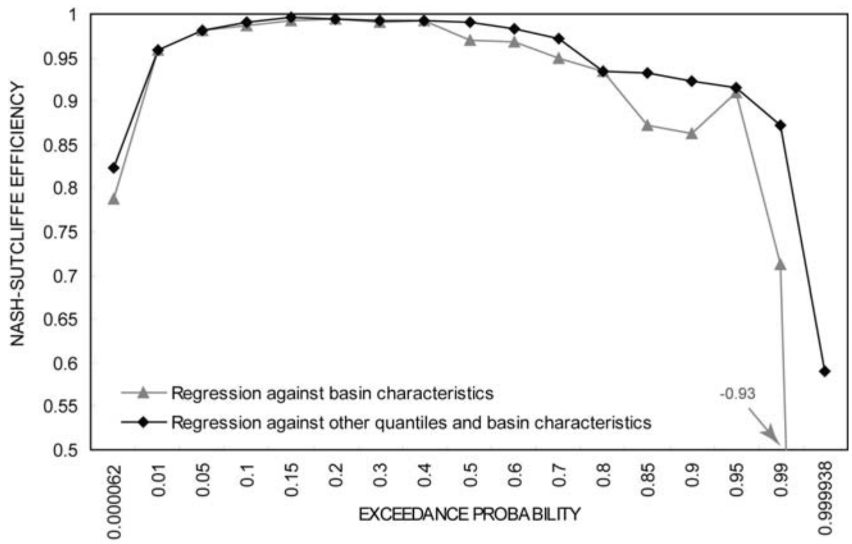

Figure 5. Comparison of methods to estimate selected quantiles of an FDC at 66 basins

Estimation of continuous, daily streamflow at ungaged sites

To convert an estimated FDC at an ungaged site to a time series of streamflow, an index gage is used to assign a date to each exceedance probability of the FDC at the ungaged site. The QPPQ-transform method developed by Fennessey (1994) relates the FDC at the ungaged site to an index gage by equating the FDC at the ungaged site with the FDC at the index gage. The timing of the flows at the index gage is transferred to the ungaged site through this relation between FDCs; however, no guidance was provided in Fennessey (1994) on how to select the index gage.

An experiment using 28 of the 66 study gages is conducted to determine how to select the index gage when using the QPPQ-transform method. A subset of 28 gages was chosen because they were the largest set of gages that contained the longest, continuous, and concurrent period of observed record of observed streamflow out of the 66 gages. For each of the 28 gages, a FDC was created at the gage from the observed streamflows. The remaining 27 gages were considered potential index gages and used to construct a time series of flows at the gage. The observed and QPPQ-estimated time series were compared using the Nash-Sutcliffe-efficiency value. Therefore, for each of the 28 gages, there were 27 Nash-Sutcliffe-efficiency values with each value representing the agreement between the estimated and observed time series resulting from the use of each of the 27 gages as an index gage.

Differences between basin characteristics, distances between the gages, and correlation between the gages – calculated from the natural log of the streamflows – were computed and compared to the Nash-Sutcliffe-efficiency values. It is important to note that the period-of-record was identical for these 28 gages and therefore was not a consideration in the calculation of the correlation between the gages. Figure 6 compares the Nash-Sutcliffe-efficiency values against distances, correlations, and the relative percent difference between selected basin characteristics between each of the 28 gages and the 27 potential index gages.

