APPLICABILITY OF CONCURRENT BASEFLOW MEASUREMENT FOR ESTIMATION OF LOW-FLOW STATISTICS IN NEW HAMPSHIRE AND VERMONT

By

S. Lawrence Dingman and Jennifer S. Putscher

Department of Earth Sciences
University of New Hampshire
Durham, NH 03824

December 1991

APPLICABILITY OF CONCURRENT BASEFLOW MEASUREMENT FOR ESTIMATION OF LOW-FLOW STATISTICS IN NEW HAMPSHIRE AND VERMONT

by

S. Lawrence Dingman and Jennifer S. Putscher

Earth Sciences Department

University of New Hampshire

Durham, NH 03824

ABSTRACT

This study evaluates aspects of the accuracy of estimates of the annual 7-day-minimum flow with a non-exceedence probability of 0.1 (7Q10) at ungaged stations in NH and VT based on regression of baseflow measurements at the ungaged station with concurrent flows at stations with long-term streamflow records. We identified 48 stations with suitable low-flow records. Among these, fewer than half the randomly selected concurrent baseflows had significant correlations and statistically identical slopes and intercepts for regressions between: (1) logarithms of annual 7-day minimum flows (7Q); and (2) logarithms of randomly selected concurrent independent baseflows. Correlations exceeded 0.7 for 250 of these pairs, and 10 stream pairs, chosen for their areal and size distribution, were intensively studied. A method proposed by Stedinger and Thomas was found to give the lowest root-mean-square and bias among the three methods we compared and was used in the remainder of the study. One station of each of the pairs was considered to be "ungaged", and we evaluated

prediction accuracy by comparing 7Q10 estimated for the "ungaged" stations by regression with concurrent flows at a gaged station with the actual 7Q10 at the "ungaged" stations. Using a bootstrap procedure with 50,000 iterations, the bias of the 7Q10 estimated from concurrent flows was determined as a function of the number, M, of concurrent flows used; this bias decreased with M as expected but leveled off at M =12. The absolute value of the bias with the concurrent-baseflow method was less than or essentially equal to that obtained with regression on basin characteristics in 8 of the 10 streams tested as long as M > 12. Bias was not a function of inter-basin distance or of inter-basin differences in area, elevation, or geology. Thus the concurrent-baseflow method has potential as a useful estimation method in the region, but it should be tested on a larger sample of streams.

INTRODUCTION

The general problem addressed here is that of estimating low-flow statistics for stream reaches lacking long-term streamflow measurements. There are two general approaches to this problem: (1) establish regression relations between low-flow statistics and readily measurable drainage-basin characteristics ("regression on basin characteristics"); and (2) establish regression relations between baseflows at the ungaged reach and concurrent baseflows at stations with long-term streamflow records ("concurrent-baseflow method"). A previous study [1] developed and evaluated an approach to regression on basin characteristics for New Hampshire and Vermont. The goal of the present study is to assess the accuracy and applicability of the concurrent-baseflow method for the same region.

The flow statistic selected for study herein is the annual minimum streamflow averaged over seven consecutive days which has a non-exceedence probability of 0.10 in any year, called "7Q10". This statistic is commonly used as a design flow in sizing waste-water treatment plants and for other purposes.

DATA

Lawlor [1] identified 49 gaging stations in NH and VT that had records suitable for obtaining reliable estimates of natural (unregulated) 7Q10. We eliminated one of these because its annual seven-day minimum flow was zero for one or more years. The locations and distributions of record length and drainage area for the stations included in the initial sample are shown in Figures 1 - 3.

Statistical tests by Lawlor [1] supported the hypotheses that annual seven-day minimum flows at a given station are log-normally distributed and are not autocorrelated. Because of the log-normality, all correlations and regressions in this study are done using the logarithms of the reported flow values.

For most streams in the region, the low-flow season is from July through October, and the annual seven-day minimum flows for most streams occur either in August or September. However, at high elevations, the lowest flows occur during the winter (usually February). Measurement of low flows when ice is present is difficult, and most flow data for these periods are estimated values. Because of the low precision of these data, only summer baseflow periods were considered in the sampling.

NOTATION AND UNITS

We use the following notation, with "u" or "g" added as required to distinguish flows at ungaged and gaged stations:

7Q	annual 7-day minimum flow ($ft^3\ s^{-1}$)
7Q10	7-day minimum flow with annual non-exceedence probability of 0.01 ($ft^3\ s^{-1}$)
a	intercept computed via regression
b	slope computed via regression
e	residual error in linear regression
E( )	population mean
$\hat{E}$( )	sample mean
K_{10}	frequency factor for flow with non-exceedence probability of 0.10
L7Q	base-10 logarithm of 7Q
LQ	base-10 logarithm of Q
L7Q10	base-10 logarithm of 7Q10
M	number of concurrent-flow pairs regressed
N	number of year of record
P	significance level
Q	average daily or instantaneous baseflow ($ft^3\ s^{-1}$)

$S(\ )$	population standard deviation
$\hat{S}(\ )$	sample standard deviation
α	intercept in "true" linear relation
β	slope in "true" linear relation

STATISTICAL ASPECTS OF CONCURRENT-BASEFLOW METHOD

Basic Relations

The concurrent baseflow method assumes that the logarithms of the annual 7-day minimum flows at the ungaged site, $L7Q_i$, are related to those at the gaged site, $L7Qg_i$, as

$$L7Qu_i = \alpha + \beta * L7Qg_i + e_i, \tag{1}$$

where α (intercept) and β (slope) are regression parameters and the e_i are independent residual errors that are assumed to be normally distributed with mean 0 and uncorrelated with the $L7Qg_i$. In principle, α and β can be estimated by regression analysis of a sample of L7Qu and concurrent L7Qg values.

However, since we have no record on which to determine values of 7Qu, concurrent measurements of baseflows during independent streamflow recession periods are used as proxies for the variables in Eqn. (1):

$$LQu_i = \alpha + \beta * LQg_i + e_i, \tag{2}$$

where α, β, and $S^2(e)$ have the same values as in Eqn. (1).

In principle, the values of L7Q10u and L7Q10g are given by

$$L7Q10u = E(L7Qu) + K_{10}*S(L7Qu) \quad (3)$$

and

$$L7Q10g = E(L7Qg) + K_{10}*S(L7Qg), \quad (4)$$

where K_{10} is a frequency factor giving the value with a 0.10 non-exceedence probability appropriate for the distribution of L7Qu and L7Qg (assumed identical).

The present study has examined three approaches that make use of Equations (1) - (4) to estimate L7Q10u from concurrent measurements of L7Qu and L7Qg; these approaches are described below.

Ordinary Least-Squares Regression (Riggs Approach)

Riggs [2] suggested estimating L7Q10u using the ordinary least-squares estimators of b and a in Eqn. (2). These are found as

$$b = \frac{\sum_{i=1}^{M}\{[LQu_i - \hat{E}(LQu)]*[LQg_i - \hat{E}(LQg)]}{(M - 1)*\hat{S}^2(LQg)} \quad (5)$$

and

$$a = \hat{E}(LQu) - b*\hat{E}(LQg). \quad (6)$$

Then L7Q10u is estimated as

$$L7Q10u_R = a + b*L7Q10g. \qquad (7)$$

Stedinger and Thomas [3] showed that $L7Q10u_R$ is a biased estimator of L7Q10u.

