Appendix F – Flow Metric Testing

Contents

This appendix provides a more detailed description of the analyses performed to verify the ability of the flow metrics generated from simulated time series to accurately represent those generated from observed time series.

The disc directory contains only this document.

Introduction

The ability of the HSPF model and WOOOMM routing module to represent stream flows in the Middle Potomac study area was investigated by comparing flow metrics calculated from model simulated time series and flow metrics calculated from daily time series measured at USGS gaging stations. The model is successful if simulated and observed flow metrics respond similarly to both natural environmental features and anthropogenic stressors. Flow metrics from a subset of 242 of the larger ELOHA watersheds (10 – 3,050 mi^2) were compared to a subset of 98 watersheds from the Potomac-Susquehanna gage flow dataset described in Appendix C. The subset of larger ELOHA watersheds was used in order to better match the range of watershed sizes in the Potomac-Susquehanna dataset and make comparisons between simulated and observed flow metrics equivalent. The dearth of USGS stream gages in watersheds less than 10 mi^2 precludes verifying model performance in smaller watersheds of the Middle Potomac study area. The relatively few Coastal Plain watersheds in each dataset were also excluded from the analysis because of their different hydrologic properties.

Five tests, each with a different purpose, were done to compare fifteen flow metrics calculated from the simulated and observed flow time series. The first three tests compare metric values in least-disturbed watersheds, where anthropogenic impacts do not confound flow metric responses to natural environmental factors. A baseline scenario, or flow time series representing least-disturbed conditions, is simulated for each ELOHA watershed in order to calculate flow alteration (Appendix E). Eleven gaged watersheds that meet or nearly meet the baseline scenario criteria were identified as least-disturbed, or "reference," in the Potomac-Susquehanna dataset (Table 1). If the model is accurately representing stream flows, flow metric values in the 242 baseline scenario watersheds and 11 reference watersheds should be very close. Also, the influence of natural environmental features will be most apparent in baseline and reference watersheds because they are not confounded by anthropogenic impacts. The fourth test compares the current scenario flow metrics to the observed flow metrics in the same 31 watersheds. Behaviors of the flow metrics in these watersheds should be very similar or identical. The fifth test compares flow metric responses to impervious surface, the anthropogenic factor in the Middle Potomac study area that appears to be primarily responsible for most flow alteration. Flow metrics from the 98 Potomac-Susquehanna watersheds, including the reference watersheds, were compared to the current scenario flow metric values for the 242 larger ELOHA watersheds, some of which also meet the baseline criteria. Responses of the observed and simulated metrics to anthropogenic impacts are not expected to match exactly because only 31 of the watersheds are common to both datasets and proportions of land and water uses differ in the other watersheds. The metrics should, however, show the same general responses to imperviousness.

Reference Watersheds

The eleven reference watersheds are more than 81 percent forested and except for one (South Branch Potomac River) they have less than 0.35 percent impervious surface, less than 0.5 percent surface withdrawals, and less than 1 percent impoundments (Table 1). The South Branch Potomac River has slightly higher levels of imperviousness (0.42 percent), surface withdrawals (1.90 percent), and impoundments (3.41 percent). It was included in the reference watersheds because its size (1,482 mi^2) extended the size range against which the effect of watershed size on flow metrics could be tested. The eleven watersheds range in size from 7.6 – 1,461 mi^2 with a median of 210 mi^2. Mean watershed slopes range from 3.4^o – 17.9^o with a median of 11.2^o. Elevations at gage locations range from 74 – 642 feet (22.4 – 195.6 meters). All of the reference watersheds have minimal amounts of karst (<6 percent).

Baseline Scenario vs Reference Watersheds

Comparisons of baseline scenario and reference flow metrics indicate if the model is adequately representing (1) actual values of the flow metrics in the absence of significant anthropogenic disturbance, and (2) the influence on flow metrics of natural environmental factors in the watershed. Representing the influence of natural factors is particularly important for the model to accomplish. When it does, simulated changes in the flow metrics relative to their baseline values (percent alteration) are accounting for the influence of natural factors on flow within each watershed. Watersheds therefore do not need to be classified by stream category in order to reduce natural variability in flow.

Mean Values

Mean values of the reference and baseline flow metrics were compared using the t-test assuming unequal variances (Table 2). Observed and simulated values are close (less than $\pm$10 percent) for only four of the fifteen metrics: median, flashiness, number of reversals, and high flow duration DH17. The observed values for 3-day maximum and annual mean are significantly higher than simulated values, August median is lower, high pulse duration is longer, and fall rate is faster ($p<0.01$). Means of a sixth metric, the frequency of extreme low flows, were weakly different ($p<0.05$) with more frequent extreme low flows in the observed data. Although not statistically significant, the observed rise rate is noticeably slower than simulated rise rates.

The model represents median flows very well. Comparisons of observed and modeled values of high flow duration DH17 (the average duration of flows above the median) are also good. The numbers of reversals and to some extent high pulse count and low pulse duration depend on the frequency of rain events, which is information input to the model and not generated by the model. Their good agreement should be expected. Simulated flashiness values appear to balance the faster fall rate and slower rise rate seen in the observed data and the result is simulated flashiness values that are very similar to observed.

Alteration in the flow metrics from baseline values is the preferred approach in the study for expressing change in flow. Flow alteration can be accurately represented by the model even if significant differences occur between observed and simulated flow metric values. This is true if (1) simulated baseline and reference flow metrics respond similarly to most natural and anthropogenic factors, and (2) differences between current values and their baseline pairs are proportional to those expected between observed values and a reference benchmark. The remaining comparisons evaluate these properties of the simulated flow metrics.

Responses to Watershed Size

Baseline scenario and reference flow metric responses to watershed size are shown in Figure 1 (left panel) and Table 3. Seven flow metrics do not respond to watershed size in both the Potomac-Susquehanna reference watersheds and the ELOHA baseline scenarios: mean, median, fall rate, number of reversals, extreme low flow frequency, high flow index MH21, and low pulse duration. The 3-day maximum, flashiness, and high pulse count respond significantly to watershed size along parallel courses in both datasets. Each metric decreases with increasing watershed size. The model seems to be accurately representing the responses of these ten flow metrics to watershed size in reference watersheds.

Of the remaining five metrics, it is not clear from the scatter plots if the reference and baseline flow metrics are responding differently to watershed size. Rise rate and high pulse duration show significant responses in the baseline scenarios and no responses in the reference watersheds. Despite the lack of statistical significance in the reference watersheds, the data trend in the same direction as baseline data as watershed size increases. Three metrics—August median, 3-day minimum and high flow duration DH17—show significant responses in the observed data but not in the simulated data. In each case, however, the observed data follows the lower edge of the cloud of simulated data points which suggests natural factors that are not in play in the observed watersheds are influencing watersheds represented in the upper edge of the simulated data cloud (Figure 1, left panel). A likely candidate is the percent karst geology. Removing watersheds with greater than 10 percent karst from the baseline data pool lowers the upper edges of the data point clouds for August median, 3-day minimum and high flow duration DH17, making the point distributions more like those for the observed data. Given the small sample size (11) of the reference dataset and the interplay of multiple environmental factors affecting flow, the significance of the reference data regression coefficients could change as sample size increases. Five near-reference quality watersheds from the Potomac-Susquehanna dataset (Ref+ in Table 3) were added to the reference group to determine if a slightly larger dataset would change the regressions. Regression p-values improved somewhat but the slopes of the regressions did not change substantially. Future examination of additional reference data is needed, but it appears as if the model has some difficulty representing alteration in at least these five flow metrics.

Responses to Slope

Baseline and reference flow metric responses to a gradient are shown in Figure 1 (middle panel) and Table 4. Responses of the observed and simulated flow metrics could not be directly compared.

Mean slope of the watershed was calculated for the Potomac-Susquehanna gaged watersheds[1] and mean slope of the stream channel was calculated for the ELOHA watersheds. Although related, these two parameters are not always closely related. Similar flow metric responses, however, should indicate if gradient is an influential environmental factor.

Six of the fifteen flow metrics show no effect of gradient in both the Potomac-Susquehanna reference data and the ELOHA baseline scenario: 3-day maximum, annual mean, median, fall rate, extreme low flow frequency, and low pulse duration. Two show significant, albeit weak, positive responses to increasing slope in both observed and simulated data: high pulse duration and high flow duration DH17. Rise rate and high flow index MH21 show significant relationships in baseline and discernible but non-significant relationships in reference. Of the remaining five metrics, baseline values of August median, 3-day minimum, flashiness, and high pulse count all come more into line with their reference counterparts when watersheds with percent karst greater than 10 percent are removed. Number of reversals, however, continues to show no response in baseline and a distinct and significant decrease in reference as watershed mean slope increases. [2]

Percent Karst

The influence of karst cannot be confirmed independently with observed data because no Potomac-Susquehanna watersheds with substantive amounts of karst meet the reference criteria. Karst appears to be a factor in the model that affects simulated hydrology and flow metrics (Figure 1, right panel and Table 5). In baseline watersheds (limited to those greater than 10 percent karst to minimize other influences), eight flow metrics are negatively related to percent karst, two are positively related, and five show no relationship to percent karst. The karst effects that are imposed on flow by the model explain many of the changes found when karsted watersheds are removed from baseline flow metric regressions with watershed size and slope (above). Removal of watersheds with greater than 10 percent karst typically brought several baseline flow metric regressions with size and slope into line with reference regressions. These included 3-day minimum, August median, high flow duration DH17, flashiness, and high pulse count.

The Same Watersheds

The fourth test of the model compares simulated and observed flow metrics for the same watersheds. Flow metrics simulated for the current scenario and flow metrics calculated from gaged data were available for 31 watersheds. Watershed characteristics under consideration included land uses (percent impervious cover, urban, forest, and agriculture) and water uses (fraction withdrawals of median flow, fraction discharges of median flow, and percent impoundment normal storage capacity of median annual flow volume). The Recursive Partitioning and Regression Tree, or RPART, analysis tool of the R software package was utilized for this purpose (Venables and Smith

---

[1] To calculate watershed mean slope, the ArcGIS 9.2 Spatial Analyst's Calculate Slope tool was used to measure the slope of each 30m grid cell in a watershed from the National Elevation Dataset (NED) 30m resolution raster elevation grid. Slope statistics, including the mean, are then calculated for each watershed using Spatial Analyst's Zonal Statistics tool.

[2] Only two of the reference watersheds in the Potomac-Susquehanna dataset have mean slopes less than ~10°. No conclusions about the effect on number of reversals of mean watershed slopes less than 10° can be drawn from the Potomac-Susquehanna data.

2011). During multiple RPART runs, flow metrics were evaluated as dependent variables and the anthropogenic factors were included as independent variables.

For purposes of this test, model results were considered acceptable if the RPART analysis identified the same primary anthropogenic factor (independent variable that splits the data first) and a similar splitting threshold for the factor (+/-10 percent). This screening process provides further justification that the hydrologic model is performing as expected and is able to adequately replicate the behavior of flow metrics in responses to changes in water and land uses.

The 31 ELOHA watersheds correspond to HSPF calibration locations where long term USGS gages are located. Overall, the simulated flow metrics behaved comparably to the observed flow metrics, particularly as influenced by land uses (Table 6). Only a handful of simulated and observed splits matched or were similar for withdrawals and discharges. This is likely due to the relatively small non-zero sample size of the withdrawal (n=18) and discharge (n=16) datasets. Similarly, only four of the test watersheds contained impoundments. As a result, impoundments were not identified as a primary driver of change in most flow metrics.

Similar primary breaks were not identified for the mean of the high flow volume metric. For other metrics such as median flow, low pulse duration, and 7Q10, only a small number of similar breaks were found. This suggests that either the watershed characteristics under consideration do not have a distinct impact on those flow metrics or that the model does not adequately represent those metrics.

