Appendix D - Application of ELOHA

Contents

This appendix provides a more detailed description of the modification and application of the Ecological Limits of Hydrologic Alteration framework. This chapter was originally included as part of the main study report and followed section 5.1.

The disc directory contains only this document.

1. Compilation of Stream Data

The compilation and analysis of several types of stream monitoring data was a large and important project effort. These observed data consist of daily flow time series measured at USGS gages and biological, habitat, and water quality data collected at stream monitoring stations. The daily flows were used to validate baseline and current scenario model results and relate watershed land and water uses to flow conditions. Metrics calculated from the biological data were used as ecological response variables in developing FA-E relationships. Habitat and water quality data were used to investigate environmental factors potentially masking or confounding the FA-E relationships. Appendix C describes in detail the various stream monitoring datasets compiled for the project. The data are available on the disc accompanying this report.

USGS flow gages

One hundred and seventeen USGS gages are listed for the entire Potomac River basin (http://waterdata.usgs.gov/nwis). All but approximately 20 are located in the Middle Potomac study area. A careful review of the data identified 65 gages with nearly continuous data records for 1984 – 2005, the time period on which current and baseline model scenarios are executed. This time period was selected to correspond to the hydrologic modeling time period and is limited based on the availability of observed input time series required by the hydrologic model (e.g. meteorology). Fifty-four are calibration points for the Chesapeake Bay Watershed Model P5. Flow data for the 65 gages were downloaded from the USGS website. Watersheds upstream of the selected gages were delineated and their natural features and land and water uses quantified from graphical information system (GIS) layers for stream networks, elevation, ecoregion, geology, precipitation, land cover, withdrawals, impoundments, and discharges (data sources are available in Appendix C). Fifteen flow metrics capturing different portions of the hydrograph were calculated from the observed flow time series using Indicators of Hydrologic Alteration (IHA) version 7.0 software (TNC 2007) and the USGS Hydrologic Index Tool (HIT) program (Henriksen et al. 2006). Information about the hydrologic and biologic importance of these flow metrics can be found in both TNC (2007) and Apse et al. (2008). To increase sample size and statistical rigor of the analyses, the same fifteen flow metrics for 40 stream gages in the neighboring Susquehanna River basin were added to the Potomac dataset. The associated natural features and land/water uses for the Susquehanna watersheds and most of the flow metrics were provided by Michele DePhilip and Tara Moberg of the Pennsylvania chapter of TNC and Jennifer Hoffman of the Susquehanna River Basin Commission (SRBC). Watersheds in the combined Potomac-Susquehanna dataset varied in size from 2.7 mi^2 to 27,100 mi^2 and covered a broad range of land and water uses. This dataset was used to check flow metrics calculated from model output and corroborate the relationships found between flow metrics and watershed characteristics in the modeled data (Section 3).

Biological data

The technical team reviewed the available macroinvertebrate and fish data for their usefulness as the biological element in FA-E relationships. At the time, both types of data were being acquired from federal, state, and local sources and assembled into relational database structures by ICPRB as part of an effort supported by the Chesapeake Bay Program (CBP). The macroinvertebrate database was significantly larger and more complete than the fish database due to the fact that all of the state agencies in the region collect macroinvertebrate samples as part of their routine monitoring programs. The compiled database consisted of 5,410 sampling events collected with roughly comparable field methods at 3,310 stations throughout the Chesapeake region between 1986 and

2009. Comparable family-level metrics could be generated from all of the contributed datasets and evaluations of macroinvertebrate status could be done in a consistent manner across the entire Middle Potomac region. Furthermore, family-level metrics responsive to anthropogenic disturbance had been combined in a Chesapeake basin-wide index of biotic integrity ("Chessie BIBI") for the purpose of evaluating stream health (Foreman et al. 2008; Buchanan et al. 2011). Finally, CBP had implemented an annual data call to acquire and incorporate new data and data sources. In contrast, the database of available fish data was smaller with approximately 2,600 sampling events collected at 1,667 stations. Only records from the state of Pennsylvania, Montgomery County in Maryland, and Fairfax County in Virginia were included at the time the database was reviewed. Programmatic differences in gear type and protocols made it difficult to find comparable fish data across the Middle Potomac region and little work had been done to develop a multi-metric index of fish community status. The decision was made to use macroinvertebrates as biological response variables because of the larger size, better integration, and broader coverage of the macroinvertebrate database and the comparability and utility of the family-level metrics.

Stream macroinvertebrates are a diverse group of organisms with many morphological, behavioral, and feeding adaptations for life in flowing waters. They consume detritus, algae, bacteria, and microscopic animals, and are food for fish and birds. They have a broad range of tolerances and sensitivities to different stressors and anthropogenic pollutants, and fill many ecological niches. This makes them ideal as indicators of the health of most streams and rivers.

Certain constraints come with using macroinvertebrates as the ecological variable in FA-E relationships. Aquatic macroinvertebrates are a diverse group representing numerous phylogenic taxa, but most communities in free-flowing waters are dominated by insects. Insect life cycles are relatively short—on the order of weeks to a few years—and their responses to extreme but infrequent events (extreme droughts, extreme floods) occur in the year of the event and then fade as successive generations recover. An appropriate suite of flow metrics for macroinvertebrates may be those that reflect flow alteration occurring on a regular basis. Macroinvertebrates also are sensitive to environmental stressors that do not alter flow *per se* but that can be associated with flow and can confound the effects of flow alteration, such as water chemistry (pH, conductivity, dissolved oxygen, nutrients) and habitat condition (loss of riparian zone vegetation, sedimentation, channelization). Removing the influence of these factors, many of which are anthropogenic in origin, is an important step in isolating and quantifying the true impacts of flow alteration.