Stedinger and Thomas Moment-Estimation Approach

Stedinger and Thomas [3] suggested a "reasonable, consistent, and simple" moment estimator of L7Q10u to avoid the bias problem associated with the OLS estimate. First, values of a and b are calculated via Equations (3) and (4) and used along with $\hat{E}(L7Qg)$, $\hat{S}(L7Qg)$, and $\hat{S}(e)$ to estimate E(L7Qu) and S(L7Qu) as

$$\hat{E}(L7Qu) = a + b*\hat{E}(L7Qg) \qquad (8)$$

and

$$\hat{S}^2(L7Qu) = b^2*\hat{S}^2(L7Qg) + \hat{S}^2(e)*\left[1 - \frac{\hat{S}^2(L7Qg)}{(M - 1)*\hat{S}^2(LQg)}\right]. \qquad (9)$$

The Stedinger-Thomas (ST) estimator of L7Q10 is then

$$L7Q10u_{ST} = \hat{E}(L7Qu) + K_{10}*\hat{S}(L7Qu). \qquad (10)$$

Maintenance of Variance Extension Method

Stedinger and Thomas [3] also examined the "maintenance-of-variance-extension" ("MOVE.1") method, which estimates L7Q10u as

$$L7Q10u_M = \hat{E}(LQu) + K_{10}*\hat{S}(LQu). \tag{11}$$

SPECIFIC QUESTIONS ADDRESSED AND METHODOLOGY

Specific Questions

The overall objective of the present study is to determine the applicability of the concurrent-baseflow method for estimating 7Q10 in New Hampshire and Vermont. To do this, we address the following questions:

1. Under what conditions are LQu and LQg sufficiently correlated to give acceptable precision in predictions of L7Q10u?

2. Under what conditions can one assume that α and β in Equations (1) and (2) are identical?

3. Which of the three estimates $L7Q10u_R$, $L7Q10u_{ST}$, and $L7Q10u_M$ provides the best predictions of L7Q10u?

4. How many pairs of concurrent baseflow measurements are required to obtain acceptable prediction precision and

accuracy?

Methodology

To answer the questions posed above, this study examines relations between baseflows for pairs of gaged stations, with one member of the pair consider as "ungaged". Because statistical inferences are valid only when the items in the sample are independent, we developed the following protocol for selecting baseflows for each station pair: For the period July - October, recessions (i.e., periods of continuously decreasing flow) of at least seven days duration at the "ungaged" stream were identified. Then the first three days of each recession were eliminated, and one flow was randomly selected from the remaining days. The flow for the same day at the "gaged" member of the pair was the concurrent flow. Each stream was used once as the "ungaged" stream.

Assessment of Prediction Accuracy and Precision

For comparison of various approaches to estimating L7Q10, we measure prediction accuracy as the bias (BIAS), which is calculated as

$$BIASx = \frac{1}{M} * \sum_{i=1}^{M} (L7Q10u_x - L7Q10u) \qquad (12)$$

where x represents the three methods (R, ST, M).

We also compare estimation methods using the root-mean-square error of prediction (RMSEx), defined as

$$\mathrm{RMSEx} = \left[\frac{\sum_{i=1}^{M} (\mathrm{L7Q10u}_x - \mathrm{L7Q10u})^2}{M} \right]^{1/2} . \qquad (13)$$

Equations (12) and (13) were employed in a bootstrap-sampling mode, which involves repeated sampling with replacement to evaluate BIASx and RMSEx for very large sample sizes. Here, we used bootstrap sampling with M = 50,000 samples for each pair of streams examined.

RESULTS

Number of Independent Recessions per Baseflow Season

There were an average of five recessions each year with durations of at least seven days.

Occurrence of Significantly Correlated Baseflow Pairs

For the 48 streams in the original sample, there were 2026 pairs of concurrent flow records (considering each stream as an "ungaged" stream). Table 1 summarizes the correlations among L7Q and LQ values. Note that only 42.2 % of the pairs had both: (1) significant correlations for LQ-LQ and L7Q-L7Q; and (2) statistically identical values of slopes and intercepts for the two relations. Out of the 855 pairs passing these two tests, only 250 (12.3 %) had correlation coefficients exceeding 0.7.

We attempted to see if prediction was enhanced by geographical proximity or similarity in the three watershed characteristics known to have a significant effect on L7Q10 in the region: drainage area, mean basin elevation, and percentage of basin underlain by stratified glacial drift [1]. In all cases, the effects of these factors was assessed only for stream pairs that had: (1) LQ and L7Q correlations that were significant ($p = 0.05$) and positively correlated; and (2) statistically identical values of slope and intercept for Equations (1) and (2). In all cases the L7Q10u values were predicted using 12 pairs of concurrent baseflows.

Figures 4 - 7 show BIAS as a function of distance between basin centroids and differences in the three basin characteristics. The plots show a slight tendency for BIAS to increase with inter-basin distance and differences in elevation and drift fraction, but none of the trends is statistically significant ($p = 0.05$).

Assumption of Identical Parameters for L7Q and LQ Regressions

Of the 2026 flow pairs examined, 1387 (68 %) had statistically identical values of a and b (i.e., 95 % confidence intervals for α and β in the two regressions overlapped.

Choice of Method for Estimating L7Q10u

The three methods for estimating L7Q10u described earlier, $L7Q10u_R$ (Equation (7)), $L7Q10u_{ST}$ (Equation (10)), and $L7Q10u_M$ (Equation (11)) were compared with respect to bias (BIAS) and root-mean-square error (RMSE) of predicted values. Table 2 shows the results: The Stedinger and Thomas [3] method gave the lowest BIAS and RMSE. Thus $L7Q10u_{ST}$ was used as the estimator for the remainder of the study.

Number of Concurrent Baseflow Measurements Required

The effect of the number of baseflow pairs measured on he bias as calculated via Eqn. (10) was assessed using the bootstrap-sampling method with M = 50,000. For each pair of streams tested, L7Q10u was estimated as $L7Q10u_{ST}$ and BIAS was calculated based on 4, 8, 12, 20, and 30 pairs of concurrent-baseflow measurements. Only values of BIAS calculated for significant positive LQ correlations were included in the assessment.

Figures 8 - 17 show BIAS as a function of number of concurrent-baseflow measurements used for the 10 stream examined. Each graph is for the situation when the named stream is the "ungaged" stream, and each line on the graphs represents the results when L7Q10u is predicted for the named stream from another stream with which its LQ values are significantly correlated. Also shown on each graph is a horizontal line representing the BIAS in the prediction of L7Q10u obtained using basin characteristics in the method of Lawlor [1].

Figures 8 - 17 show that very little improvement in accuracy is obtained when more than 12 concurrent flows are used in the prediction.

The BIAS using the concurrent-baseflow method is clearly better than that obtained with the basin-characteristic approach for six of the 10 streams (Black River, Diamond River, Oyster River, Pemigewasset River, South Branch Piscataquog River, Saxtons River). The basin-characteristic approach seems better for two of the streams (East Orange Branch, Wild River), and the remaining two give no clear-cut indication of the better method.

Interestingly, when the concurrent-baseflow method is better, it is better even when only four measurements are used in the prediction.

Conclusions

This study suggests that the concurrent-baseflow method may yield more accurate predictions of 7Q10 than the basin-characteristics approach of Lawlor [1] under certain conditions. A significant limitation to its practical use, though, is that correlations between concurrent baseflows will be significant for only two-thirds of the station pairs. There is no indication that minimizing inter-basin distance or maximizing basin similarity with respect to area, elevation, or geology will markedly improve the chances of selecting a gaged stream for which flows will be significantly correlated with those at the ungaged site.

Since the biases of the concurrent-baseflow method seem of comparable magnitude to those obtained with the basin-characteristic method, it is likely that the most precise predictions of L7Q10 can be obtained by computing a precision-weighted average of the two methods, as suggested by Tasker [4].

Based on the results reported here, the optimum strategy for estimating L7Q10u would appear to be:

1. Measure baseflows (LQu) for 12 independent recessions (two to three baseflow seasons) at the ungaged site.

2. Test the correlation of LQu values with LQ values at a number of unregulated gaging stations.

3. Use all the significantly correlated LQu-LQg relations to estimate L7Q10u using the Stedinger and Thomas method ($L7Q10u_{ST}$).

4. Calculate an average L7Q10u value from the estimates of Step 3.

We are conducting further studies to confirm these results and conclusions, and to provide an assessment of the precision weighted estimate as proposed by Tasker [4] for New Hampshire and Vermont.

REFERENCES CITED

1. Lawlor, S.C. 1990. Estimation of low flows from basin characteristics in New Hampshire and Vermont. Durham, NH:

University of New Hampshire M.Sc. Thesis.