Current Scenario vs Observed Watersheds

The fifth test of the model compares current scenario and observed flow metric responses to an important and common landscape feature in the watershed, the percent of impervious surface area. Imperviousness is a criterion used to simulate baseline scenarios, along with greater than 78 percent Forest cover, no surface withdrawals, impoundments, or discharges. The project's initial Category and Regression Tree (CART) analyses of the Potomac-Susquehanna dataset showed that impervious surface area is a strong and frequently occurring factor associated with significant flow alteration (Appendix E). A very low percentage of imperviousness (0.35 percent) is required for baseline conditions. As mentioned earlier, responses of simulated and observed flow metrics to anthropogenic impacts are not expected to match exactly because proportions of land and water uses in individual watersheds differ. They should, however, show the same general responses and these responses should explain a similar amount of the variability in the data. Current scenario data was limited to watersheds greater than 10 mi^2, making them comparable in size to the Potomac-Susquehanna watersheds. Coastal Plain watersheds were again excluded from the Potomac-Susquehanna data. The test results are summarized in Table 7. Scatter plots of the current and observed flow metrics versus percent impervious are provided in Figure 2. Apparent in the graphs are the differences noticed between the baseline and reference mean values (Table 2).

Seven of the fifteen current and observed flow metric responses to percent imperviousness match each other closely. Both responses show significant linear regressions with log-transformed percent imperviousness ($p<0.01$) that explain more than 10 percent of the flow metric's variability ($r^2>0.10$). They are: August median, 3-day minimum, rise rate, number of reversals, high pulse count, high flow index MH21, and high flow duration DH17. The annual mean and low pulse duration also have significant regressions but have shallow regression slopes and explain little of the variability. They are better described as having little or no response to percent imperviousness despite the significance of their regressions, and their regression lines are removed from Figure 2. Flashiness matches closely when two current outliers and one observed outlier are removed.

Of the five remaining flow metric comparisons of current and observed responses, three show differences in the level of significance of their regressions but the regressions explain so little of the variability ($r^2<0.10$) that these metrics are also better described as having little or no response to percent imperviousness. They are the median, fall rate, and extreme low flow frequency. Their regression lines are removed from Figure 2. The last two metrics—high pulse duration and 3-day maximum—both show significant, positive regressions in the current scenario and observed data, however the current scenario regressions have shallow slopes and explain little of the variability while the observed regressions have steeper slopes and explain much more of the variability. The model is not accurately representing magnitude of the highest annual flows or the duration of high flow events.

Summary Table

Table 8 merges the results of Tables 3-7 and gives qualitative ratings of how well simulated flows, as expressed by the flow metrics, represent actual flows. Five of the fifteen tested flow metric received an "excellent" rating because simulated and observed results showed good agreement in all five tests. The metrics are: median, flashiness, high pulse count, high flow duration DH17, and low pulse duration. Three of the tested metrics received a "good" rating because simulated and observed results of the five tests showed a combination of good agreement and qualified agreement. The metrics are: 3-day minimum, extreme low flow frequency, and high flow index MH21. Simulated values of the eight flow metrics rated "excellent" and "good" are thought to most accurately represent actual flows in Middle Potomac ungaged watershed when generated by the modeling and flow routing steps described in this report.

Three of the fifteen tested flow metrics received "fair" ratings because their baseline mean values significantly misrepresent reference mean values. The simulation overestimates baseline values for August median and underestimates baseline values for annual mean and fall rate. Although these same three metrics typically show good agreement in the four other tests, the large discrepancies in baseline values call into question the percent flow alteration values which are calculated as a percent of the baseline value.

Two other flow metrics also received "fair" ratings. Rise rate failed the paired comparison test on 31 watersheds, and number of reversals was significantly related to slope in the reference dataset but not in the baseline dataset. The reasons for assigning a "fair" rating may or may not be justified, however the metrics should have more evaluation before they are considered useful for generating flow alteration-ecology response curves.

Finally, two of the fifteen tested flow metrics failed two tests and consequently received a rating of "poor." They are the high magnitude metric 3-day maximum and the metric high pulse duration (average duration above the 90^{th} percentile). Baseline values of both significantly underestimate reference values, and responses of their current values to percent imperviousness do not match those of the observed values.

Table 1. Reference watersheds identified in the Potomac-Susquehanna gaged watershed dataset. Sub-basin: P, Potomac; S, Susquehanna. Abbreviations: %Urb, percent urban area; %Agr, percent agricultural area; %For, percent forest area; %Impv, percent impervious surface area; %Karst, percent karst surface geology; %SurfWith, surface withdrawals expressed as a percent of median annual flow volume; %Impnd, impoundment storage capacity expressed as a percent of median annual flow volume.

Tributary Short Name	USGS Gage #	Sub-basin	Area (mi^2)	Slope (^{o})	%Urb	%Agr	%For	%Impv	%Karst	%Surf With	%Impnd
Quantico	01658500	P	7.62	3.4	0.3%	0.0%	99.4%	0.01%	0.0%	0.00%	0.00%
U1_Tonoloway	01613050	P	10.7	9.6	2.9%	1.8%	89.7%	0.18%	0.0%	0.00%	0.00%
U1_North River	01620500	P	17.3	16.5	2.3%	16.1%	81.5%	0.11%	0.0%	0.00%	0.00%
Savage River	01596500	P	49.1	11.2	4.5%	13.3%	81.3%	0.32%	0.0%	0.00%	0.31%
Cedar Creek, Shenandoah	01634500	P	102	10.4	3.0%	5.1%	90.5%	0.05%	5.7%	0.00%	0.67%
U1_NorthFork	01632000	P	210	14.7	2.7%	13.1%	84.1%	0.19%	0.0%	0.09%	0.02%
Wills Creek	01601500	P	247	11.2	1.7%	3.3%	91.8%	0.20%	0.7%	0.00%	0.56%
U2_South Branch	01606500	P	651	16.2	2.4%	13.8%	81.4%	0.30%	2.4%	0.47%	0.00%
Cacapon River	01611500	P	675	10.9	1.0%	9.3%	87.5%	0.16%	0.0%	0.03%	0.98%
Pine Cr. nr Waterville PA	01549700	S	944	11.8	2.6%	6.6%	90.0%	0.20%	0.2%		0.44%
South Branch Potomac Riv.	01608500	P	1,461	14.5	1.5%	13.4%	81.8%	0.42%	1.0%	1.90%	3.41%

Some watersheds or subwatersheds that meet the baseline land and water use criteria are missing daily mean flow data. The North Fork of the Potomac River South Branch gage is missing daily records for 1984-March 1998 and was excluded. The Cacapon River, included above, is missing 366 daily records from water year 1996. This was a particularly high flow year and the Cacapon flow metrics representing high flows differ slightly from other reference watersheds. U1_Tonoloway is missing one daily record. All the remaining reference watersheds have complete daily mean flow records for water years 1984 through 2005. Pine Creek near Waterville PA met the land use and impoundment criteria for least-disturbed watersheds but no actual surface withdrawal data was available.

Table 2. Mean flow metric values in baseline and reference watersheds (t-test assuming unequal variance). Agreement: flow metric are statistically very similar (●), somewhat similar (O), and different (X) in baseline and reference watersheds. p-value: ns, not significant; **, $p<0.01$; *, $p<0.05$.

Flow Metric	Baseline[a] mean (n=242)	Reference mean (n=11)	p-value	Agree-ment
3-day maximum (cfs/mi^2)	8.455	12.501	**	X
annual mean (cfs/mi^2)	0.997	1.279	**	X
median (cfs/mi^2)	0.529	0.553	ns	●
August median (cfs/mi^2)	0.222	0.135	**	X
3-day minimum (cfs/mi^2)	0.075	0.063	ns	●
flashiness (ratio)	0.402	0.406	ns	●
rise rate (cfs/mi^2)	0.114	0.090	ns	●
fall rate (cfs/mi^2)	-0.027	-0.054	**	X
number of reversals (#)	87.44	93.36	ns	●
high pulse count (#)	8.5	10.5	ns	●
extreme low flow frequency (#)	0.98	1.82	*	O
high flow index, MH21 (days)	50.42	62.47	ns	●
high pulse duration (days)	2.4	5.25	**	X
high flow duration, DH17 (days)	21.3	19.7	ns	●
low pulse duration (days)	7.80	9.15	ns	●

[a] only watersheds >10 mi^2 and not located in the Coastal Plain bioregion

Table 3. Regressions between flow metrics and ln watershed size. All regressions are linear except for 3-day maximum which is exponential. Agreement: flow metric relationships with size are statistically very similar (●), somewhat similar (O), and opposing (X) in baseline watersheds >10mi^2 (n=242) and reference watersheds (n=11). p-value: ns, not significant; **, p<0.01; *, p<0.05. Slope: neg, regression slope is negative; pos, regression slope is positive. Parentheses indicate non-significant p-values with a discernable regression slopes in the reference data. Ref+, reference enhanced with five near-reference quality watersheds from the Potomac-Susquehanna gage dataset (see text for details). [1] Removal of watersheds with >10% karst makes baseline watershed characteristics align better with those of the reference watersheds.

	Baseline[a]			Reference			Ref+	
	p-value	r^2	slope	p-value	r^2	slope	p-value	Agree-ment
3-day maximum (cfs/mi^2)	**	0.143	neg	*	0.482	neg	*	●
annual mean (cfs/mi^2)	ns			ns			ns	●
median (cfs/mi^2)	ns			ns			ns	●
August median (cfs/mi^2)	ns[1]			*	0.389	pos	*	O[1]
3-day minimum (cfs/mi^2)	ns[1]			*	0.533	pos	**	O[1]
flashiness (ratio)	**	0.158	neg	*	0.468	neg	**	●
rise rate (cfs/mi^2)	**	0.136	neg	ns		(neg)	ns	O
fall rate (cfs/mi^2)	ns			ns			ns	●
number of reversals (#)	ns			ns			ns	●
high pulse count (#)	**	0.135	neg	0.051	0.382	(neg)	0.053	●
extreme low flow frequency (#)	ns			ns			ns	●
high flow index, MH21 (days)	ns			ns			ns	●
high pulse duration (days)	**	0.245	pos	ns		(pos)	ns	O
high flow dur, DH17 (days)	ns[1]			*	0.431	pos	*	O[1]
low pulse duration (days)	ns			ns			ns	●

[a] only watersheds >10 mi^2 and not located in the Coastal Plain bioregion

Table 4. Linear regressions of flow metrics versus slope. Agreement: the baseline scenario flow metric relationship with mean channel slope (%) in watersheds greater than $10 mi^2$ and the reference flow metric relationship with watershed mean slope (°) are similar (●),somewhat similar (O), and different (X). p-value: ns, not significant; **, p<0.01; *, p<0.05. Slope: neg, regression slope is negative; pos, regression slope is positive. Parentheses indicate weak p-values or non-significant p-values with a discernable regression slopes in the reference data. Ref+, reference enhanced with five near-reference quality watersheds from the Potomac-Susquehanna gage dataset (see text for details). **[1]** Removal of watersheds with more than 10 percent karst makes baseline watershed characteristics align better with those of the reference watersheds.