Habitat and water quality data

Another feature of the macroinvertebrate database influenced the decision to use this biological group. Several habitat and water quality parameters are measured concurrently with the macroinvertebrate samples and are included in the database. Habitat and water quality data are crucial for independently identifying high quality, or "reference" sites. Some physical habitat parameters are directly impacted by flows (e.g., bank stability). Connections between flow, habitat, and macroinvertebrate status could be explored in the future with this dataset, and would facilitate development of environmental flow standards where habitat and water quality data are available but biological data are scarce.

2. Hydrologic Modeling

The ELOHA framework is predicated on the ability to couple flow and biological data. In the Potomac River basin, biological monitoring data is most often available for headwaters, streams, and small and medium sized rivers. Conversely, flow data is most often collected at USGS gages on medium and large sized rivers. To this end, hydrologic modeling efforts were undertaken to estimate flows at ungaged locations corresponding to select biological monitoring points. The hydrologic modeling efforts included the Middle Potomac study area as well as the North Branch of the Potomac River, an upstream tributary that is required for efficient modeling of the study area (Figure 1). Detailed descriptions of key steps in the modeling approach are provided in Appendix E.

Biological monitoring locations in the Middle Potomac study area were selected from the CBP macroinvertebrate database. Of the 3,310 biological sampling locations in the database, 869 sites were selected for the MPRWA that (1) represented a range of hydrologic alteration, (2) represented a range of watershed sizes, (3) included broad spatial distribution across the study area, (4) were located within 200 feet of a National Hydrography Dataset (NHD) stream (deviation from the stream is indicative of erroneous monitoring point location information, or less commonly, spatial accuracy problems in the NHD dataset), and (5) were sampled between 2000 and 2008 to correspond to the current hydrologic model scenario (the current scenario is defined later in this document). Watersheds that drain to the 869 biological monitoring points were delineated with two methods: selection of the NHDPlus catchments associated with biological monitoring points in ArcGIS and implementation of the Utah State University Multi Watershed Delineation Tool. Three hydrologic modeling tools were evaluated for potential use in the MPRWA including the US

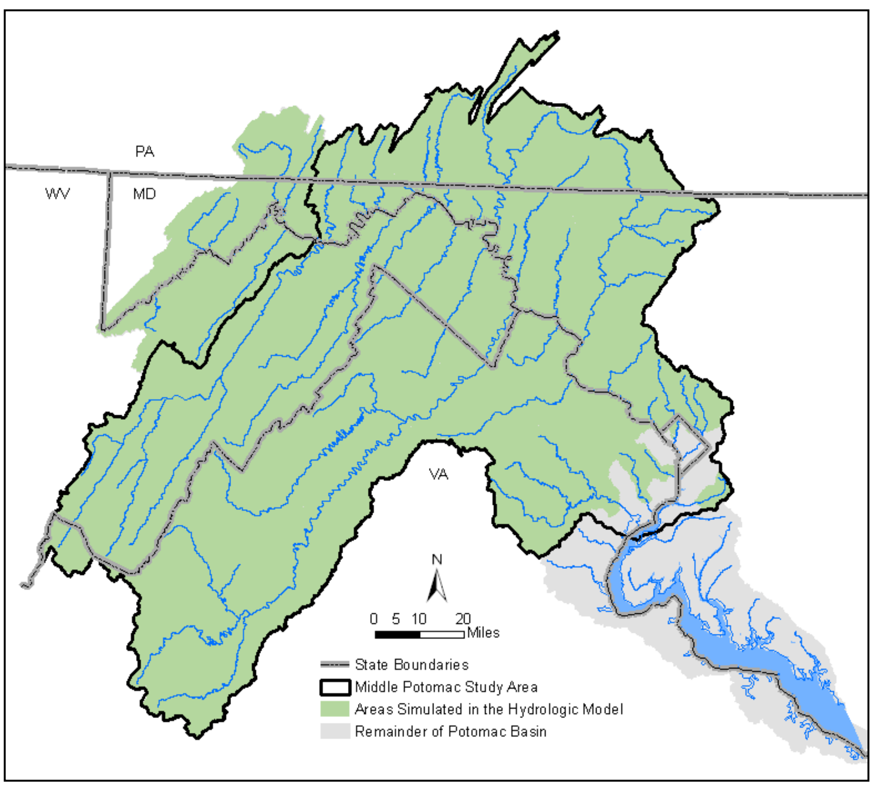

Figure 1. Geographic areas simulated in the hydrologic model in relation to the Middle Potomac study area.