2. Riggs, H.C. 1965. Estimating probability distributions of drought flows. Water and Sewage Works 112(5):153-157.

3. Stedinger, J.R., and Thomas, W.O., Jr. 1985. Reston, VA: U.S. Geological Survey Open-File Report 85-95.

4. Tasker, G.D. 1975. Combining estimates of low-flow characteristics of streams in Massachusetts and Rhode Island. Journal of Research of the U.S. Geological Survey: 3(1) 107-112.

Table 1. Summary of Significant ($p = 0.05$) Correlations Among Gaging-Station Flows

Flows Correlated	Significant Correlations Number	Percent
LQ - LQ	1934	95.5
L7Q - L7Q	1129	55.7
LQ-LQ and L7Q-L7Q*	855	42.2

----

* Significantly correlated and slopes and intercepts for the two regressions not significantly different

Table 2.

Summary of Bias and Root-Mean- Square for LQ-LQ and L7Q-L7Q* Stream Pairs

Summary for "Y" stream	OLS BIAS	ST BIAS	MOV BIAS	OLS RMSE	ST RMSE	MOV RMSE
AMM	-0.0780	0.0040	-0.0090	0.0780	0.0120	0.0130
AYS	-0.7683	-0.0467	-0.0926	0.7683	0.1861	0.2424
BER	8.0225	2.5291	2.9983	8.0225	2.6545	3.2319
BEV	0.4316	0.1577	0.1226	0.4316	0.1577	0.1530
BKL	-0.0381	0.0320	0.0328	0.0428	0.0320	0.0334
BKW	-0.1529	-0.0097	0.0050	0.1529	0.0411	0.0496
BLK	-0.0806	0.0468	0.0285	0.0806	0.0468	0.0358
CDB	0.3008	-0.0361	0.0032	0.3008	0.0586	0.0886
CDR	-0.3061	-0.0077	-0.0008	0.3061	0.0443	0.0762
DIA	-0.2037	-0.0421	-0.0667	0.2037	0.0573	0.0803
DOG	-0.0952	0.0841	0.1050	0.1029	0.0882	0.1130
EBP	-0.1355	-0.0498	-0.0494	0.1355	0.0498	0.0540
EOB	0.6103	-0.1514	-0.0974	0.6131	0.2607	0.3304
FLD	0.7584	0.2309	0.2583	0.7584	0.2357	0.2922
KBY	0.4330	0.2080	0.2185	0.4330	0.2080	0.2185
KNT	0.5184	0.1151	0.0576	0.5184	0.1255	0.1386
LCY	0.6145	-0.0381	-0.0331	0.6145	0.0727	0.1234
LWS	0.2338	-0.0768	-0.0493	0.2338	0.0768	0.0771
MAD	-0.0896	0.0345	0.0168	0.0923	0.0402	0.0393
MAS	-0.3260	-0.0637	-0.0783	0.3260	0.0786	0.1098
MBK	-0.1281	-0.0080	-0.0109	0.1281	0.0180	0.0247
MNT	-0.0682	0.0100	0.0025	0.0697	0.0205	0.0337
MOS	-0.2146	-0.0160	-0.0146	0.2163	0.0327	0.0514
MSV	-0.3382	-0.0374	-0.1115	0.3382	0.0452	0.1115
OTT	-0.7519	0.5079	0.3501	0.7834	0.5475	0.4501
OYS	0.4789	-0.2140	-0.2103	0.4879	0.2193	0.2509
PEM	-0.0657	-0.0039	-0.0017	0.0666	0.0125	0.0188
PMW	-0.0738	0.0008	-0.0158	0.0738	0.0073	0.0183
SAC	4.5337	1.7265	1.7396	4.5337	1.7265	1.8082
SAX	-0.2948	0.0118	0.0483	0.2948	0.0634	0.0903
SBP	-0.2682	-0.0016	0.0086	0.2682	0.0504	0.0789
SCK	-0.2820	0.0481	0.0307	0.2884	0.0625	0.0881
SCO	-0.0624	0.0042	-0.0020	0.0624	0.0127	0.0187
SHG	-0.0435	0.1050	0.1008	0.0699	0.1050	0.1008
SMI	-0.1800	0.0040	0.0062	0.1800	0.0543	0.0767
STB	1.0013	0.2804	0.0836	1.0013	0.2804	0.2047
STN	1.0385	0.0069	0.0038	1.0385	0.2204	0.2954
SUN	0.2619	0.0984	0.0888	0.2619	0.1150	0.1130
WBW	0.2229	-0.0007	-0.0109	0.2229	0.0378	0.0583
WHB	-0.0675	0.0224	0.0230	0.0690	0.0236	0.0334
WHI	-0.0371	0.0388	0.0291	0.0412	0.0396	0.0331
WIL	-0.2557	-0.0629	-0.0512	0.2557	0.0765	0.0781
WLD	-0.0652	0.1625	0.1469	0.0816	0.1625	0.1469
WLM	-0.0639	0.0136	0.0109	0.0639	0.0152	0.0255
WRN	-0.3529	-0.0335	-0.0291	0.3553	0.0613	0.0965
Average for all eligible stream pairs	0.3480	0.1449	0.1541	0.6430	0.2128	0.2494