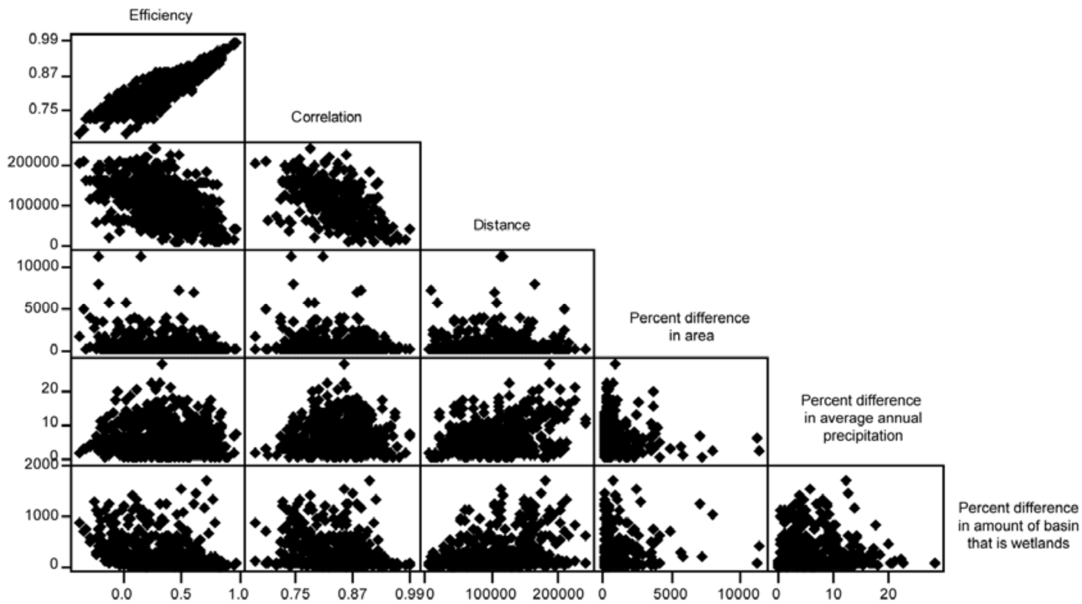

Figure 6. Scatter plots of Nash-Sutcliffe efficiency values, correlation, distances, and relative percent differences in basin characteristics between 28 gages in southern New England.

Inspection of figure 6 shows that correlation between the gages is the only variable tested that has a strong relation to Nash-Sutcliffe-efficiency values; however, correlation cannot be used to select an index gage because it cannot be calculated at an ungaged site. Other possible selection criteria (such as distance between the gages) did not show any strong relation to the Nash-Sutcliffe-efficiency values. If the closest index gage was chosen, the best or nearly best (within 0.15 of the highest efficiency value obtained) index gage would have been selected at only 17 of the 28 gages. Figure 6 also shows that the choice of an index gage can have a substantial impact on the errors in the estimated time series; therefore, finding selection criteria that correctly chooses the best or near-best index gage is critical to the success of the method.

Because no simple relation between the variables was evident in figure 6, two multivariate selection criteria were tested. A multivariate regression relation was developed from the variables shown in figure 6 (with the exception of correlation) to estimate Nash-Sutcliffe-efficiency values at ungaged sites; however, this relation selected the best or nearly best index gage at less than half the gages. A second multivariate selection criterion was to select the gage that minimizes the sum of the distance and relative percent differences between the basin characteristics. This multivariate selection criterion resulted in the best or nearly best gage being chosen at 22 of the 28 gages. Therefore, this selection criterion is used in the SYE to choose an index gage to be used to transform the estimated flow-duration at the ungaged site to a time series of streamflows.

Model validation

To evaluate the uncertainty in estimates of time series of streamflow at an ungaged site resulting from the regression equations and QPPQ-transform method the following jack-knife cross-validation experiment is performed. A subset of 19 sites having observed streamflow for the period-of-record of interest (October 1, 1960 to September 30, 2004) were compared to streamflows estimated from jack-knifed FDCs generated at the respective gages and time series estimated from the index gage selected from the selection criteria determined from above. In effect, this validation experiment evaluates the ability of the methodology to generate a time series of streamflow at a site which was NOT used in the development of the FDC regression. Comparisons of the FDCs and hydrographs for one year are shown in figure 7 for the: (A) best and (B) worst cases considered.