Geological Survey (USGS) Sustainable Yield Estimator (SYE), the Chesapeake Bay Program's HSPF model, and the Virginia Department of Environmental Quality's (VADEQ) Online Object Oriented Meta Model (WOOOMM) (Appendix E, Application of Modeling Tools). SYE was evaluated as it has been used in similar projects (Waldron and Archfield 2006; Archfield et al. 2010). SYE is capable of simulating baseline but not future scenarios, a desired capability for the Potomac project. The Potomac River Basin, being a part of the Chesapeake Bay Watershed, is fortunate to have a fully developed HSPF model through the Chesapeake Bay Program (Linker et al. 1999). The HSPF model has undergone collaboration, development, peer review, research, revision, and implementation over the past 30 years. It was designed to understand freshwater inflows to the Chesapeake Bay as well as the nutrients and sediments transported by the flows. Phase 5.2 of the HSPF model was utilized because it was the most recent, complete version available at the time of analysis. The HSPF model is capable of simulating current, baseline, and future scenario flows. The model is divided into more than 2,000 simulated segments across the 64,000 square mile Chesapeake Bay Watershed (Chesapeake Bay Program 2010). The spatial resolution of the modeled segments averages 89 square miles in the study area. This is a much coarser spatial resolution than needed to simulate the range of watershed sizes in this study – from small creeks to large rivers.

WOOOMM complements the capabilities of the HSPF model by using the latter's land simulation coupled with a USGS channel morphology module and a WOOOMM channel routing routine to estimate flows at locations of interest (Kudlas 2009). In combination, the HSPF model and the WOOOMM enable simulation of streamflows at locations in the study area corresponding to biological monitoring points. The result was the selection of the HSPF model,[1] in combination with the WOOOMM routing module.[2] The channel routing routine utilizes Manning's equation to simulate flow through a trapezoidal channel utilizing input parameters such as channel length, side-slope ratio, Manning's roughness coefficient, base width, and slope. The channel morphology module estimates the channel properties for input into the channel routing routine such as the side-slope ratio, base width, and Manning's roughness coefficient using regression equations developed by the USGS. These equations describe the relationship of the channel properties to drainage area and physiographic province.[3] The strength of combining HSPF with WOOOMM is the ability to effectively simulate flows at any selected stream location, enabling estimation of flows at selected biological monitoring points in the study area. Further details about the WOOOMM module are available in Appendix E (WOOOMM Inputs).

Two enhancements were made to the HSPF model prior to use in this study. The first enhancement was re-segmentation at "significant" impoundments (Appendix E, Resegmentation at Impoundments). Significant, in this case, indicated that: (1) the normal storage capacity of the impoundment was greater than 10 percent of the mean annual flow volume OR the impoundment was used for hydroelectric purposes, AND (2) biological monitoring points were located upstream and downstream of the impoundment. This effort was conducted to eliminate the influence of major sources of hydrologic alteration within a model component. Of the 481 impoundments found in the National Inventory of Dams in the Potomac River basin, 12 impoundments were

---

[1] The HSPF model is freely available from the Chesapeake Community Modeling Program http://ches.communitymodeling.org/models/CBPhase5/index.php). Model documentation is also provided through the program.

[2] The WOOOMM is an online tool accessible at http://deq1.bse.vt.edu/wooomm/login.php. A Wiki describing the tool and its capabilities are online at http://sifn.bse.vt.edu/sifnwiki/index.php/Creating_Flow-Ecology_Relationships.

[3] Personal communication, R. Burgholzer (VADEQ) in memo "modeling an unaltered flow regime" (2010).

selected for inclusion in the model. The HSPF model was re-segmented so that these 12 impoundments were located at river segment outlets. The dam operations such as pass-by requirements and whitewater releases were also included in the model where information could be obtained. The result was a revised HSPF model with a total of 16 simulated impoundments, 4 of which were previously included. The second enhancement was implementation of a nonlinear groundwater recession algorithm (Schultz et al. In review). The HSPF model is typically utilized to understand nutrient and sediment transport during high flows. As a result, low flows are often under-simulated. To improve the simulation of low flows, the traditional linear groundwater recession algorithm was replaced with a nonlinear algorithm. The nonlinear algorithm was included in the subsequent CBP Phase 5.3 version of the HSPF model.

Hydrologic model calibration

The HSPF model was calibrated using the CBP auto-calibration routine (USEPA 2010). The HSPF model was then evaluated for simulation efficiency using the Nash Sutcliffe efficiency (NSE) coefficient (Nash and Sutcliffe 1970) and the coefficient of determination, or R^2. The NSE compares simulated and observed flows to quantify the variation from the 1:1 line. Values range from negative infinity to one, with one indicating that simulating values are identical to observed values. Decreasing NSE values are an indication of poorer model efficiency, with negative values indicating that it is more reliable to use the observed mean value than the model results. The NSE is considered by some to be the most commonly used metric to evaluate model efficiency (Gassman et al. 2007). R^2 values ranged from 0.39 to 0.82 at the 43 model calibration locations. Daily NSE values at the same locations ranged from 0.33 to 0.82 (Figure 2). The calibration locations are shown in Figure 3. Consistent with a "weight of evidence" model evaluation approach (Lumb et al. 1994; Donigian 2002), the results of the calibration represent an acceptable range of model error and show an overall modest improvement in the HSPF model performance.