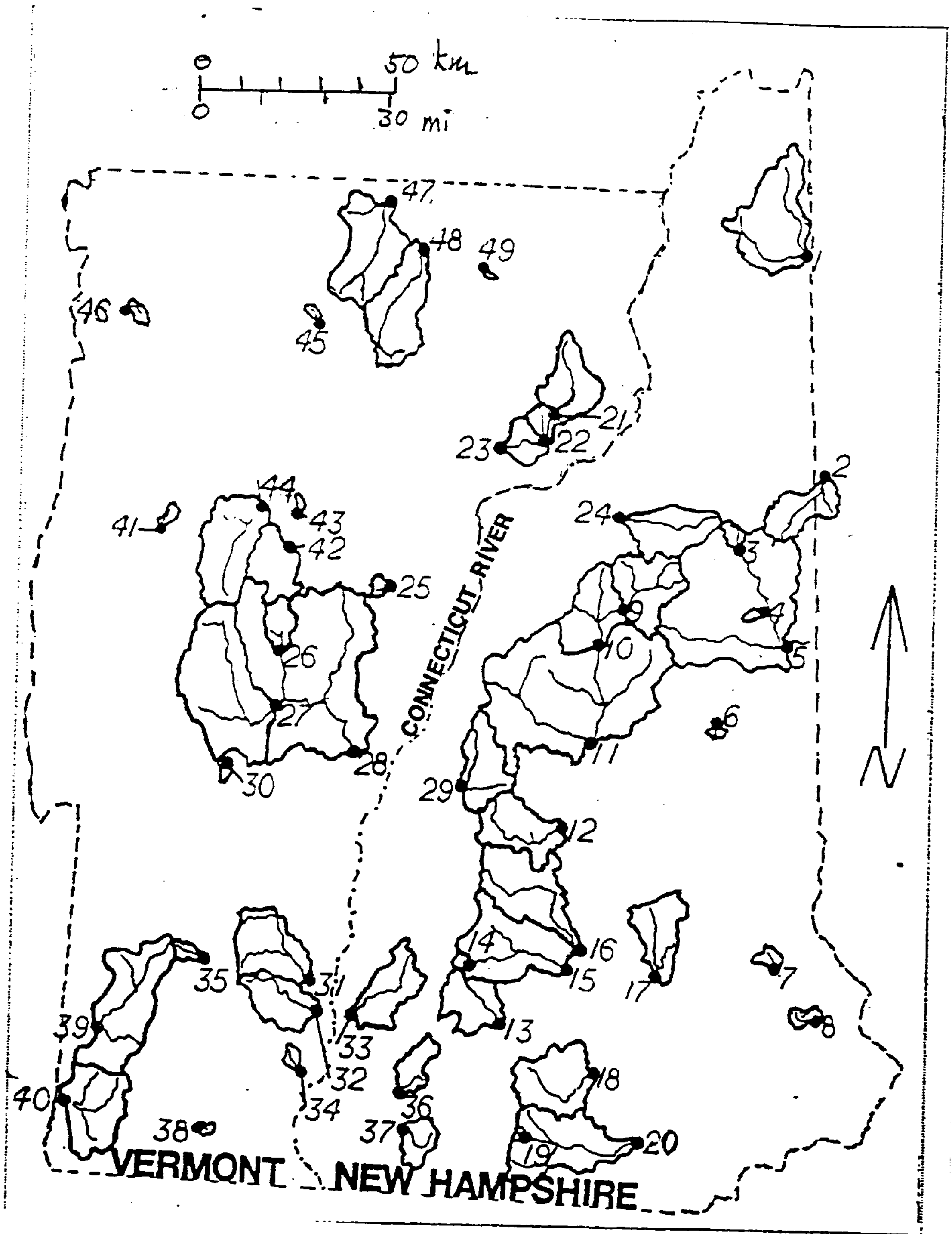
FIGURES

14. Bias in $L7Q10u_{ST}$ as a function of number of flow pairs for bootstrap experiment with Saxtons River as "ungaged" stream using concurrent flows at 25 gaged sites. Heavy line is bias of basin-characteristics estimate of L7Q10u.

15. Bias in $L7Q10u_{ST}$ as a function of number of flow pairs for bootstrap experiment with South Branch Piscataquog River as "ungaged" stream using concurrent flows at 15 gaged sites. Heavy line is bias of basin-characteristics estimate of L7Q10u.

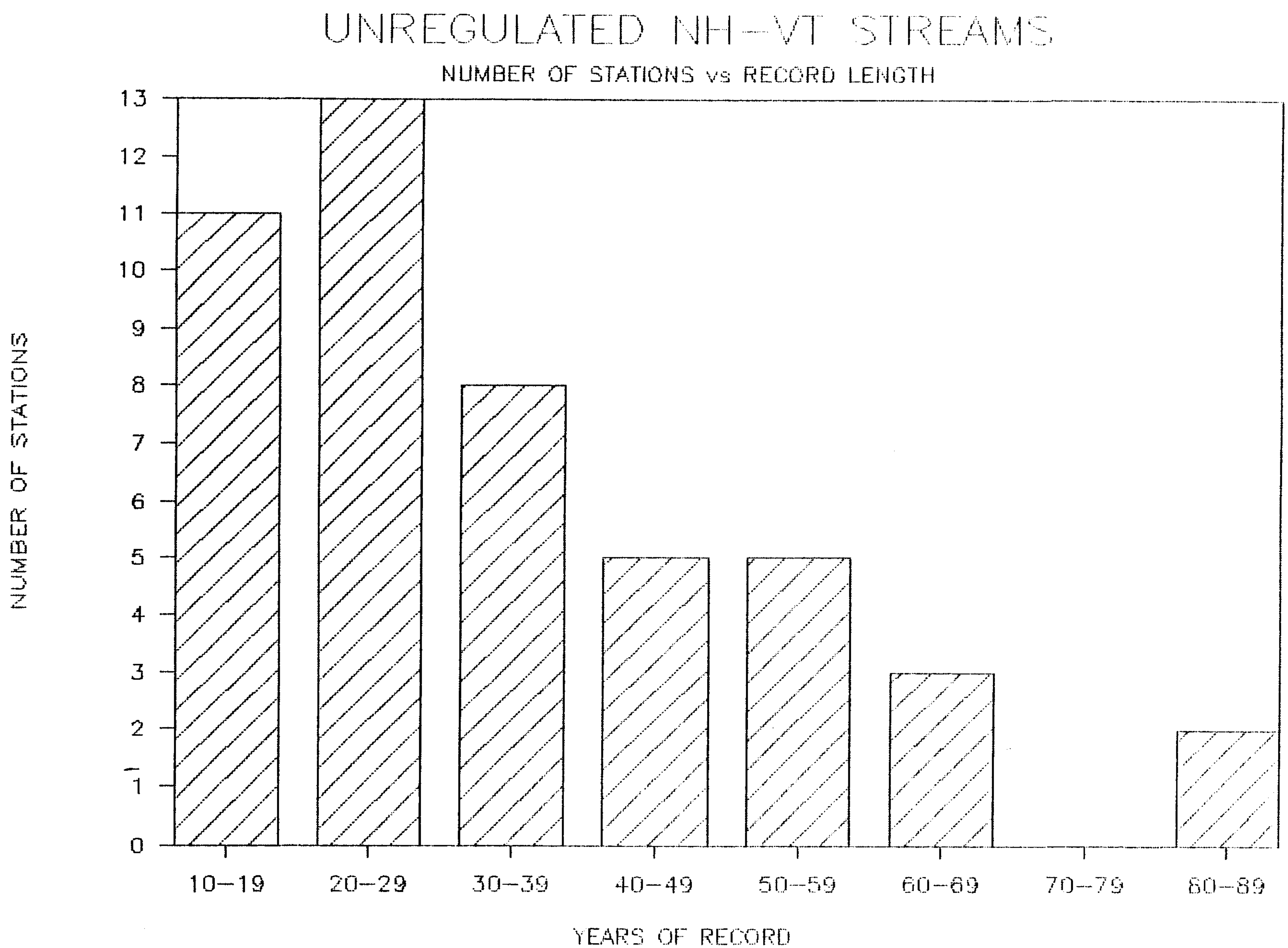
16. Bias in $L7Q10u_{ST}$ as a function of number of flow pairs for bootstrap experiment with White River as "ungaged" stream using concurrent flows at 29 gaged sites. Heavy line is bias of basin-characteristics estimate of L7Q10u.

17. Bias in $L7Q10u_{ST}$ as a function of number of flow pairs for bootstrap experiment with Wild River as "ungaged" stream using concurrent flows at 17 gaged sites. Heavy line is bias of basin-characteristics estimate of L7Q10u.



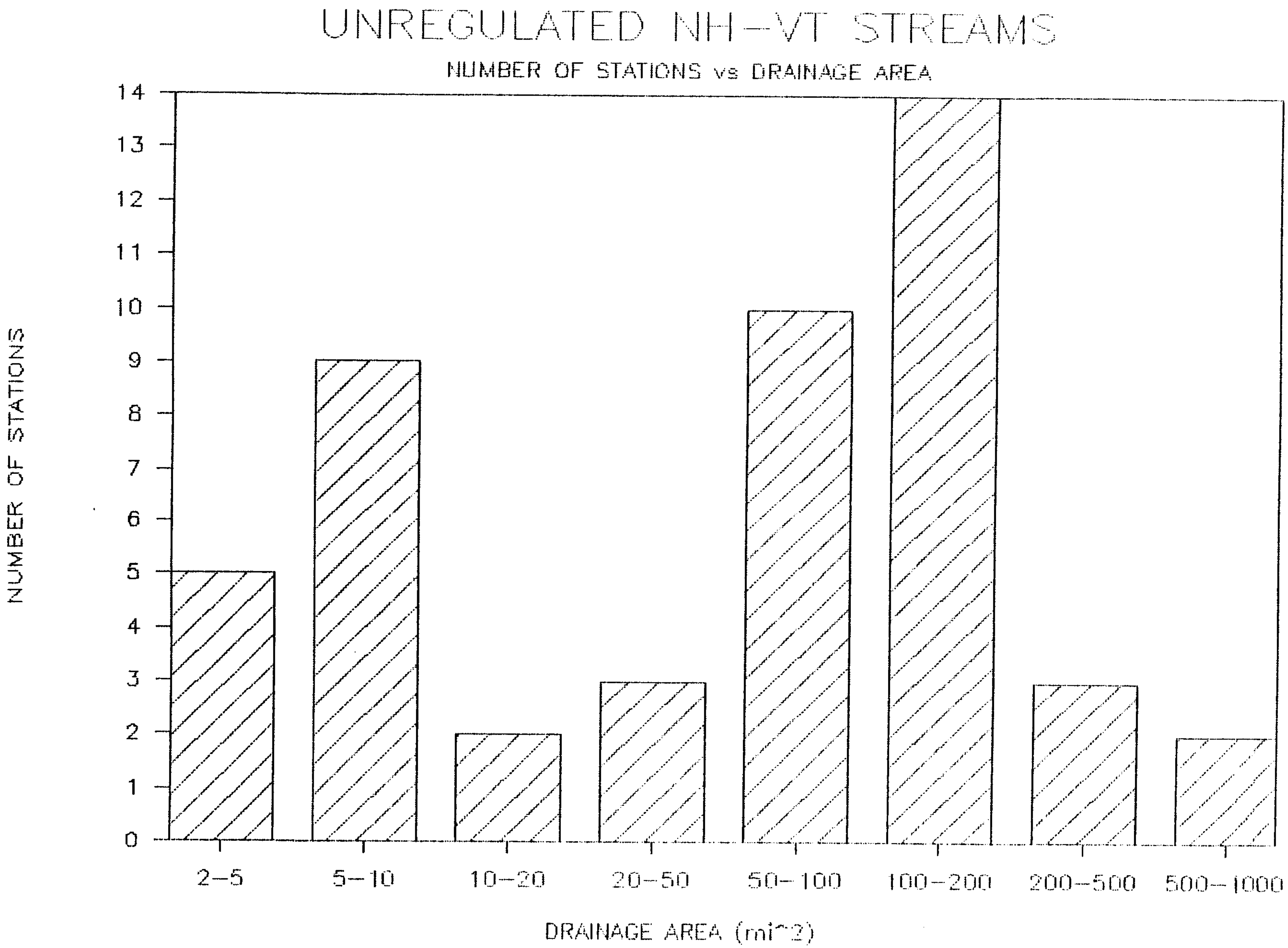
0 50 km
0 30 mi
CONNECTICUT RIVER
VERMONT
NEW HAMPSHIRE
N
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Figure 1.