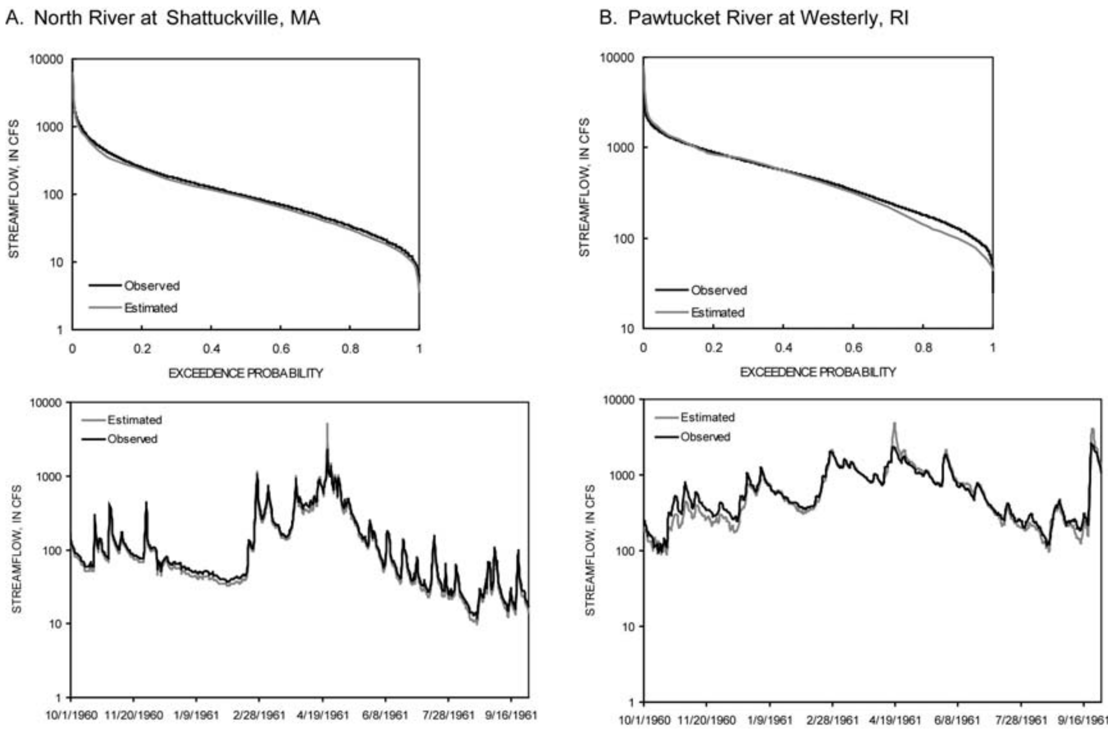

Figure 7. Best (panel A) and worst (panel B) agreements between estimated and observed flow-duration curves and hydrographs at 19 gages in southern New England.

The estimated streamflows – resulting only from the FDC regression against quantiles and basin characteristics and the selection of an index gage combined with the QPPQ transform method compare remarkably well to the observed streamflows and are comparable in many cases to the results obtained from a calibrated rainfall-runoff model.

LIMITATIONS

Although the plots comparing observed and estimated streamflows compare well, the methods have been tested only for southern New England and it is unclear of the structure of the ordered flows holds for other areas of the U.S. We expect that the methods introduced here are applicable to all regions of the U.S., however, Kroll et al. (2004) document that regional regressions for low flows are likely to be much more accurate in the northeastern U.S. than in any other region of the country. It is unclear how well the QPPQ-transform method will perform for other areas of the country or if the selection criteria can be improved upon through use of multiple index gages. Although weighted-least-squares regression was used to account for differences in record length at the study gages, the spatial correlation between sites is not considered and it is also likely that the set of regression equations have cross-correlated errors that may require the use of seemingly-unrelated regression methods.

CONCLUSIONS

An interactive, decision-support system was developed for Massachusetts to estimate unimpacted daily streamflow at any ungaged location on a perennial stream in Massachusetts. Regional regressions were developed to estimate a flow-duration curve at the ungaged site and an index gage is then chosen and used to transform the estimated flow-duration into a time series of streamflows. The selection of an index gage is a choice based the closest gage in terms of both distance and multivariate basin characteristics. A jack-knife cross-validation experiment showed that estimated streamflows are comparable to observed streamflows at 28 gages in southern New England. Future work will focus on improving the selection of the index gage and computing prediction intervals for estimated streamflows.

REFERENCES

Archfield, S.A., R.M. Vogel, and S.B. Brandt, 2007, Estimation of flow-duration curves at ungaged sites in southern New England, *Proceedings of the American Society of Civil Engineers World Water and Environmental Resources Congress*, Tampa, Florida, 2007.

Fennessey, N.M. (1994). *A hydro-climatological model of daily streamflow for the northeast United States*, Tufts University, Medford, MA, Ph.D. dissertation.

Fennessey, N. and Vogel, R.M. (1990). Regional Flow Duration Curves for Ungaged Sites in Massachusetts, *Journal of Water Resources Planning and Management*, ASCE, 116(4), 530-549.

Kroll, C.N., J. Luz, B. Allen and R.M. Vogel, Developing a Watershed Characteristic Database to Improve Low Streamflow Prediction, *Journal of Hydrologic Engineering*, Vol. 9, No. 2, 116-125, 2004.

Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles, *Journal of Hydrology*, 10 (3), 282–290.

Ries, K.G. and Friesz, P.J. (2000). Methods for Estimating Low-Flow Statistics for Massachusetts Streams, *US Geological Survey WRIR 00-4135*.

Stedinger, J.R., R.M. Vogel and E. Foufoula-Georgiou, Frequency Analysis of Extreme Events, Chapter 18, *Handbook of Hydrology*, McGraw-Hill Book Company, David R. Maidment, Editor-in-Chief, 1993.

ACKNOWLEDGEMENTS

The authors would like to thank the Massachusetts Department of Environmental Protection and the U.S. Geological Survey Cooperative Water Program for their funding and support of this work. The authors would also like to thank Peter Steeves, Philip Zarriello, Sara Brandt, Peter Weiskel, and Dave Armstrong of the U.S. Geological Survey's Massachusetts-Rhode Island Water Science Center and Stephen Garabedian of the U.S. Geological Survey's Conte Anadromous Fish Laboratory for their technical advice and project support.