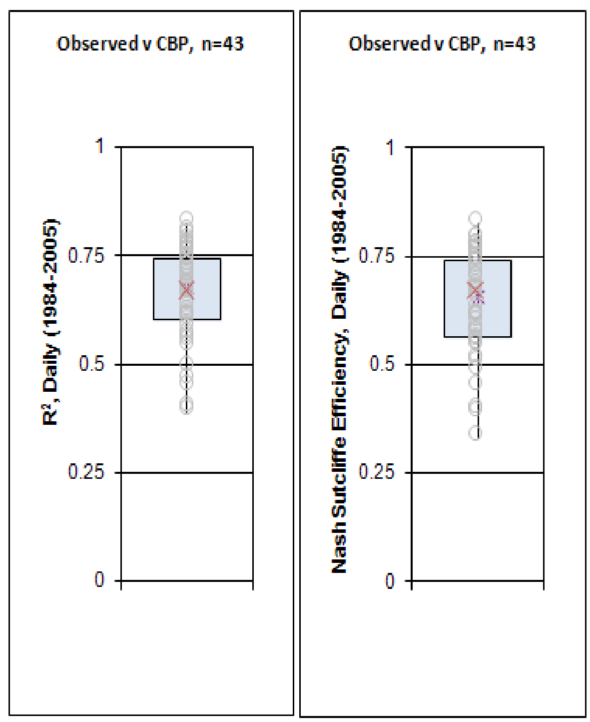

Figure 2. Distribution of daily Nash Sutcliffe Efficiency (NSE) coefficient and coefficient of determination (R^2) values for 43 model calibration locations.

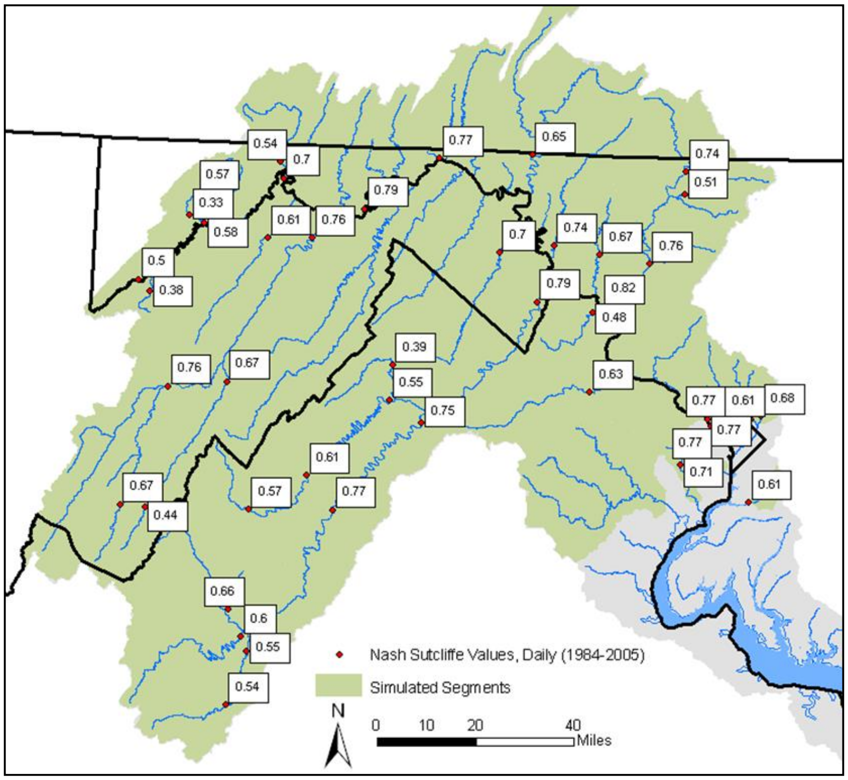

Figure 3. Daily NSE values at model calibration locations.

Hydrologic model implementation

Once the HSPF model was calibrated utilizing the approach described in the model calibration methodology (Chesapeake Bay Program 2010), the watersheds draining to biological monitoring points that are the subject of this study's stream flow needs analysis were established in the WOOOMM routing module (Appendix E, Watershed Delineation). The watersheds represent a sub-set of the originally delineated 869 watersheds. Specifically, a total of 28 duplicate watersheds were identified (Appendix E) and removed. Duplicates occurred when biological sample locations were so close their delineated watersheds were effectively the same for modeling purposes. An additional 94 delineated watersheds needed for simulating flows in nested or adjacent watersheds were also removed because there were no biological data available to couple with them. (Simulated baseline and current flows for these 94 "assisting" watersheds are included in the Master spreadsheet in Appendix C). The remaining 747 watersheds constitute the "ELOHA" analysis dataset. Each is associated with one or more biological samples.

Two scenarios were developed and executed within the WOOOMM routing module. The baseline scenario simulated the hydrologic conditions given nominal anthropogenic influence. All withdrawals, discharges, and impoundments were turned off for the baseline scenario. Land uses were returned to mostly forested conditions (greater than 78 percent forest) and impervious surface area was minimized (less than 0.35 percent). These two land use thresholds were initially identified in the large river environmental flow needs report (Appendix A). The thresholds were a result of a recursive partitioning analysis, called "Category and Regression Tree Analysis" (CART), performed on the Potomac-Susquehanna dataset. CART is similar to cluster analysis and indicates breakpoints where the clusters divide (Appendix E, Potomac-Susquehanna CART analysis). A complete description of how these thresholds for baseline land use conditions were applied in modeling baseline flows is available in Appendix E (Baseline Landuse and Baseline Scenario). The second scenario executed in the WOOOMM was a current conditions scenario. The current scenario was a snapshot of current watershed conditions, including withdrawals, discharges, land uses, and impoundments, run over a 21-year period (1984-2005) of meteorological conditions (the common time period of necessary observed model input time series). A comparison of baseline and current scenario streamflows provided an indication of how much hydrologic alteration is currently found in watersheds draining to the biological monitoring points.