UNREGULATED NH—VT STREAMS
NUMBER OF STATIONS vs RECORD LENGTH
NUMBER OF STATIONS
13
12
11
10
9
8
7
6
5
4
3
2
1
0
10–19
20–29
30–39
40–49
50–59
60–69
70–79
80–89
YEARS OF RECORD

Figure 2



UNREGULATED NH—VT STREAMS
NUMBER OF STATIONS vs DRAINAGE AREA
NUMBER OF STATIONS
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
2—5
5—10
10—20
20—50
50—100
100—200
200—500
500—1000
DRAINAGE AREA (mi^2)

Figure 3

Figure 4

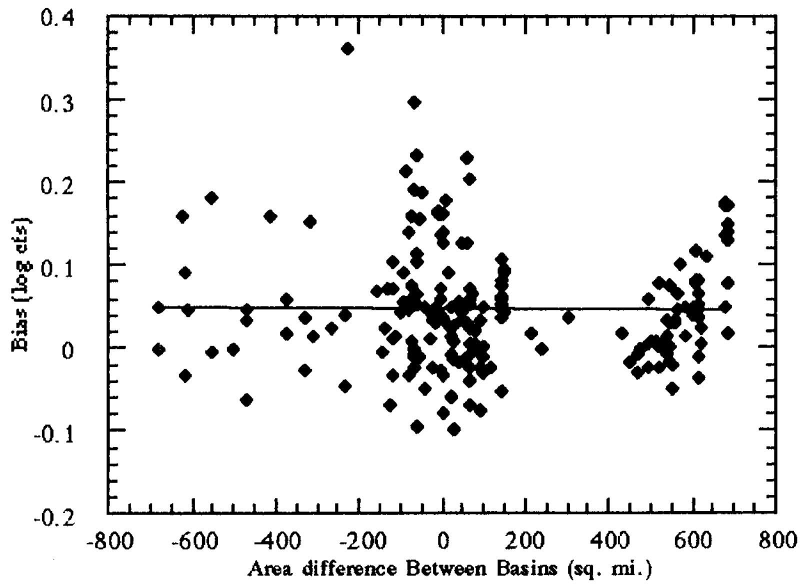


Figure 5