	Baseline[a] (n=242)			Reference (n=11)			Ref+	
	p-value	r^2	slope	p-value	r^2	slope	p-value	Agree-ment
3-day maximum (cfs/mi^2)	ns			ns			ns	●
annual mean (cfs/mi^2)	ns			ns			ns	●
median (cfs/mi^2)	ns			ns			ns	●
August median (cfs/mi^2)	ns[1]			*	0.384	pos	*	O[1]
3-day minimum (cfs/mi^2)	*[1]	0.017	neg	ns			ns	O[1]
flashiness (ratio)	ns[1]			*	0.431	neg	**	O[1]
rise rate (cfs/mi^2)	*	0.026	neg	ns		(neg)	ns	O
fall rate (cfs/mi^2)	ns			ns			ns	●
number of reversals (#)	ns			**	0.702	neg	**	X
high pulse count (#)	ns[1]			**	0.655	neg	**	O[1]
extreme low flow frequency (#)	ns			ns			ns	●
high flow index, MH21 (days)	**	0.050	pos	ns		(pos)	ns	O
high pulse duration (days)	**	0.029	pos	0.06	0.340	pos	**	●
high flow dur, DH17 (days)	*	0.018	pos	*	0.375	pos	*	●
low pulse duration (days)	ns			ns			ns	●

[a] only watersheds >10 mi^2 and not located in the Coastal Plain bioregion

Table 5. Linear relationships between flow metrics and percent karst in baseline[a] watersheds with greater than 10 percent karst (n=77). p-value: **, p<0.01; *, 0.01<0<0.05; ns, non-significant. Pos, metric value increases with increasing %karst; neg, metric value decreases with increasing %karst.

Flow Metric	p	r^2	
3-day maximum (cfs/mi^2)	**	0.403	neg
annual mean (cfs/mi^2)	**	0.248	neg
median (cfs/mi^2)	ns		
August median (cfs/mi^2)	**	0.166	pos
3-day minimum (cfs/mi^2)	**	0.136	pos
flashiness (ratio)	**	0.322	neg
rise rate (cfs/mi^2)	**	0.248	neg
fall rate (cfs/mi^2)	**	0.449	neg
number of reversals (#)	ns		
high pulse count (#)	**	0.366	neg
extreme low flow frequency (#)	ns		
high flow index, MH21 (days)	**	0.183	neg
high pulse duration (days)	ns		
high flow duration, DH17 (days)	**	0.136	neg
low pulse duration (days)	ns		

[a] only watersheds >10 mi^2 and not located in the Coastal Plain bioregion

Table 6. Results of the RPART analysis to determine if simulated current scenario flow metrics behave like observed flow metrics for the same watersheds.

	Forest	Ag	Urban	Impervious	Withdrawals	Discharges	Impnd
High flow index, MH21							
High flow duration		*	†				
High pulse count	*	*	*			†	
High flow frequency		*	*		†		
Flood frequency	*	*	*	*			
Skewness of ann. maximum	*	*	*			†	
Median		*				†	
Flood free season	†	*	*	†	†		
Fall rate	†				†		†
Flashiness	*	*	*	*	†		
4B3/area	†		†	*			
Q85Seas	*	*	*	*			
Low pulse duration						†	
Extreme low flow duration	*	*	*	*			
CV, low flow duration	†	†	†	*	†		
Low pulse count	*	*	†	†			
Extreme low flow frequency	†	*					
7Q10						†	

*=identical RPART thresholds in simulated versus observed flows

†=RPART thresholds +/-10% in simulated versus observed flows

Green=no primary splits on these watershed factors (the watershed characteristic was not important in discerning the variability in the flow metric values)

RPART equation: ~ Curr_For + Curr_Agr + Curr_Urb + Curr_Imp + Frac_With + Frac_Disch + Frac_Imp, data = Simulated, control = rpart.control(minsplit = 20, maxcompete=3))

Table 7. Comparison of current and observed flow metric responses to watershed percent impervious surface area. The corresponding graph for each flow metric comparison is presented in Figure 2. Flow metric responses to imperviousness in current and observed datasets: good agreement (●), either a) current and observed regressions with land use (solid lines in graphs) are nearly parallel or are overlapping, or b) current and observed data show no relationship to land use and their mean values (dashed lines in graphs) are shown; qualified good agreement (O[1]), as above after highly impervious watersheds (greater than 10%) are removed; fair agreement (O), current and observed regression lines are only roughly parallel in the land use relationships; and poor agreement (X), current and observed relationships with land use are different.
[1] relationships are comparable if one or a few extreme value are removed from the observed data.

Flow Metric	Agreement
3-day maximum (cfs/mi^2)	X
annual mean (cfs/mi^2)	●
median (cfs/mi^2)	●
August median (cfs/mi^2)	●
3-day minimum (cfs/mi^2)	●
flashiness (ratio)	O[1]
rise rate (cfs/mi^2)	●
fall rate (cfs/mi^2)	●
number of reversals (#)	●
high pulse count (#)	●
extreme low flow frequency (#)	●
high flow index, MH21 (days)	●
high pulse duration (days)	X
high flow duration, DH17 (days)	●
low pulse duration (days)	●

Table 8. Summary table of how well simulated flows expressed as the flow metrics represent actual flows. Baseline and current are model-simulated daily flows; reference and observed are daily flows measured at USGS gaging stations. Baseline and reference are minimally disturbed watersheds; current and observed are watersheds with the full range of present-day land and water use conditions. Overall agreement: excellent, all categories are ● or ●+O[1]; good, all categories are ● + O; fair, X in one category but not in "response to land use;" poor, X in two or more categories, including "response to land use."
[1] relationships are comparable if one or a few extreme value are removed from the observed data.

	Baseline vs Reference			Current vs Observed		
Flow Metric	Mean value	Relationship with watershed size	Relationship with slope	Paired comparison	Response to land use	Overall Agreement
3-day maximum (cfs/mi²)	X	●	●	●	X	poor
annual mean (cfs/mi²)	X	●	●	●	●	fair
median (cfs/mi²)	●	●	●	●	●	excellent
August median (cfs/mi²)	X	O[1]	O[1]	●	●	fair
3-day minimum (cfs/mi²)	●	O[1]	O[1]	O	●	good
flashiness (ratio)	●	●	O[1]	●	O[1]	excellent
rise rate (cfs/mi²)	●	O	O	X	●	fair
fall rate (cfs/mi²)	X	●	●	O	●	fair
number of reversals (#)	●	●	X	●	●	fair
high pulse count (#)	●	●	O[1]	●	●	excellent
extreme low flow frequency (#)	O	●	●	O	●	good
high flow index, MH21 (days)	●	●	O	O	●	good
high pulse duration (days)	X	O	●	●	X	poor
high flow duration, DH17 (days)	●	O[1]	●	●	●	excellent
low pulse duration (days)	●	●	●	●	●	excellent

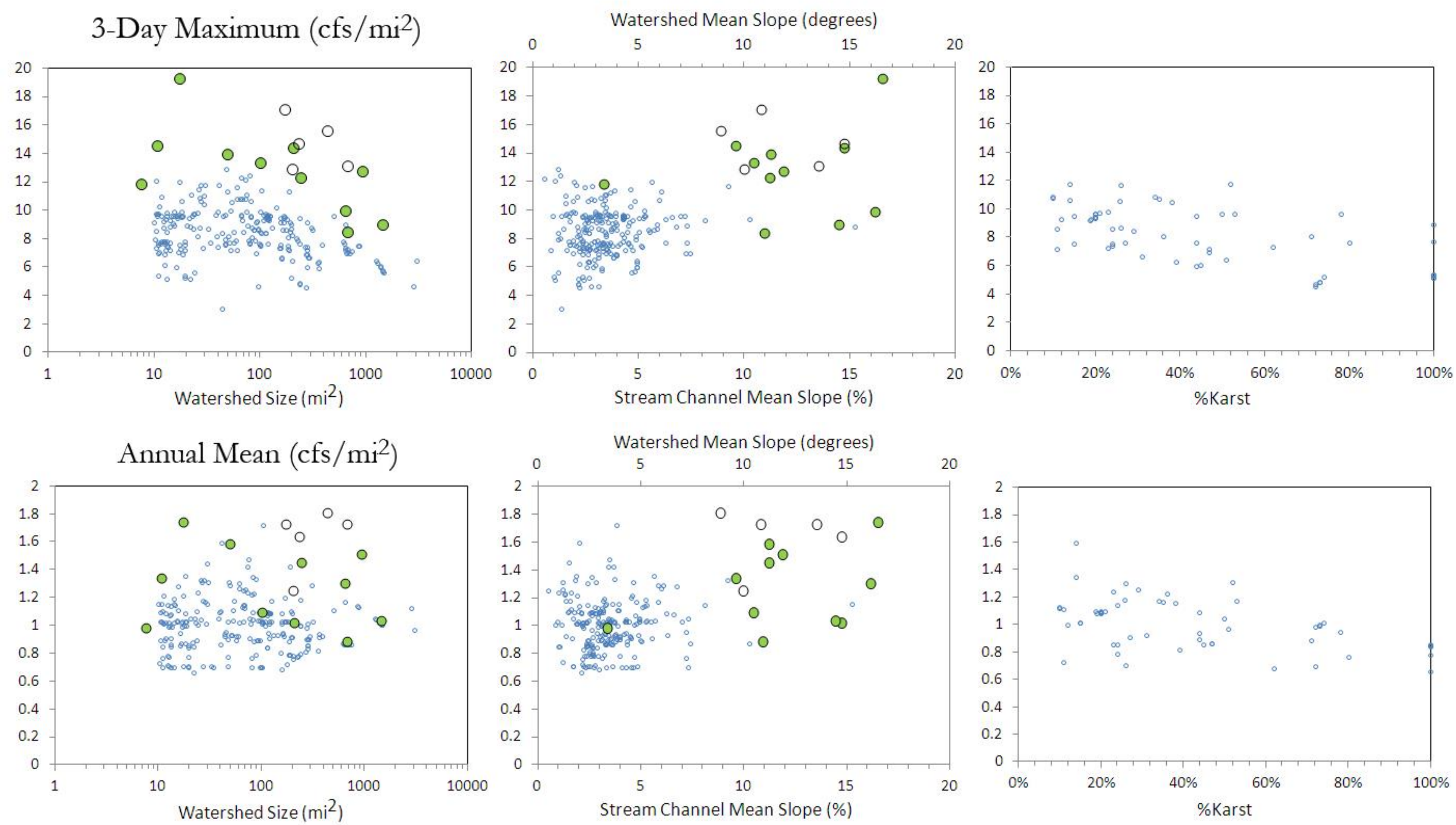


Figure 1. Scatter plots of baseline (o), reference (●), and near-reference (O) values for 3-day maximum and annual mean. Left panel, flow metric values versus watershed size; middle panel, baseline values versus stream channel mean slope and reference/near-reference values versus watershed mean slope; right panel, baseline values versus percent karst for watersheds with greater than 10% karst.