Five future scenarios were also developed and run in the HSPF model at the river segment scale. The purpose was to understand the hydrologic effects of various possible future (2030) conditions. Future scenarios included DP1, DP2, power, climate change, and hot and dry, and are briefly described in 4.1. The scenarios included combinations of changes to land use, population in terms of increases in withdrawals and discharges, per capita water use, temperature, evapotranspiration, and precipitation. A detailed description of the modeling efforts for the future scenarios is available in Appendix E (Future Scenario).

Each of the seven model runs (current, baseline, DP1, DP2, power, climate change, and hot and dry) resulted in a 21-year daily flow time series. The time series were utilized to calculate a suite of 256 hydrologic metrics representing various portions of the hydrograph and including all flow levels (high, medium, and low flows). Metrics were calculated utilizing IHA (TNC 2007), EPA's DFLOW (Rossman 1990), and HIT (Henriksen 2006) as well as several metric calculations developed by ICPRB. The number of flow metrics was reduced through a process of identifying those that are responsive to alteration, not correlated, most efficiently modeled, easily understood, and correlated with biological health (Figure 4). Metrics were also selected to ensure that all parts of the hydrograph were represented. Flow metrics that best meet the screening criteria up to this point are listed in Table 1. Subsequent testing and analysis refined and expanded this list to finalize the list of flow metrics for which FA-E relationships were developed and spatial analyses were conducted.

Hydrologic model uncertainty

"How well does the model simulate stream flow?" This question lies at the heart of decisions about when and how to use simulated stream flows in watershed management. Simulated flows are credible when they accurately represent observed flows across a range of environmental conditions. The Potomac and its neighbor river basin the Susquehanna are fortunate to have roughly 100 USGS flow gages in continuous operation between 1984 and 2005. Their watersheds range from comparatively undisturbed to heavily urban and/or agricultural. Responses of flow metrics to different land and water uses in these observed watersheds provide a standard against which to test flow metrics calculated from simulated flows.

There are numerous sources of uncertainty in the model results including, but not limited to, mathematical errors, errors in observed data utilized during the modeling process, and limited knowledge of the system being modeled (Rode and Suhr 2006). Evidence of this uncertainty was present in the MPRWA in several areas. When comparing flow metrics calculated from observed and simulated datasets, the differences provided an indication of the magnitude of model uncertainty (Figure 5). Some differences were expected between simulated and observed values as the model was a simplified representation of a complex reality. Overall, the model results were shown to adequately represent reality in response to changes in watershed characteristics such as land and water uses for the purposes of this project (Appendix F).

During an investigation of the behavior of simulated flow metrics, a sub-set of watersheds were identified as having conspicuously different flow metric values. The sub-set of watersheds were identified as either being located in the karst regions of the basin or in modeled land segments containing karst (Figure 6). This was an indication that the hydrology of karst watersheds was different than non-karst regions, the watershed model had difficulties simulating the effects of karst geology, or a combination of the two. Unfortunately, sufficient observed flow data from karst

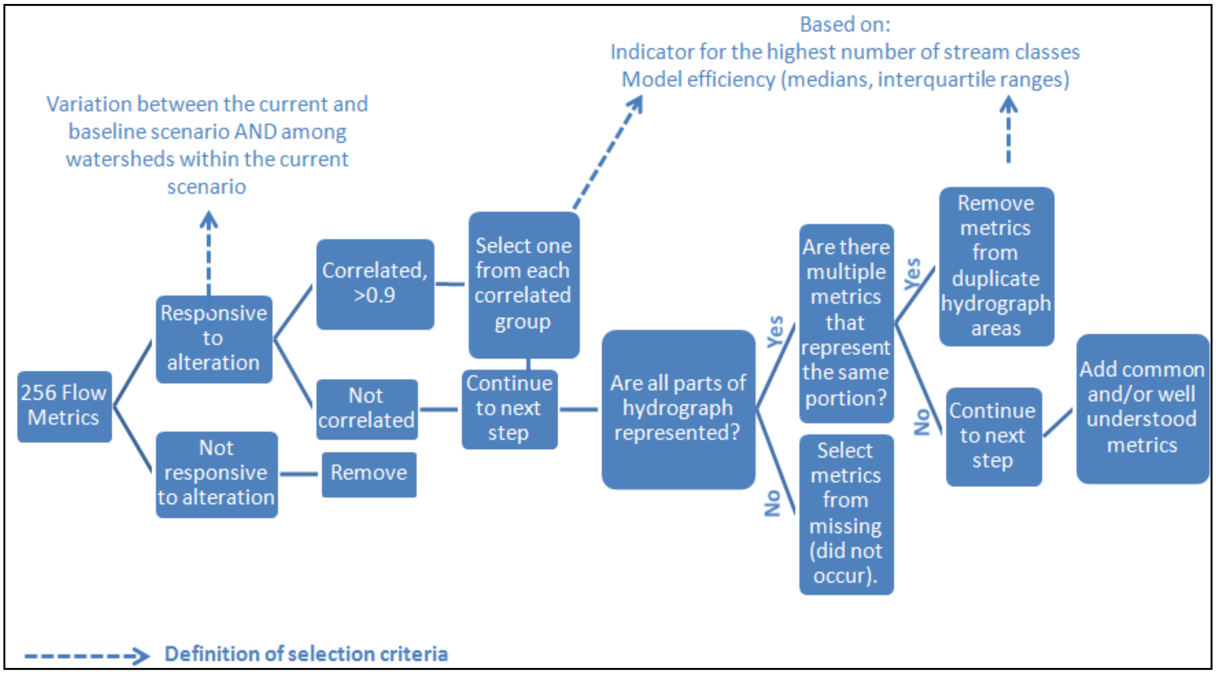

Figure 4. Decision-tree for selection of flow metrics.