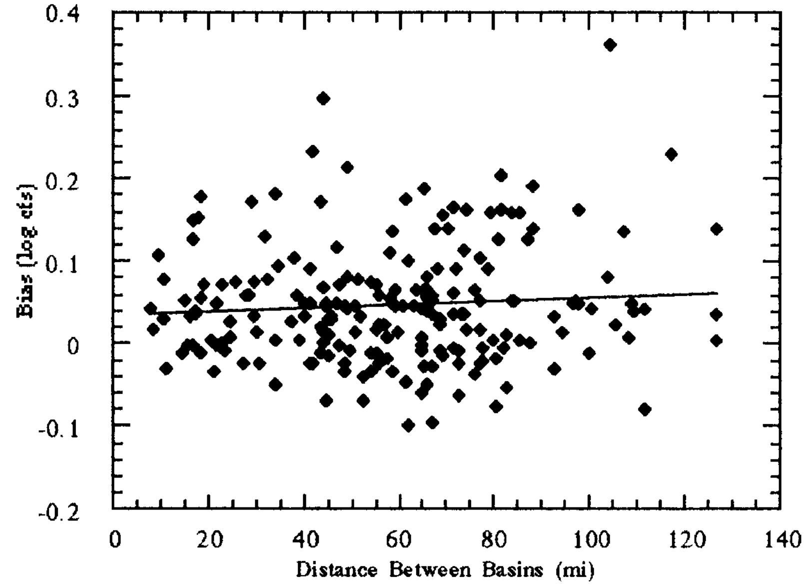


Figure 6

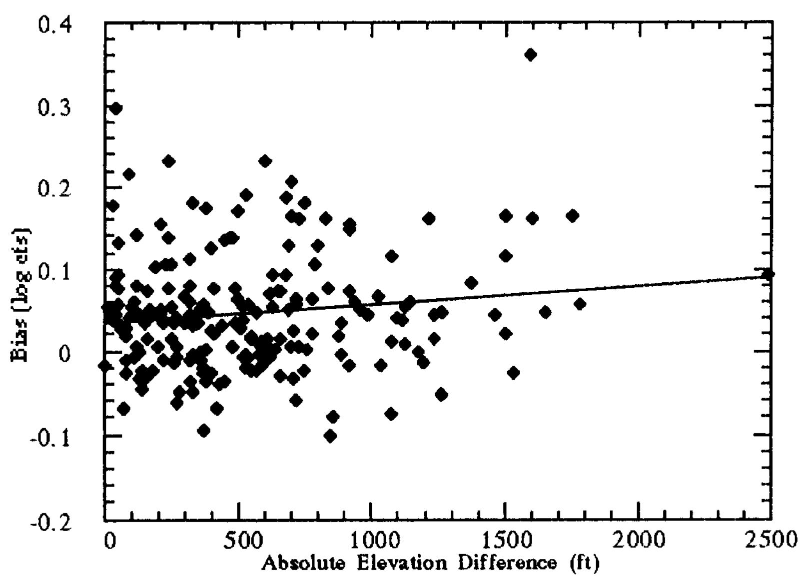


Figure 7

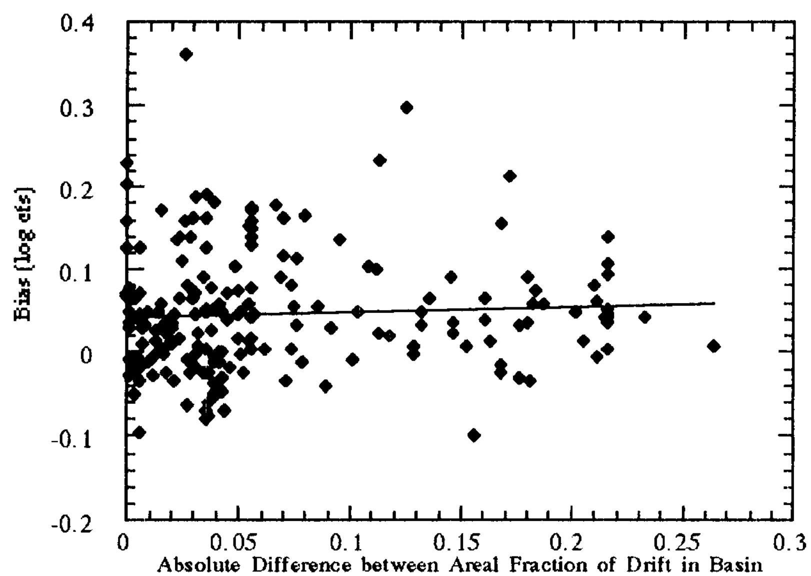
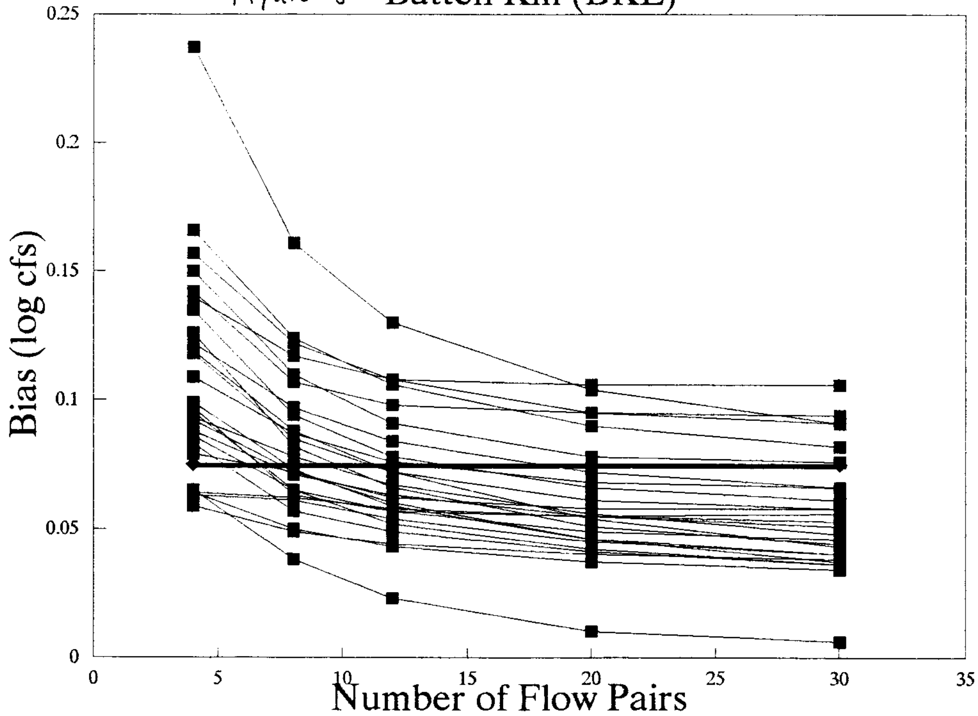


Figure 8. Batten Kill (BKL)



Comparison of difference in estimate from actual 7Q10 for
correlation of concurrent base flows and regression of basin characteristics

Figure 9. Black River (BLK)

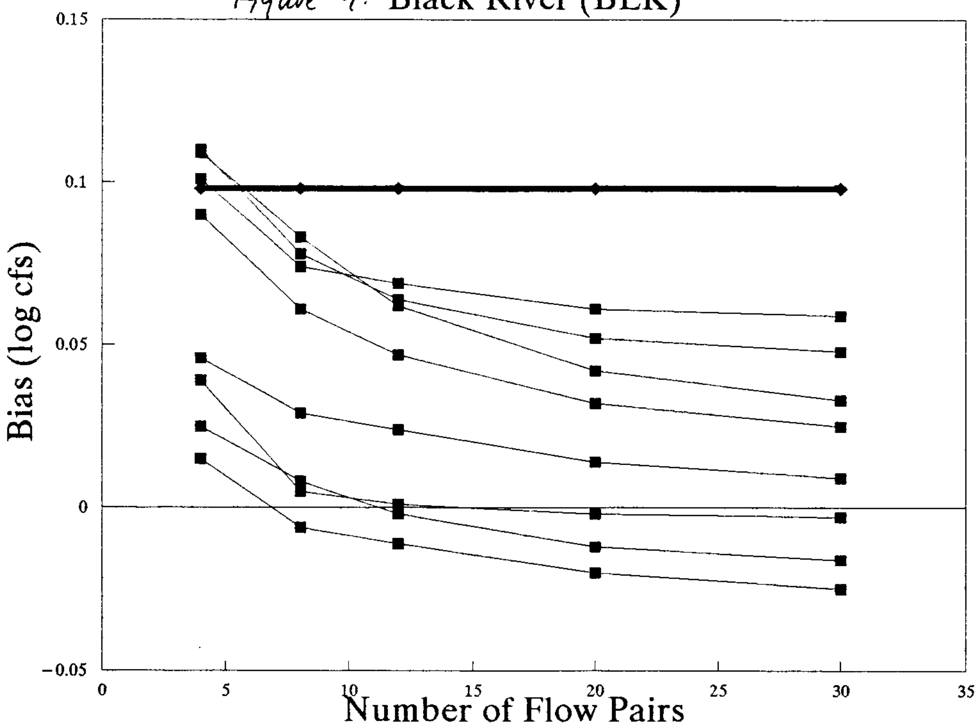
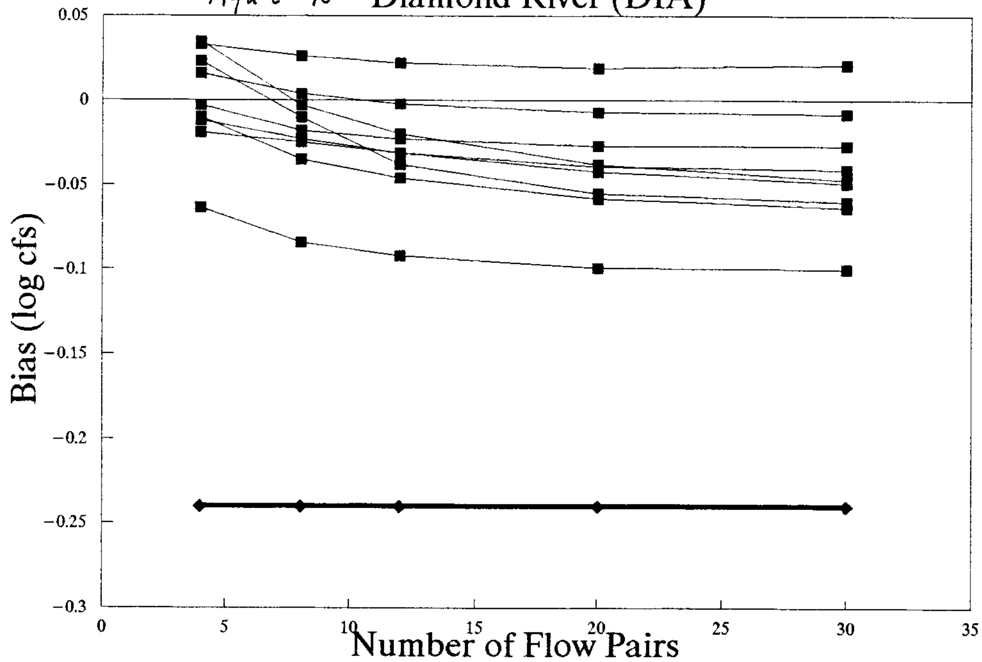
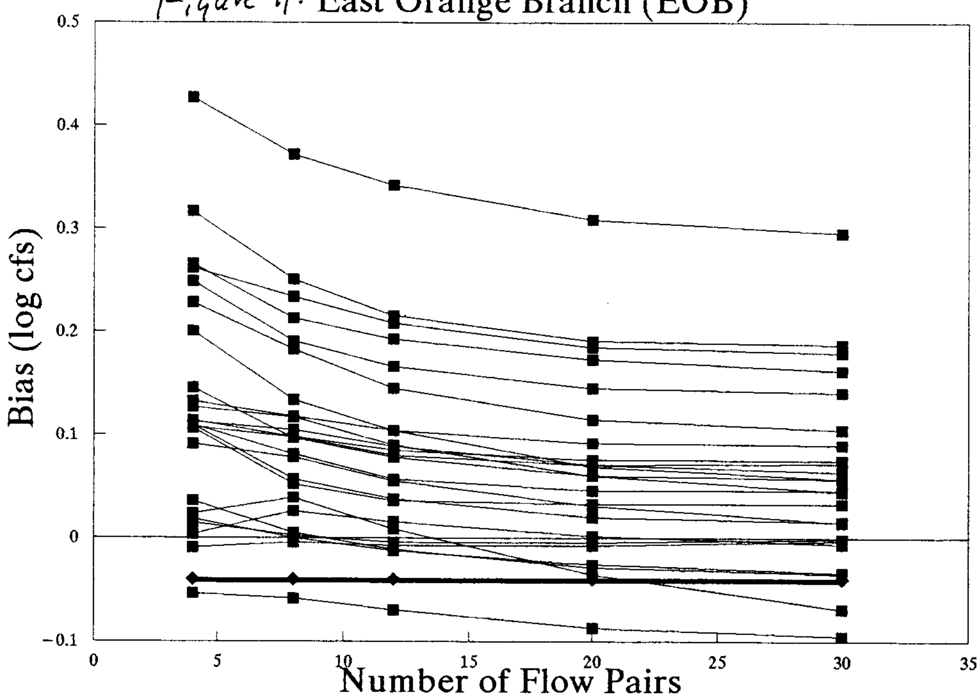