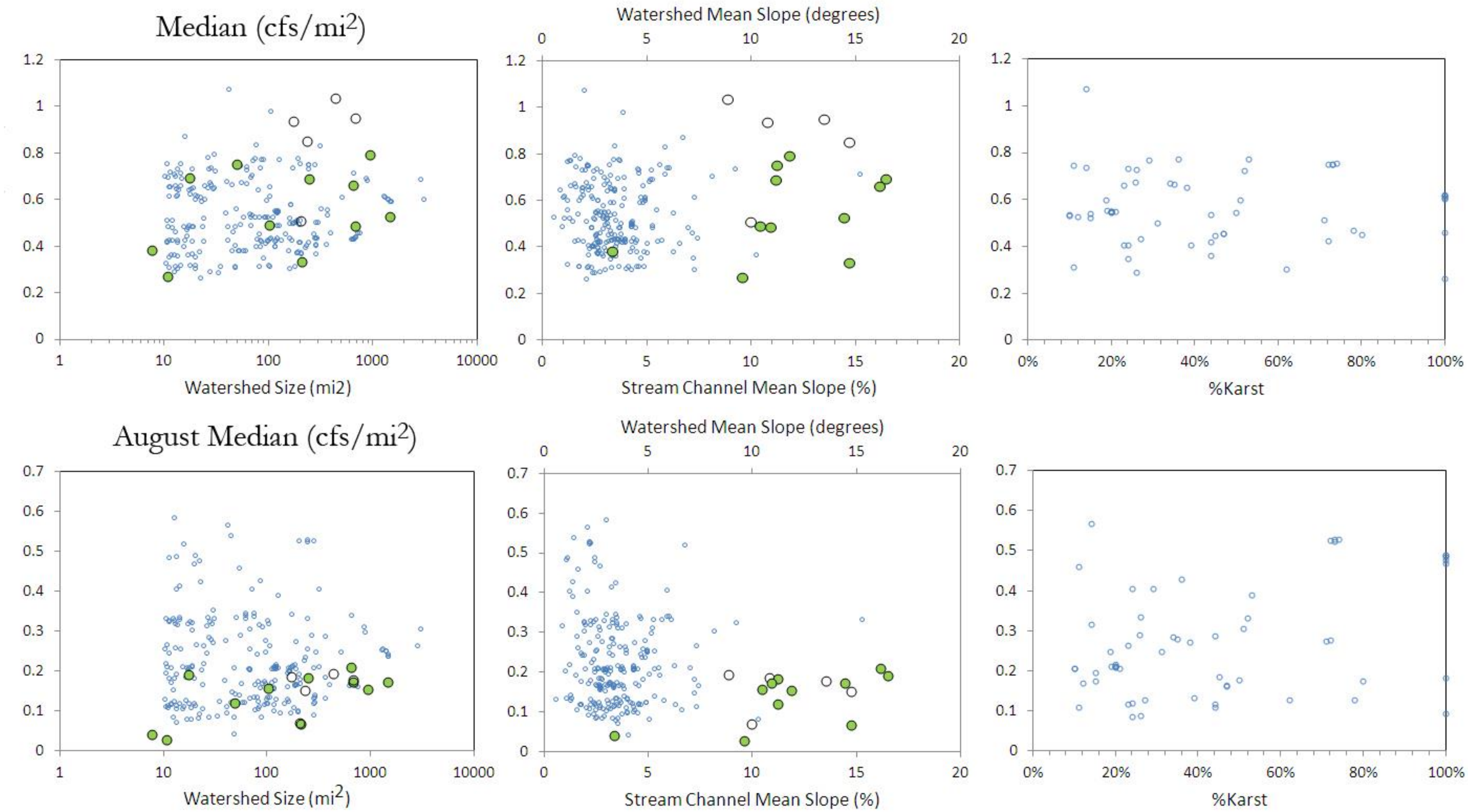


Figure 1 (cont.) Scatter plots of baseline (o), reference (●), and near-reference (O) values for median and August median. Left panel, flow metric values versus watershed size; middle panel, baseline values versus stream channel mean slope and reference/near-reference values versus watershed mean slope; right panel, baseline values versus percent karst for watersheds with greater than 10% karst.

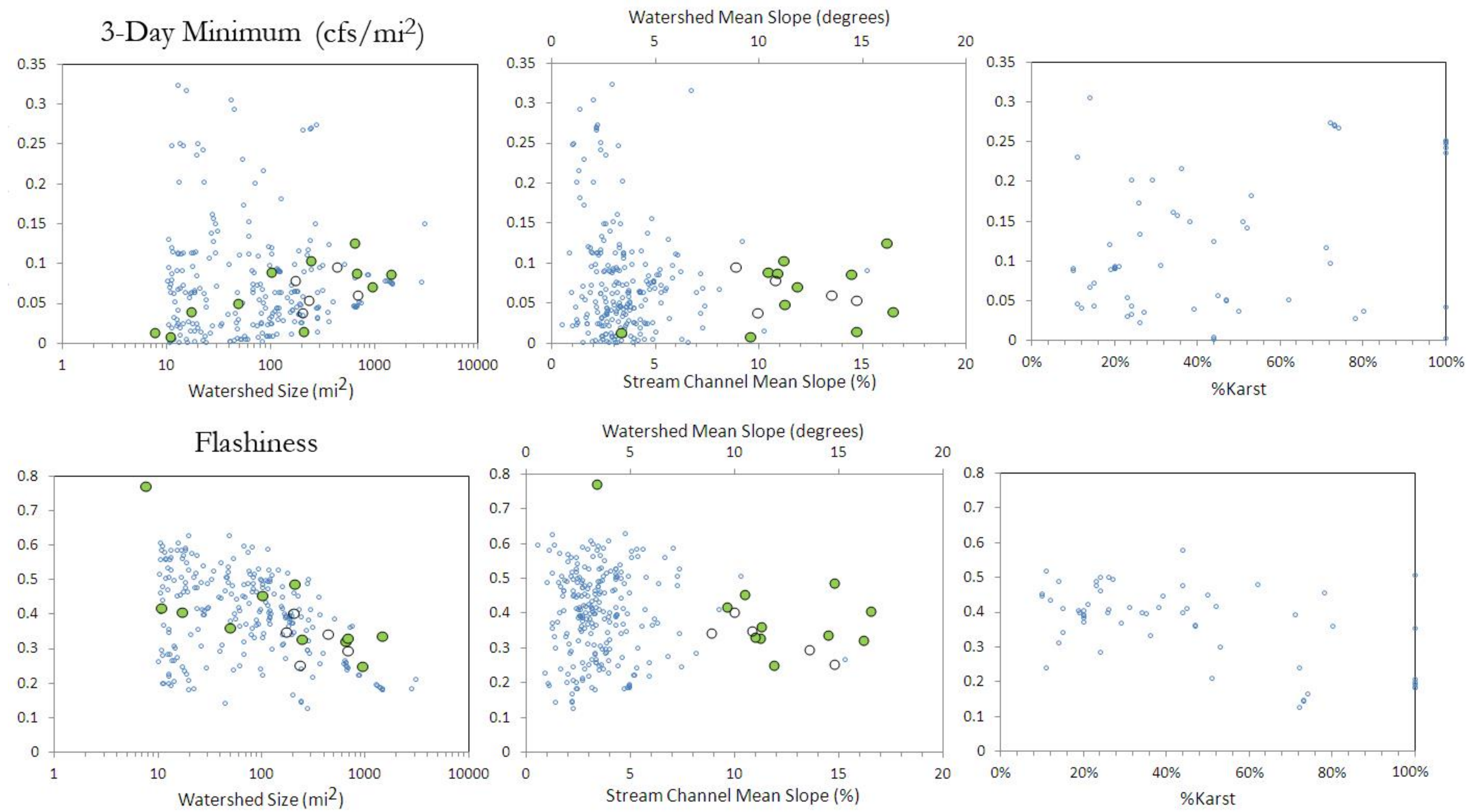


Figure 1 (cont.). Scatter plots of baseline (o), reference (●), and near-reference (O) values for 3-day minimum and flashiness. Left panel, flow metric values versus watershed size; middle panel, baseline values versus stream channel mean slope and reference/near-reference values versus watershed mean slope; right panel, baseline values versus percent karst for watersheds with greater than 10% karst.

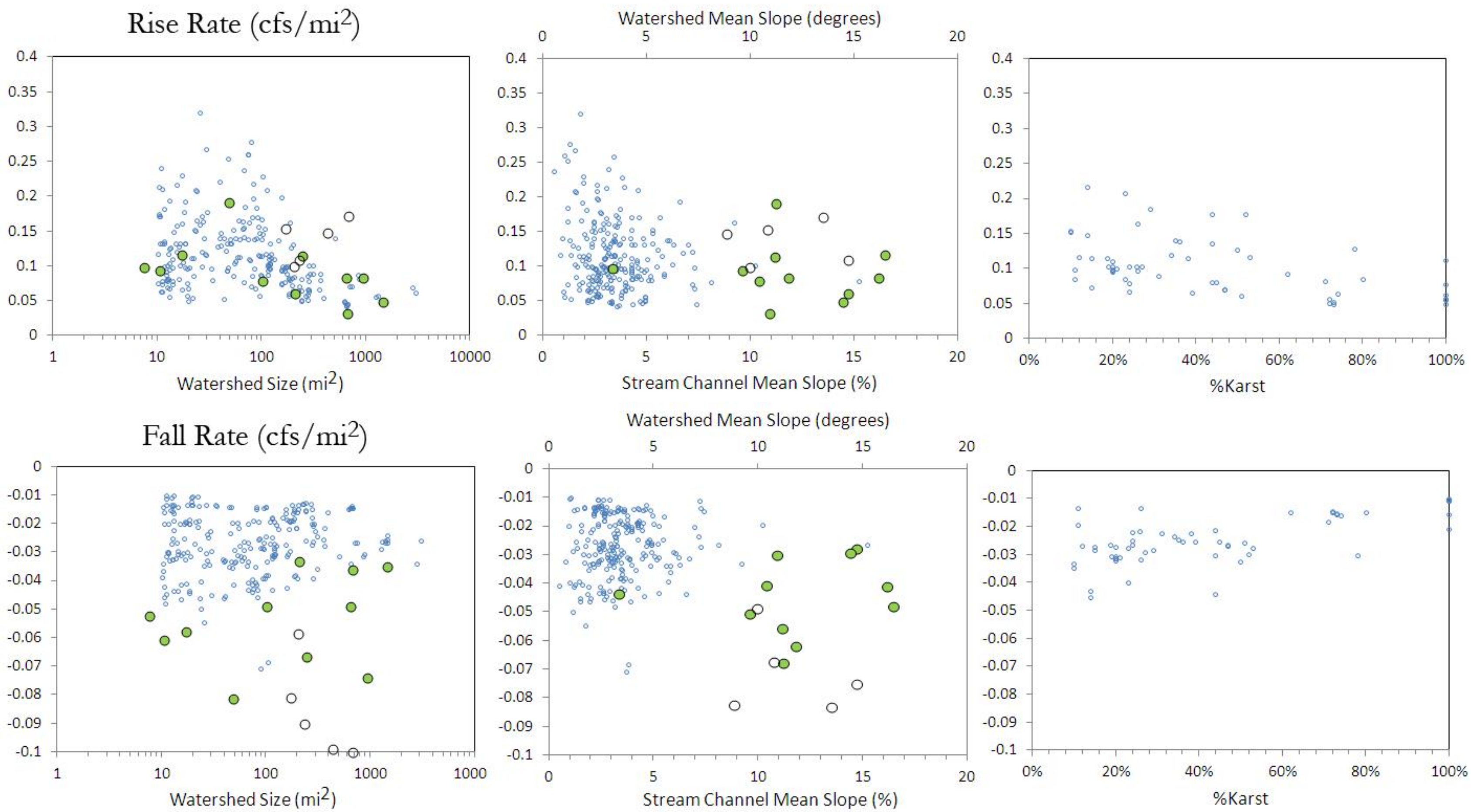


Figure 1 (cont.). Scatter plots of baseline (o), reference (●), and near-reference (O) values for rise rate and fall rate. Left panel, flow metric values versus watershed size; middle panel, baseline values versus stream channel mean slope and reference/near-reference values versus watershed mean slope; right panel, baseline values versus percent karst for watersheds with greater than 10% karst.