Table 1. Flow metrics that best meet the criteria used in the Middle Potomac study to evaluate model simulations of flow time series.

Statistic (units)	Source	Flow Range	Characteristic
3 day maximum (cfs/mi^2)	IHA	High	Magnitude
High flow duration (days)	HIT	High	Duration
High pulse count (#/year)	IHA	High	Frequency
Flashiness (unitless)	ICPRB	All	Rate-of-Change
August median (cfs/mi^2)	IHA	Mid	Magnitude
7Q10 (cfs)	DFLOW	Low	Magnitude
Seasonal Q85 (Maryland Method) (cfs/mi^2)	ICPRB	Low	Magnitude
Low pulse duration (days/year)	IHA	Low	Duration
Low pulse count (#/year)	IHA	Low	Frequency

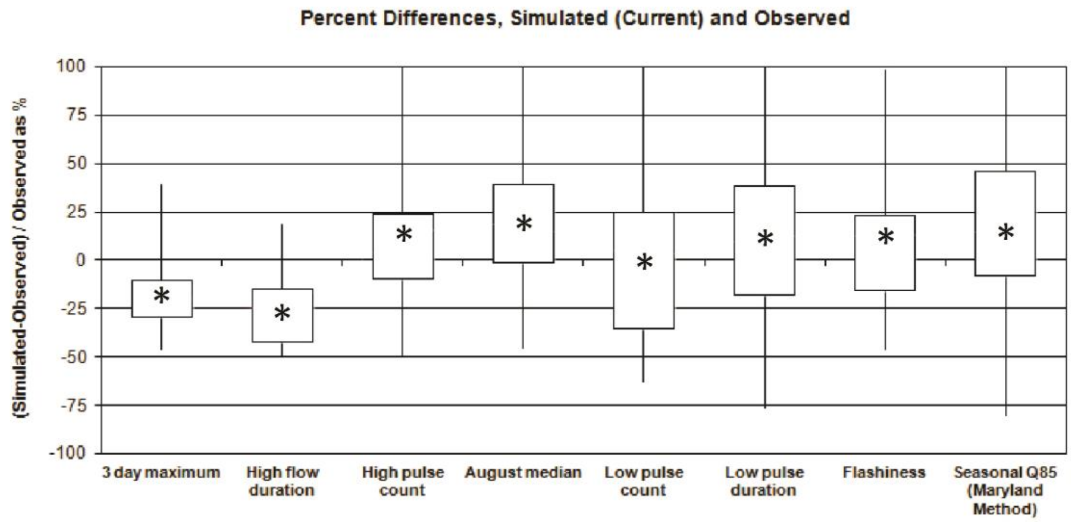

Figure 5. Comparison of observed flow metrics.
Calculated at 31 model calibration locations and HSPF simulated, current scenario flow metrics at the same locations.

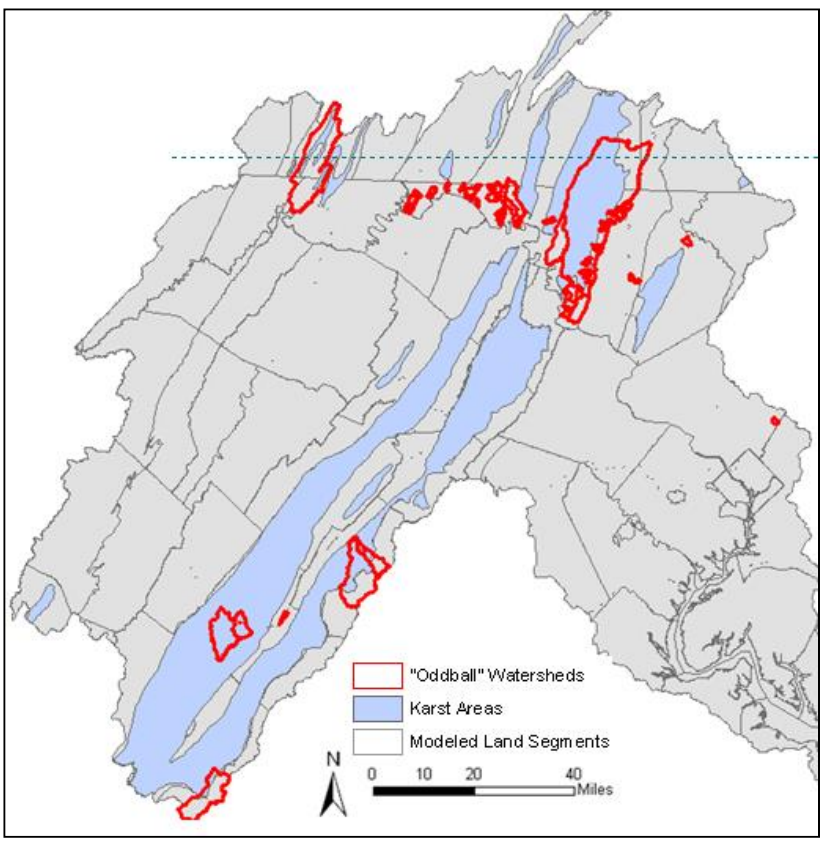

Figure 6. Location of watersheds with one or more flow metrics that behave differently than those in the observed dataset.
These watersheds are primarily located in regions of karst geology or in modeled land segments that contain karst geology.

watersheds was not available for comparison to determine the extent to which these conspicuous flow metrics were a result of model uncertainty or physical differences in the behavior of karst systems. Similarly, the availability of observed data to test model efficiency in small watersheds was limited. Most USGS gages were located in watersheds greater than 10 mi^2; however, roughly two thirds of the ELOHA watersheds were smaller than this threshold, with a median watershed size of just over 3.5 mi^2. Additional observed hydrologic data on smaller watersheds would enhance understanding of the extent to which these smaller watersheds were being adequately simulated.