Figure 10. Diamond River (DIA)



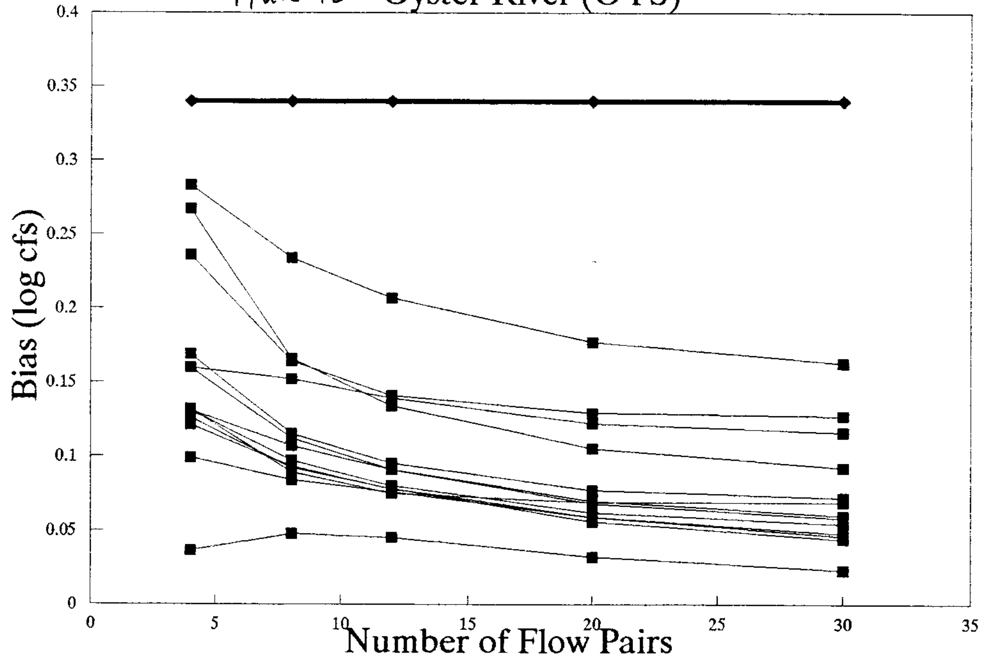
Comparison of difference in estimate from actual 7Q10 for
correlation of concurrent base flows and regression of basin characteristics

Figure 11. East Orange Branch (EOB)



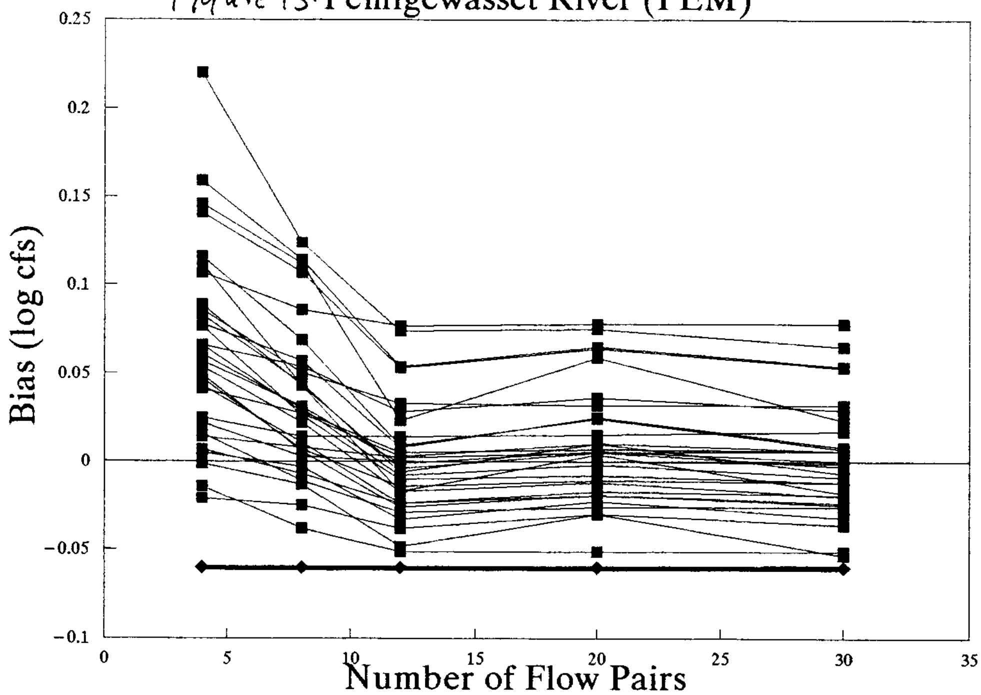
Comparison of difference in estimate from actual 7Q10 for
correlation of concurrent base flows and regression of basin characteristics

Figure 12. Oyster River (OYS)



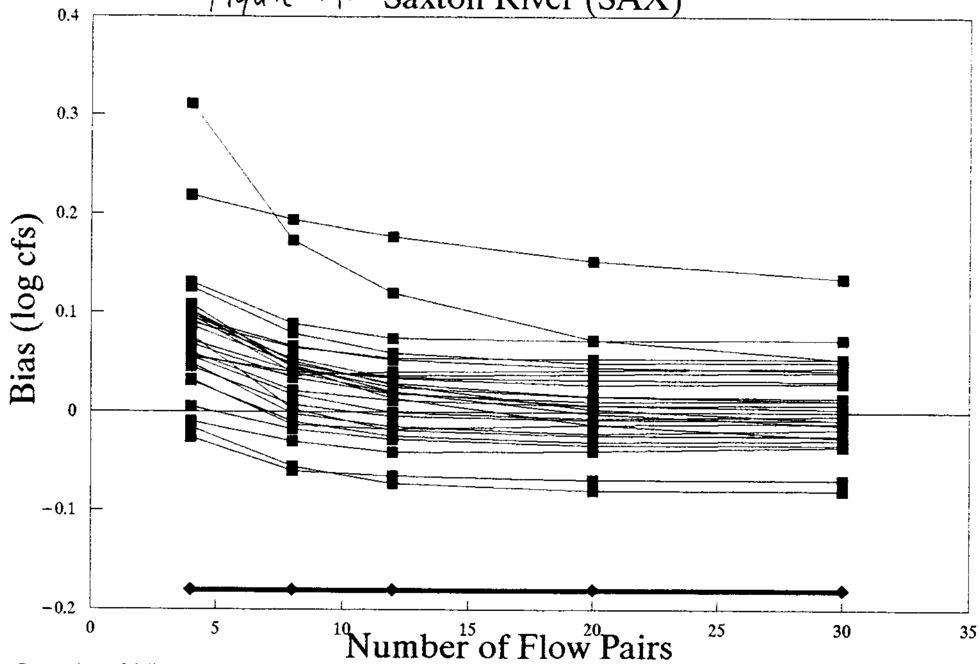
Comparison of difference in estimate from actual 7Q10 for
correlation of concurrent base flows and regression of basin characteristics