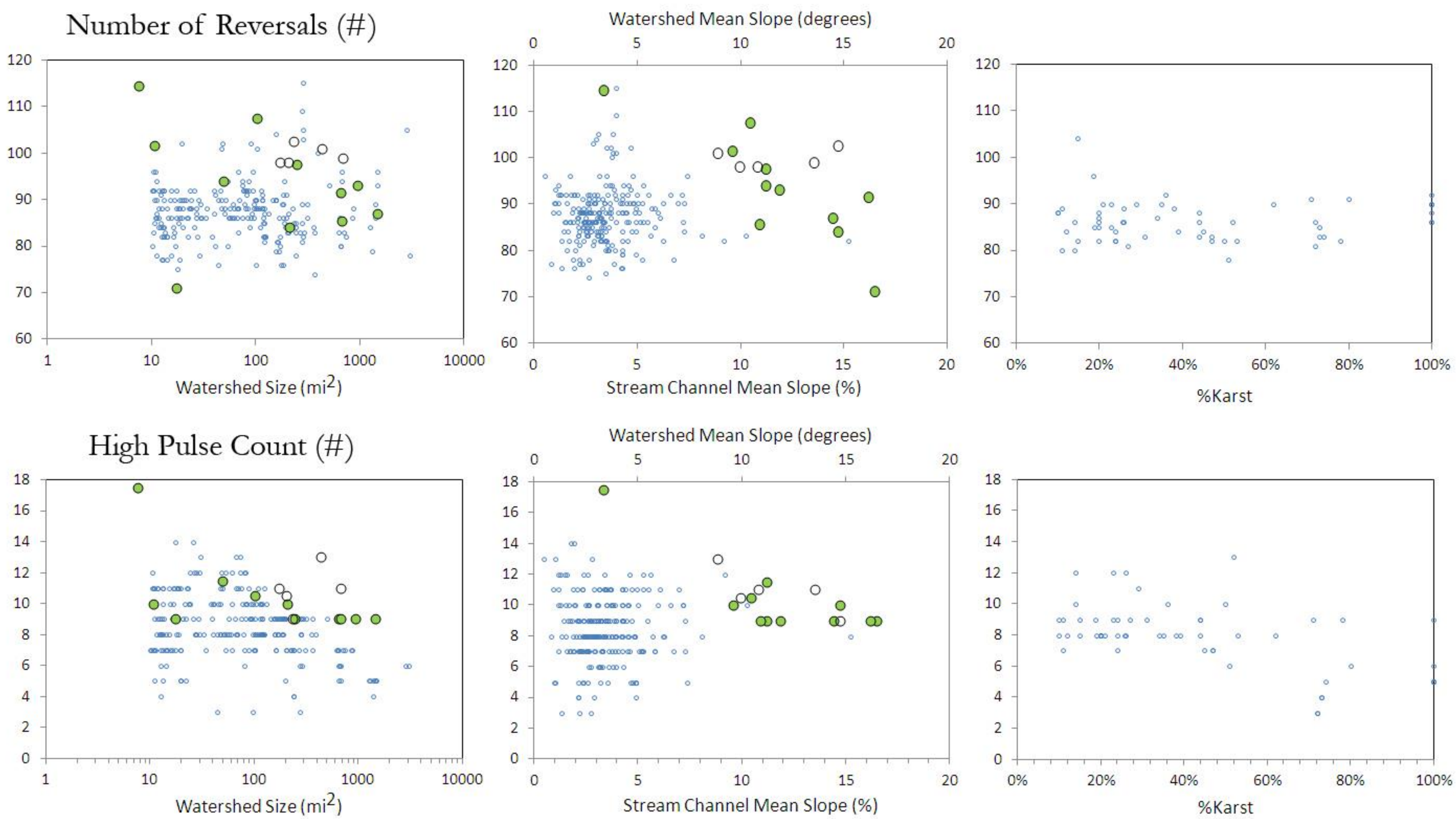


Figure 1 (cont.). Scatter plots of baseline (o), reference (●), and near-reference (O) values for number of reversals and high pulse count. Left panel, flow metric values versus watershed size; middle panel, baseline values versus stream channel mean slope and reference/near-reference values versus watershed mean slope; right panel, baseline values versus percent karst for watersheds with greater than 10% karst.

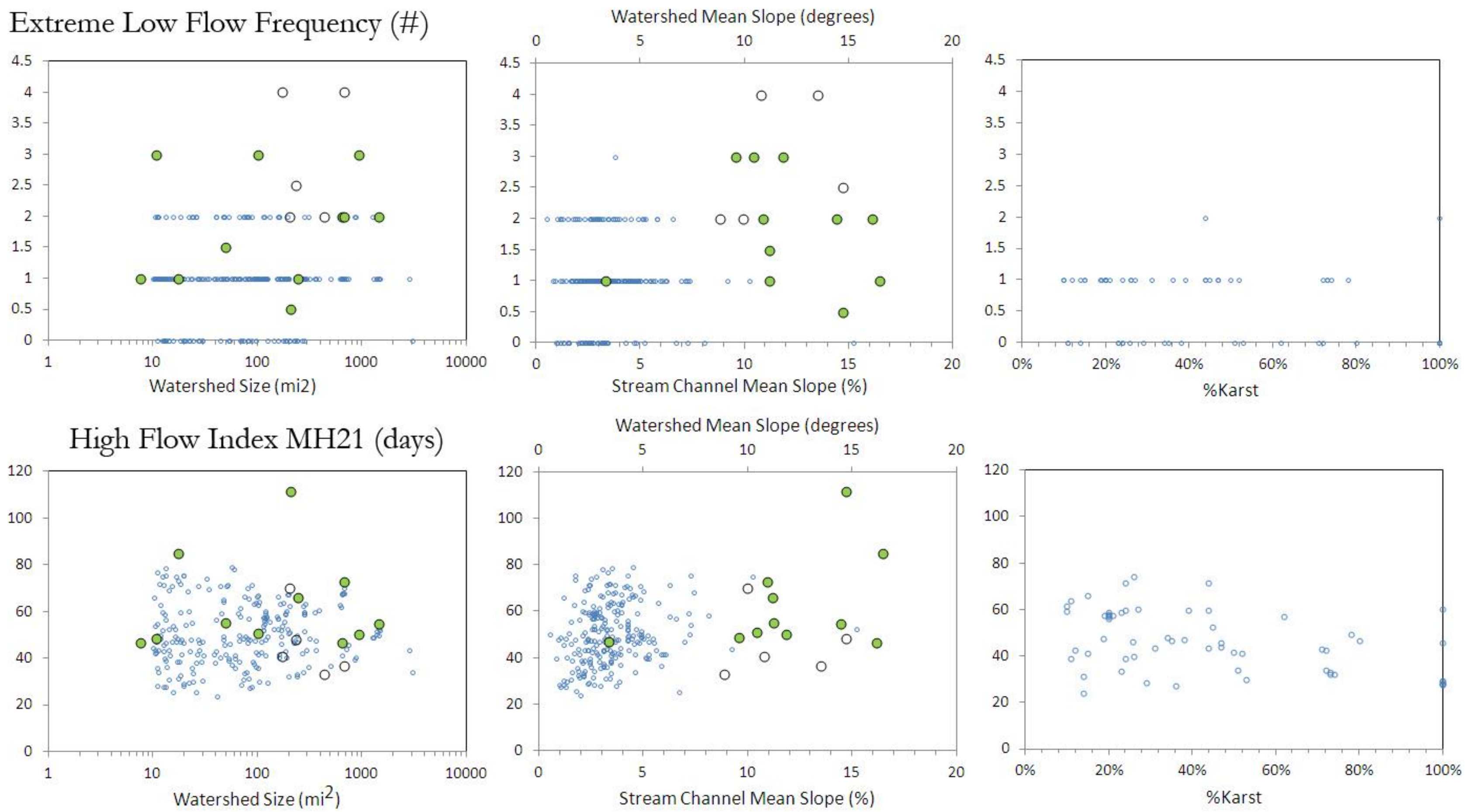


Figure 1 (cont.). Scatter plots of baseline (o), reference (●), and near-reference (O) values for extreme low flow frequency and high flow index MH21. Left panel, flow metric values versus watershed size; middle panel, baseline values versus stream channel mean slope and reference/near-reference values versus watershed mean slope; right panel, baseline values versus percent karst for watersheds with greater than 10% karst.

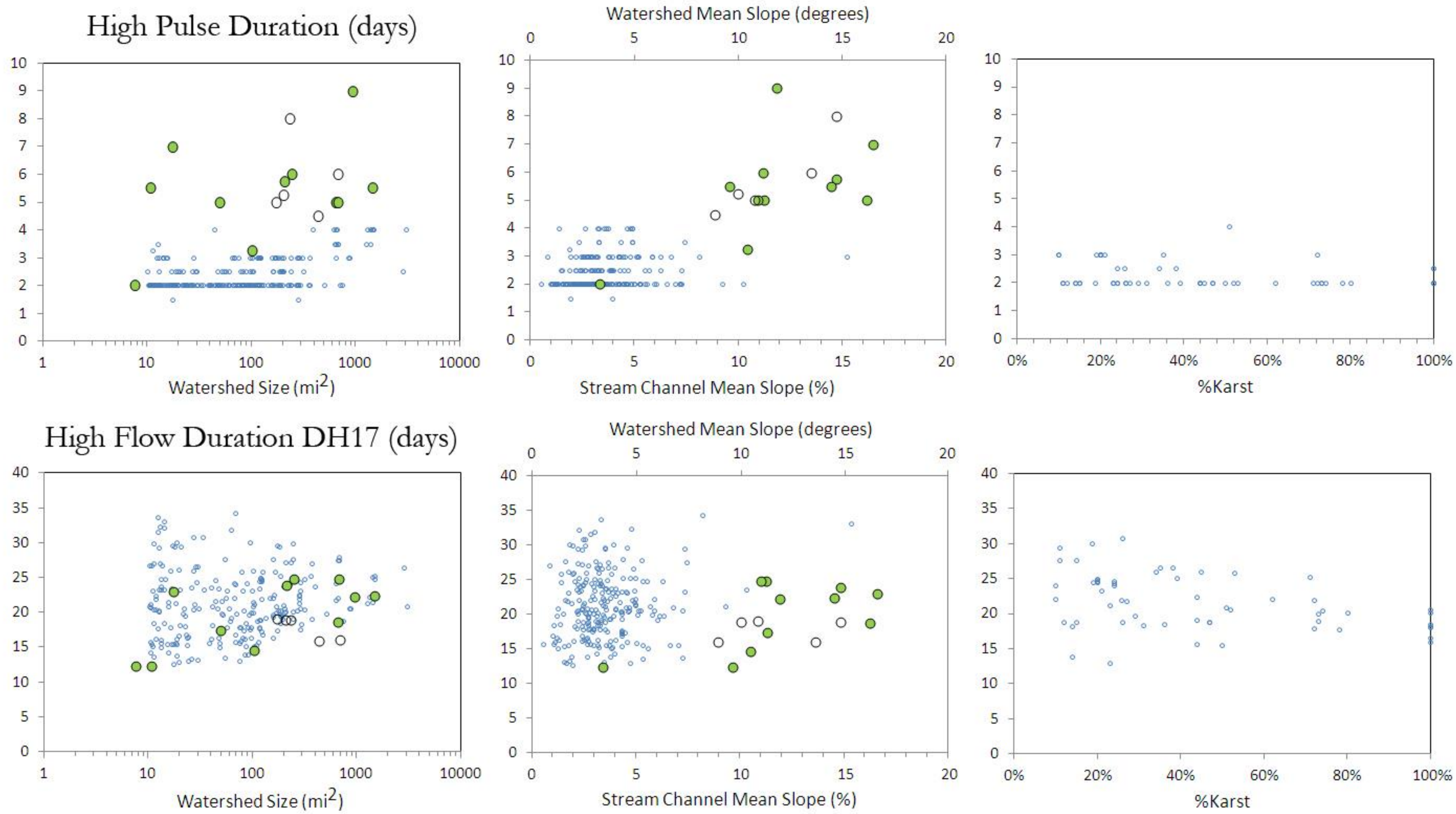


Figure 1 (cont.). Scatter plots of baseline (o), reference (●), and near-reference (O) values for high pulse duration and high flow duration DH17. Left panel, flow metric values versus watershed size; middle panel, baseline values versus stream channel mean slope and reference/near-reference values versus watershed mean slope; right panel, baseline values versus percent karst for watersheds with greater than 10% karst.