Spatial distribution and the scale of the model inputs, particularly meteorological inputs, compared to the scale of modeled watersheds is also a potential source of model uncertainty. Meteorological inputs to the HSPF model were developed by the CBP from regional observation stations (USEPA 2010). Hourly meteorological input time series for each land segment in the Potomac basin study area were developed using 52 stations with daily data and 8 stations with hourly data (Figure 7). These meteorological data drive the hydrology of the HSPF modeled watersheds, and subsequently the WOOOMM routing module for the ELOHA watersheds. The meteorology in many of the

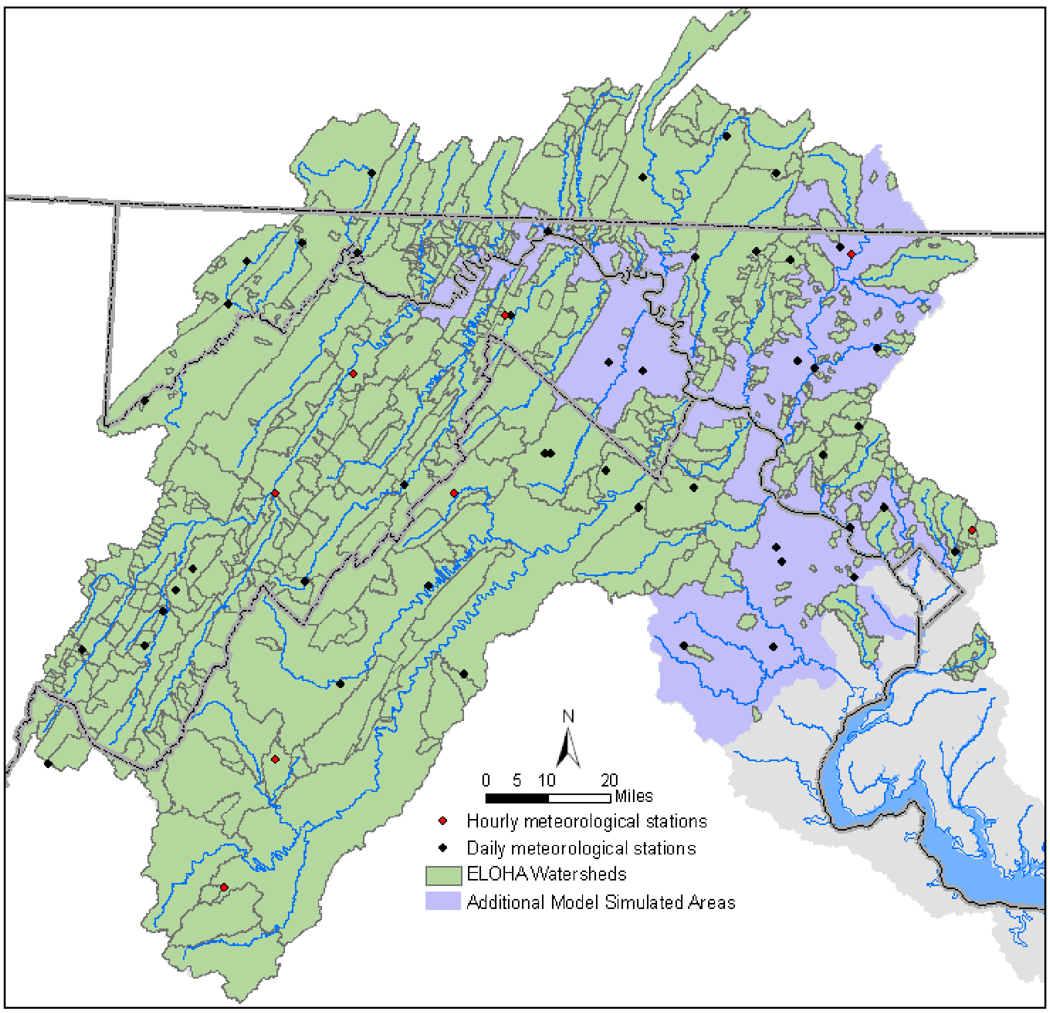

Figure 7. HSPF meteorological stations, major rivers, and the ELOHA watersheds. The additional model simulated areas indicated in the map are areas that were simulated in the HSPF model that were not coupled with biological data in the ELOHA analysis.

ELOHA watersheds, however, may be quite different from the regionally assigned meteorological station due to spatial differences. Conceptually, it is quite likely that the actual precipitation in a small headwaters watershed is not well represented by regional meteorological data. The regional generalization of meteorological data is, therefore, likely a source of model uncertainty.

Future population growth, changes in meteorological conditions, per capita water use, and land use change at differing rates than are represented in the future scenarios was an additional source of uncertainty (Semmens et al. 2006). For example, if climate change affects precipitation amounts by 2030, the predictions of this climate change scenario will differ from actual 2030 conditions as precipitation changes were not simulated in this scenario.