Figure 13. Pemigewasset River (PEM)



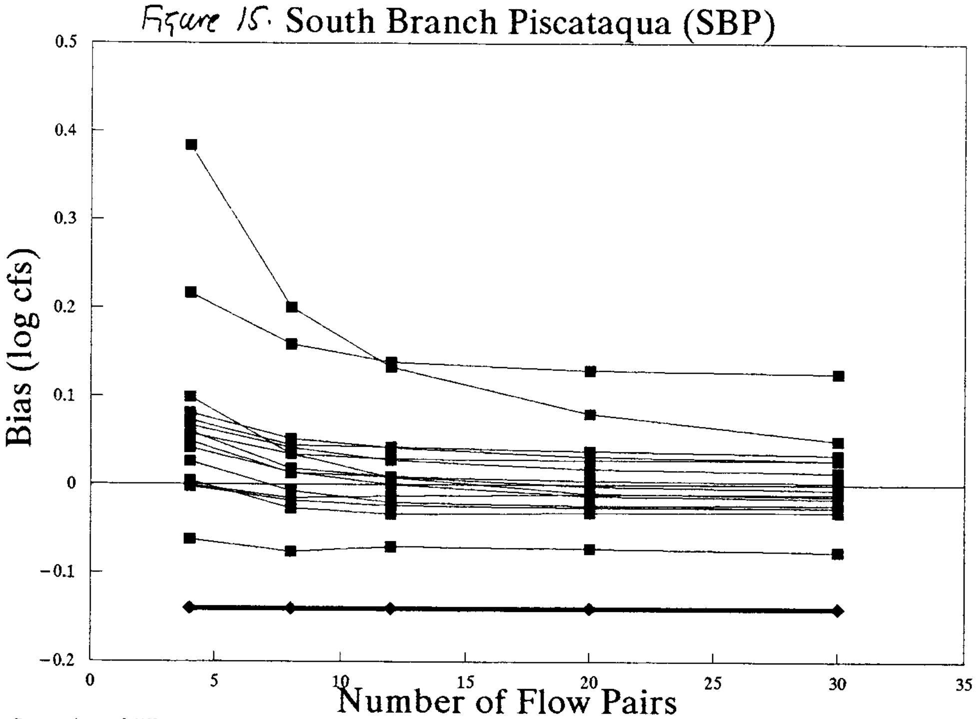
Comparison of difference in estimate from actual 7Q10 for
correlation of concurrent base flows and regression of basin characteristics

Figure 14. Saxton River (SAX)



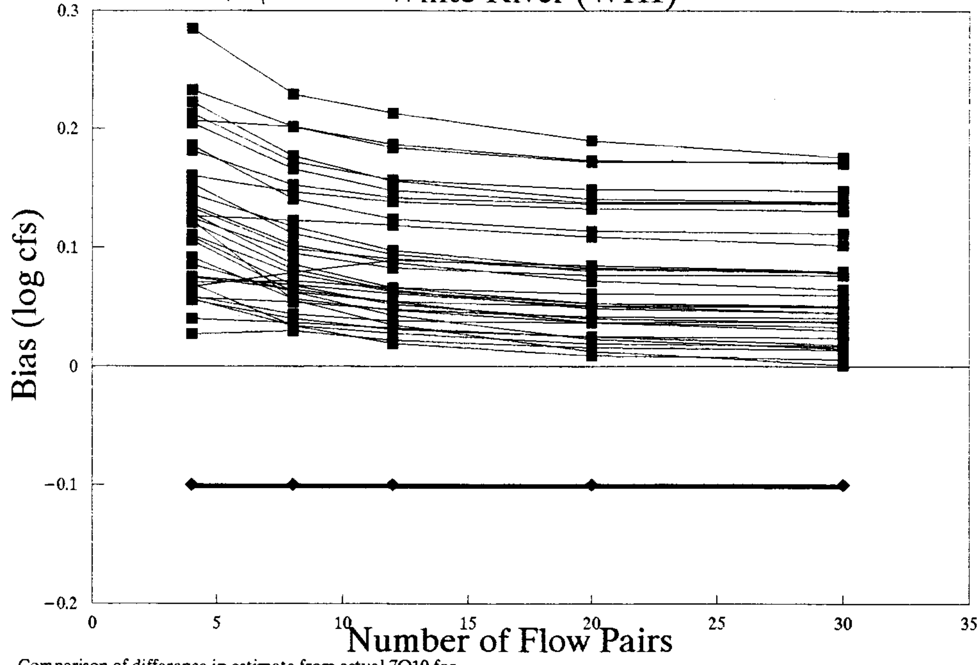
Comparison of difference in estimate from actual 7Q10 for correlation of concurrent base flows and regression of basin characteristics

Figure 15. South Branch Piscataqua (SBP)



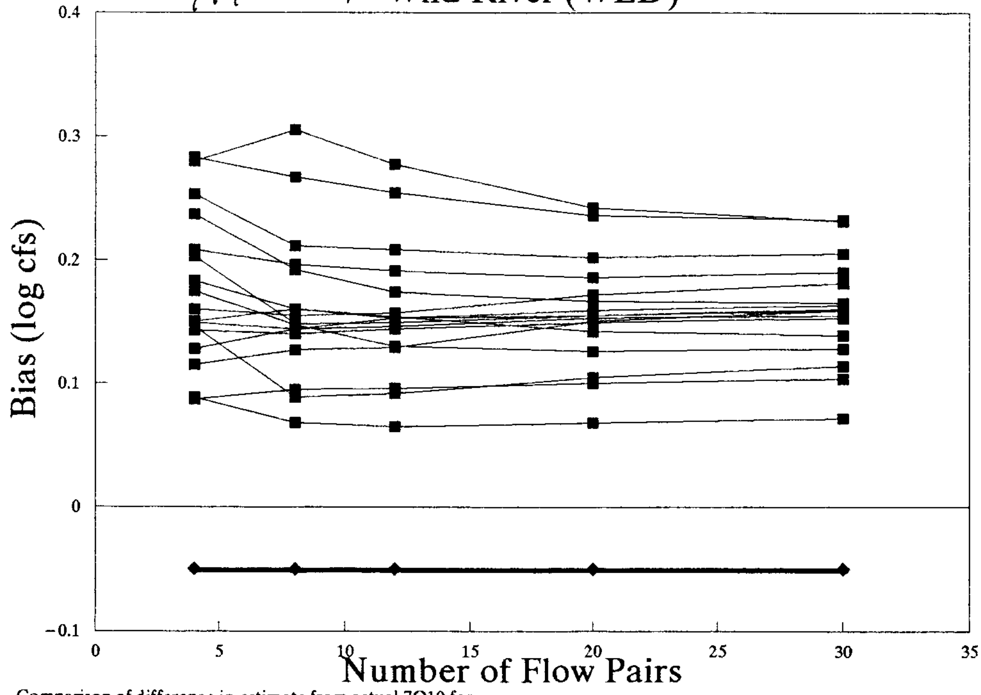
Comparison of difference in estimate from actual 7Q10 for correlation of concurrent base flows and regression of basin characteristics

Figure 16. White River (WHI)



Comparison of difference in estimate from actual 7Q10 for correlation of concurrent base flows and regression of basin characteristics

Figure 17. Wild River (WLD)



Comparison of difference in estimate from actual 7Q10 for correlation of concurrent base flows and regression of basin characteristics