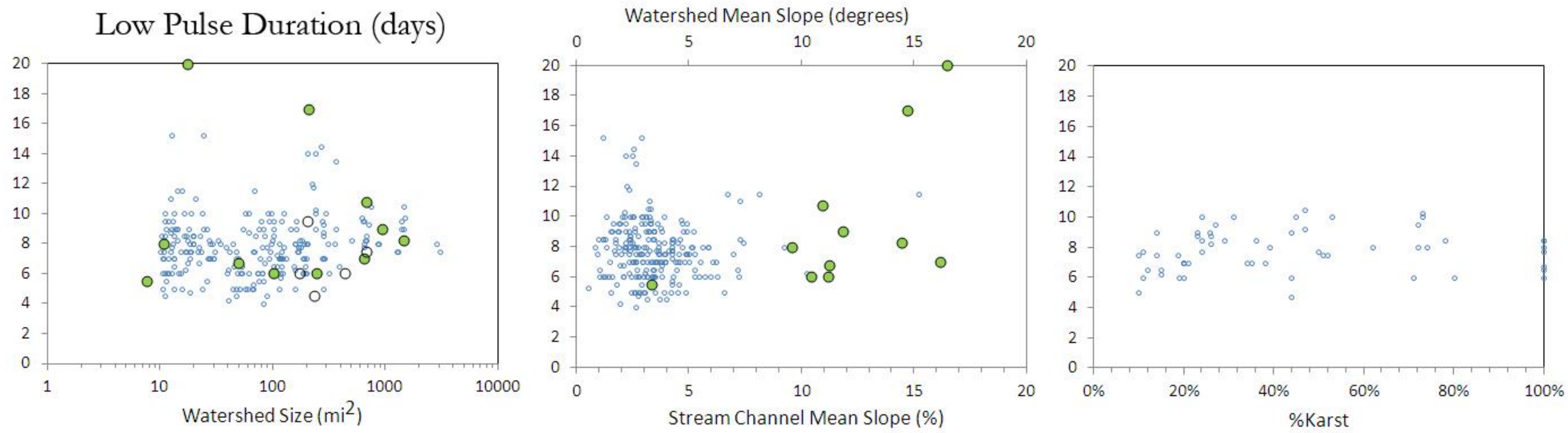


Figure 1 (cont.). Scatter plots of baseline (o), reference (●), and near-reference (O) values for low pulse duration. Left panel, flow metric values versus watershed size; middle panel, baseline values versus stream channel mean slope and reference/near-reference values versus watershed mean slope; right panel, baseline values versus percent karst for watersheds with greater than 10% karst.

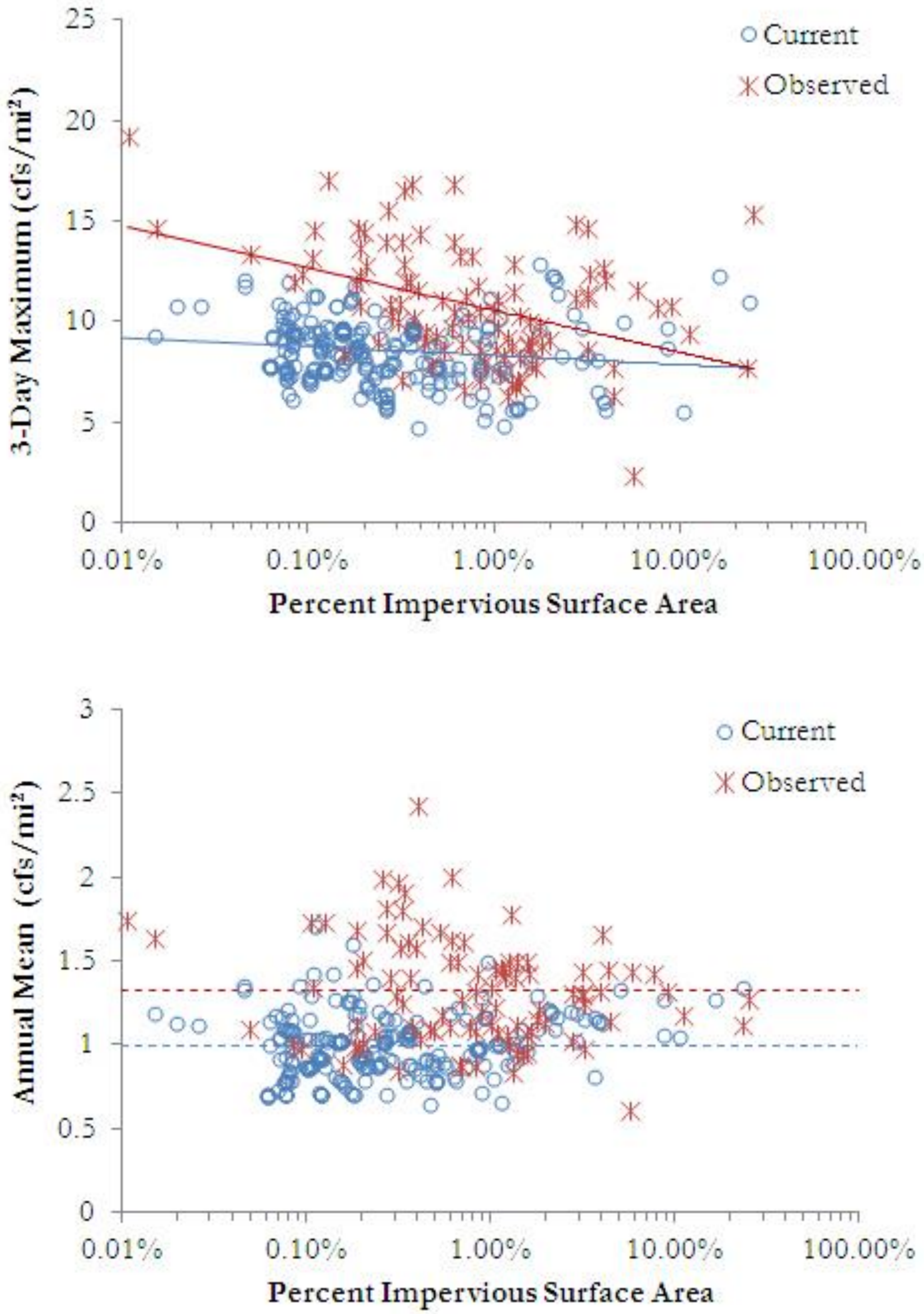


Figure 2. Current and observed values for 3-day maximum and annual mean versus percent impervious surface area. See Table 7 heading for details.

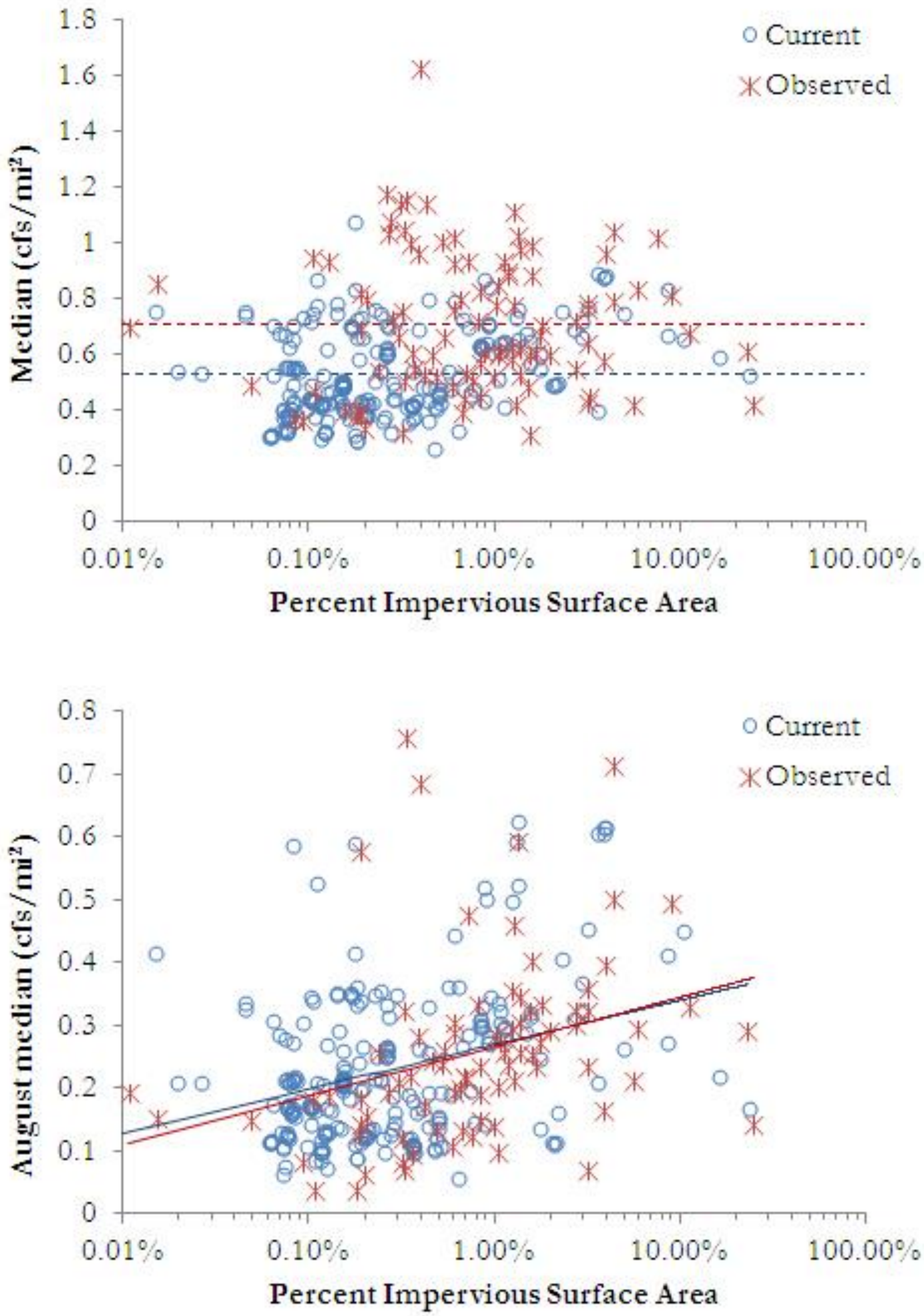
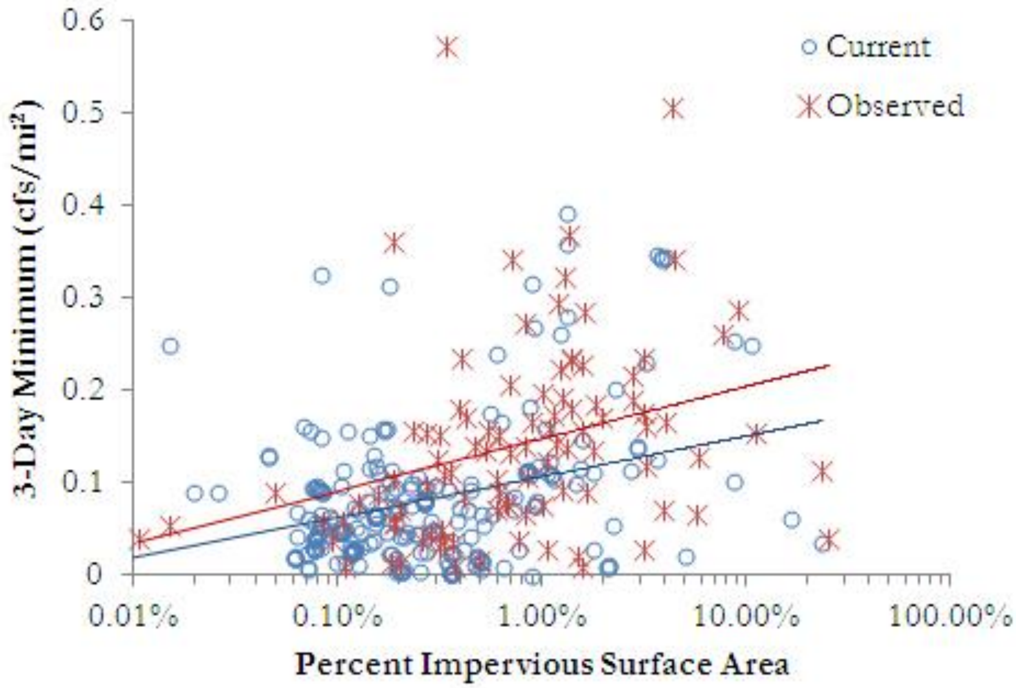


Figure 2 (cont.). Current and observed values for median and August median versus percent impervious surface area. See Table 7 heading for details.



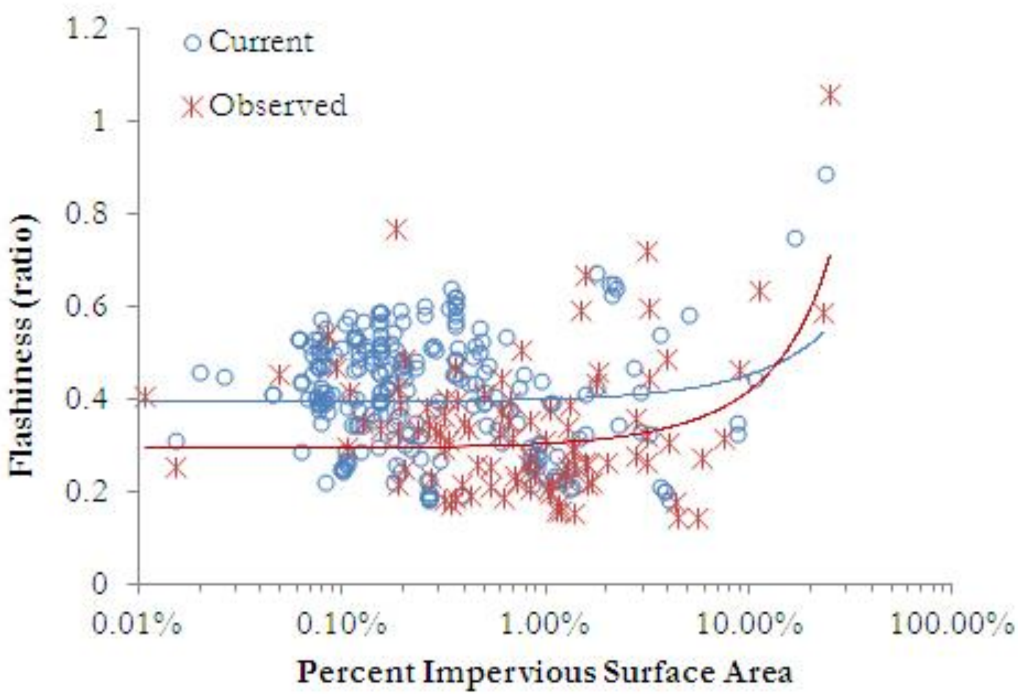


Figure 2 (cont.). Current and observed values for 3-day minimum and flashiness versus percent impervious surface area. See Table 7 heading for details.

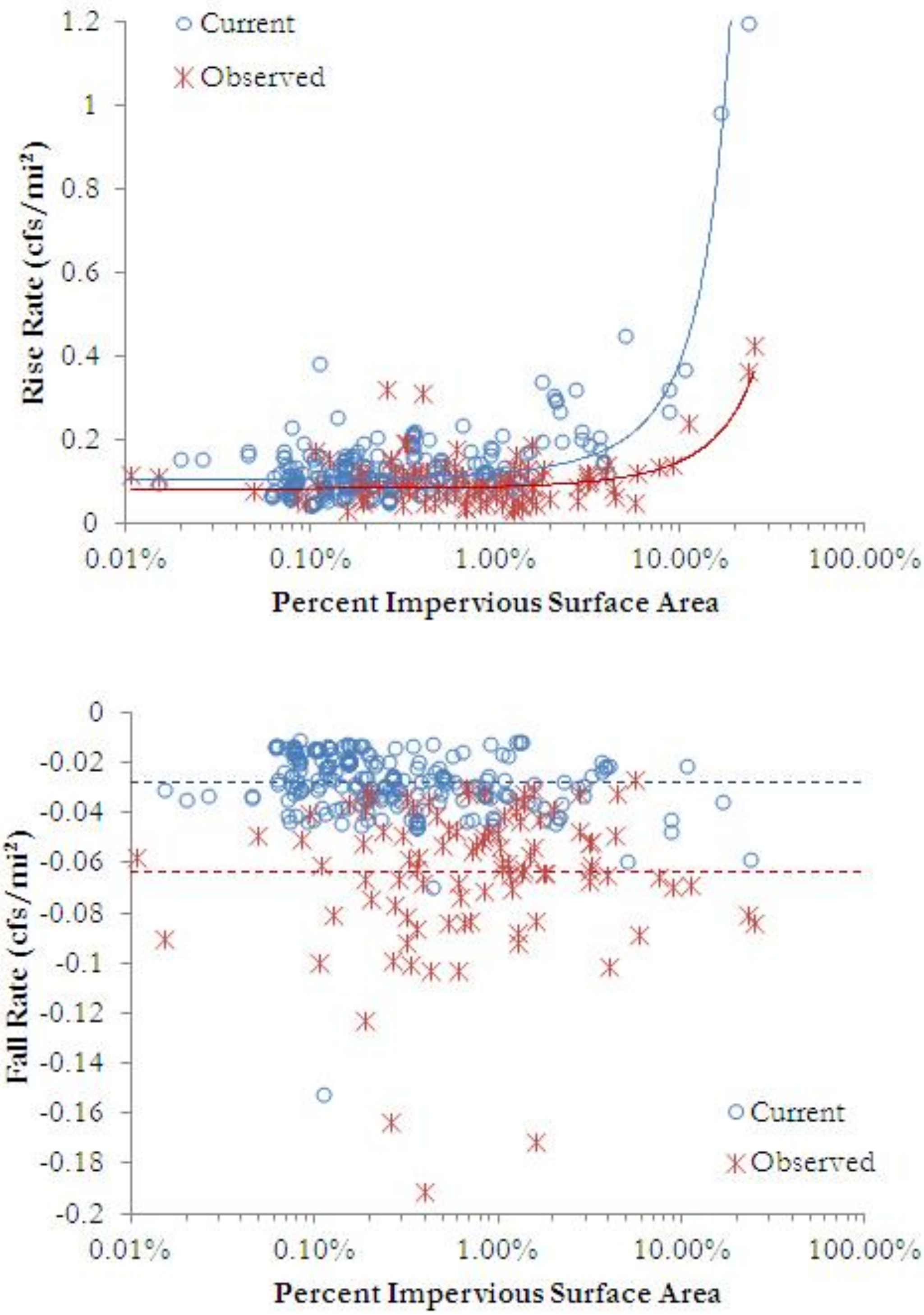
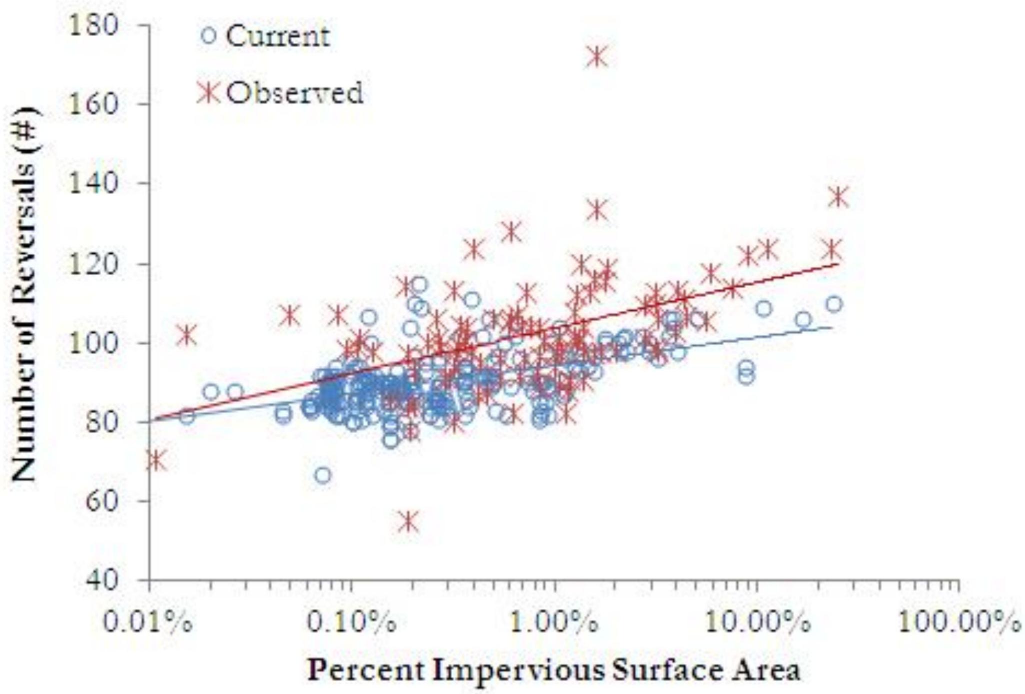


Figure 2 (cont.). Current and observed values for rise rate and fall rate versus percent impervious surface area. See Table 7 heading for details.



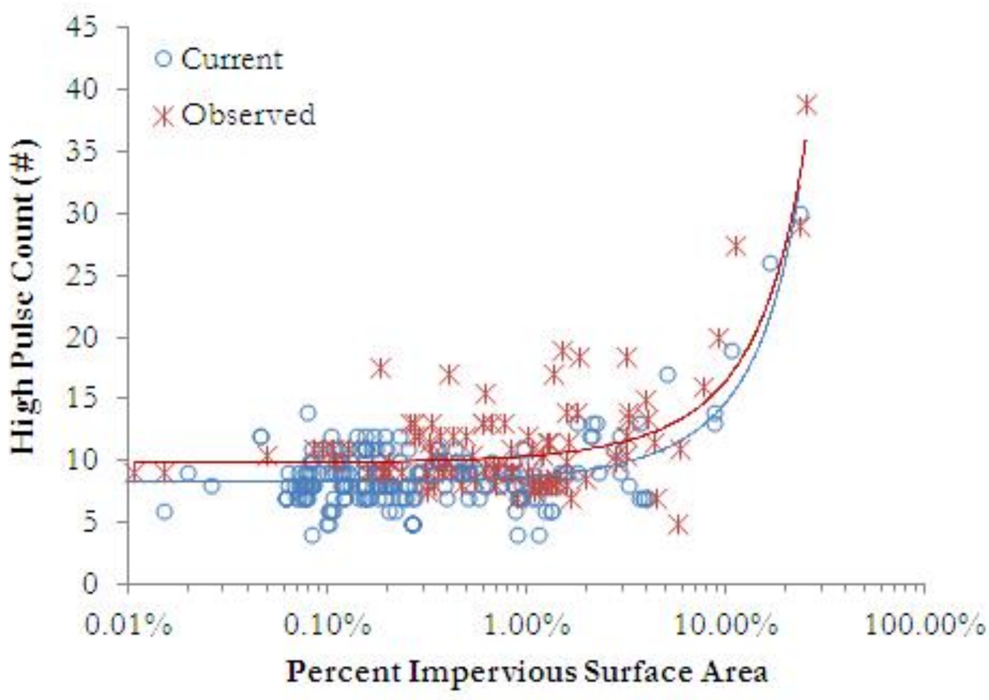


Figure 2 (cont.). Current and observed values for number of reversals and high pulse count versus percent impervious surface area. See Table 7 heading for details.

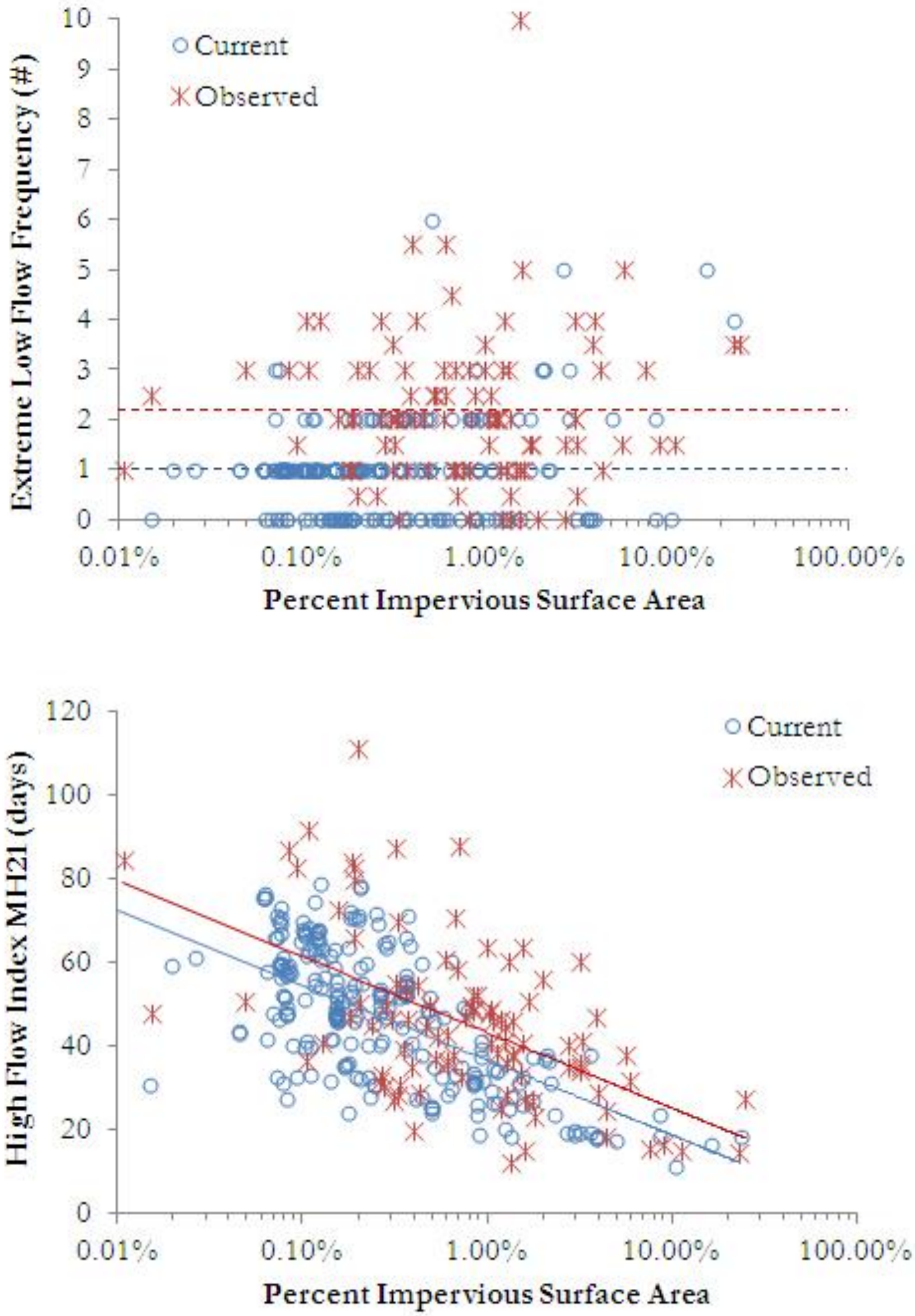
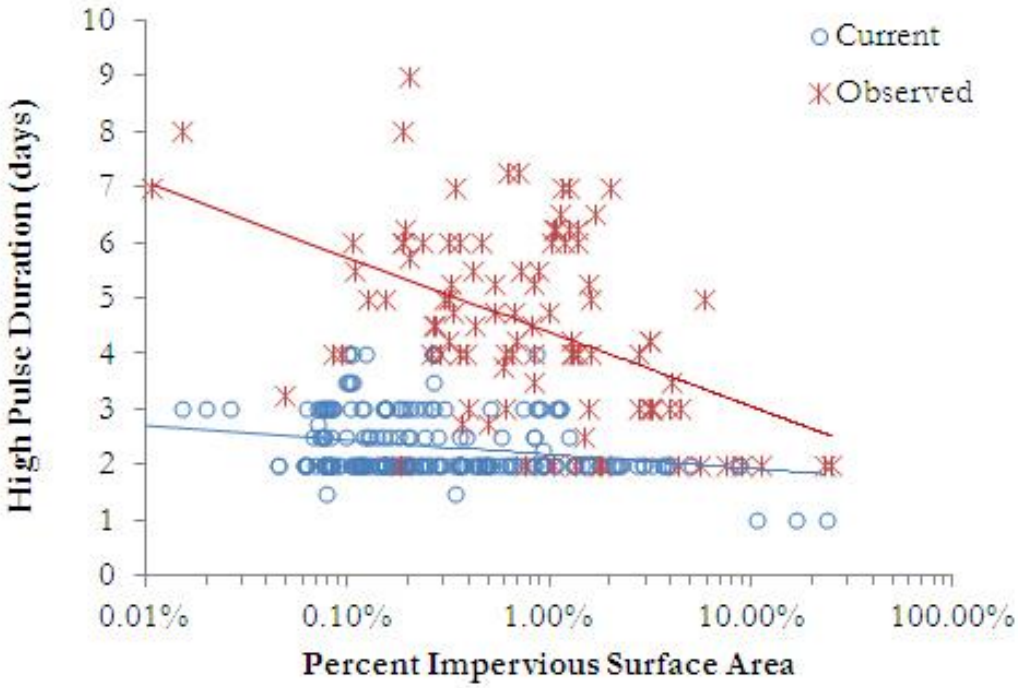


Figure 2 (cont.). Current and observed values for extreme low flow frequency and high flow index MH21 versus percent impervious surface area. See Table 7 heading for details.



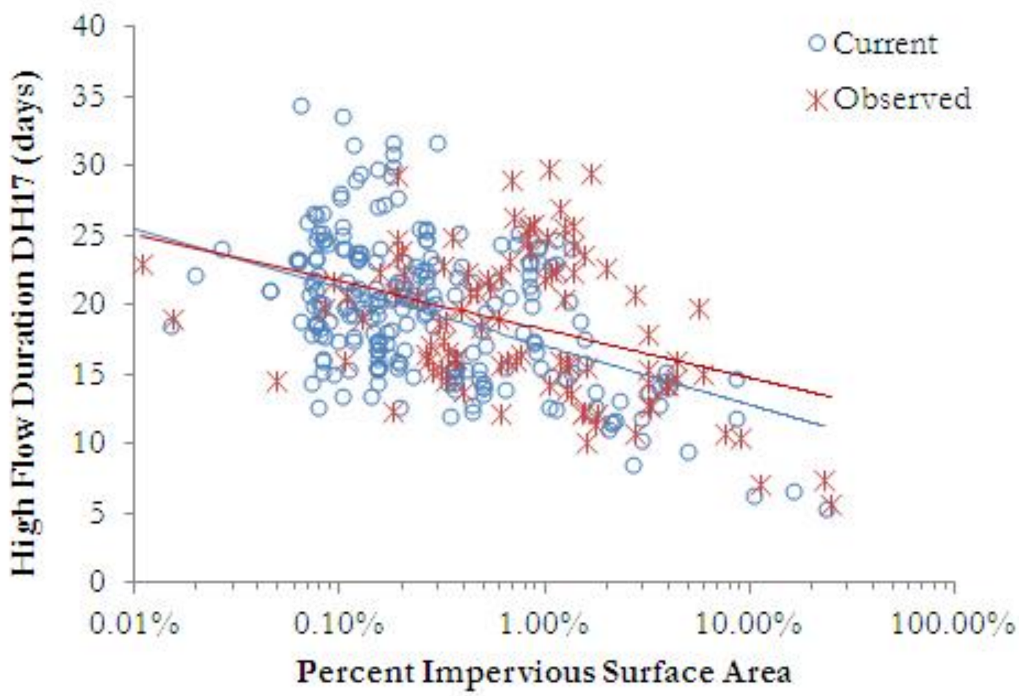


Figure 2 (cont.). Current and observed values for high pulse duration and high flow duration DH17 versus percent impervious surface area. See Table 7 heading for details.

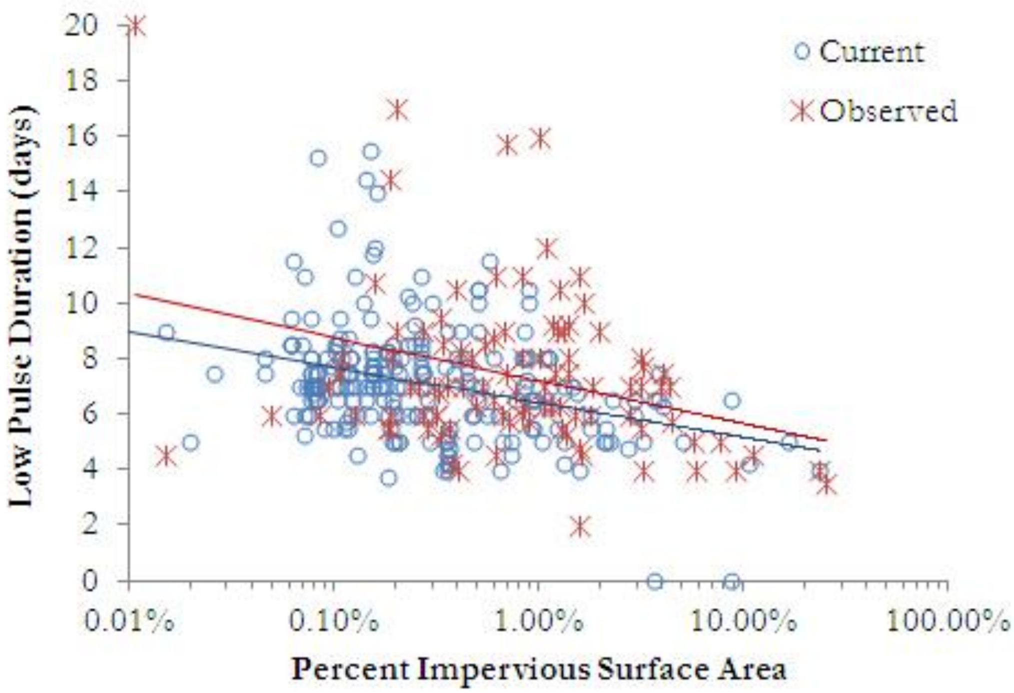


Figure 2 (cont.). Current and observed values for low pulse duration versus percent impervious surface area. See Table 7 heading for details.