3. Flow Metric Testing and Model Validation

The model's ability to represent actual flows can be further validated through quantitative comparisons of flow metrics which correspond to different parts of the hydrograph. Flow metrics calculated from simulated flow time series are compared to their counterparts calculated from observed flow time series in the Potomac-Susquehanna gage dataset (described above and in Appendix C). Before these comparisons were made, the ELOHA dataset of simulated flows was further winnowed to remove certain types of watersheds.

Of the 747 ELOHA watersheds, 59 had "oddball" flow metric values in the baseline scenario. (Twelve of the 94 assisting watersheds also had oddball baseline values.) Baseline flow metric values should show consistent, tight relationships to natural landscape features because their watershed environments are very similar—heavily forested, little or no imperviousness, and no impoundments, withdrawals, or discharges. Oddball watersheds have baseline values of one or more flow metrics that are significantly different in their relationships to watershed size and gradient. The discrepancies put into question the accuracy of the percent flow alteration calculations derived from those values. Most of the affected watersheds are small, less than 20 mi^2, or located in specific, western regions of the Middle Potomac study area. Oddballs are described in detail in Appendix C. Close examination of the individual oddball watersheds could uncover acceptable reasons for their differences but at this stage it was deemed prudent to remove them from the analyses.

Coastal Plain watersheds in the Middle Potomac study area were also excluded from the final analysis in consideration of that bioregion's unique hydrologic properties. Coastal Plain watersheds comprise relatively small portions of both the simulated flow data (n=32 or 4.3 percent) and the observed Potomac-Susquehanna data (n=7 or 6.7 percent). Model performance in these few watersheds could not be adequately tested.

Removal of the oddball and Coastal Plain watersheds left a total of 656 ELOHA watersheds in the analysis dataset. Simulated flows for these watersheds formed the basis for the FA-E relationships. To equitably compare simulated and observed flow metrics, however, a final selection step was applied which compensated for differences in the watershed size distributions. All but two of the 98 Potomac-Susquehanna watersheds with observed (gaged) flows are greater than 10 mi^2 whereas just 239 of the 656 selected ELOHA watersheds (36.4 percent) with simulated flows are greater than 10 mi^2. This is a consequence of many biological sampling locations being located in headwater streams. For the purpose of comparing flow metric performance, only the 239 ELOHA watersheds with sizes greater than 10 mi^2 were compared to the 98 non-Coastal Plain watersheds in the Potomac-Susquehanna dataset.

Simulated versus observed flow metrics

Good agreement in simulated and observed flow metric values in least-disturbed watershed conditions indicates the model is accurately representing the influence of natural landscape factors governing flow. As stated previously, least-disturbed watersheds are defined for the Middle Potomac study area as heavily forested (greater than 78 percent) with little or no impervious surface (less than 0.35 percent) and no surface withdrawals, discharges, or. Flow in least-disturbed watersheds is governed primarily by natural factors such as watershed size, gradient, and geology. The baseline scenario for the ELOHA watersheds was simulated using these criteria. Eleven watersheds in the Potomac-Susquehanna dataset fully meet or almost meet these criteria. They are identified as "reference" to distinguish them from simulated baseline watersheds.

Good agreement in simulated and observed flow metric values with respect to different anthropogenic factors indicates the model is accurately representing the influence of these factors on flow. Current scenario and observed flow metrics represent the net effect of present-day land and water uses on the flow regime in their respective datasets. The intensities and proportions of land and water uses in current scenarios are not necessarily the same as those in the Potomac-Susquehanna watersheds. Therefore, flow metric responses in the current scenario cannot be expected to exactly match those found in the Potomac-Susquehanna gaged watersheds. Strong similarities and minimal differences between the two, however, should occur if the model is accurately representing anthropogenic impacts. Simulated and observed flow metric responses tested in the same watersheds should behave nearly identically.

Table 2 summarizes the comparison results of fifteen simulated and observed flow metrics (Appendix F). Five of the fifteen flow metrics tested rated "excellent" and three rated "good." They are high pulse count, high flow duration DH17, high flow index MH21, median, flashiness, low pulse duration, 3-day minimum, and the frequency of extreme low flows (Supplemental Table A has a description of all metrics referenced in this report). Simulated and observed values of these eight metrics are not significantly different from one another and the metrics respond similarly to watershed area, gradient, and percent impervious surface, a major anthropogenic cause of flow alteration. When tested together in the same watersheds, the metrics identify the same primary anthropogenic factors altering flow and in many cases, the identical thresholds of impact for each factors. They collectively represent all aspects of the hydrograph except the highest annual flows. These eight flow metrics are therefore good candidates to use in developing FA-E relationships.

High pulse duration and 3-day maximum, the two flow metrics representing highest annual flows, did poorest in the comparison tests between the modeled flow data and the Potomac-Susquehanna gaged watersheds. Simulated values of these two metrics were significantly lower than their observed counterparts and their responses to impervious surface did not parallel those of the observed values.

Despite their drawbacks, these two poor performing metrics reveal interesting relationships to land and water uses in the watershed. Baseline and current scenario values of the two metrics are internally consistent even if the accuracy of their calculated percent flow alteration in ELOHA watersheds, or the percent difference between the two scenarios, is uncertain. As such, the metrics are useful for understanding relationships between land and water uses and stream flow alteration. The same argument can be made for the five flow metrics rated "fair" - annual mean, August median, rise rate, fall rate, and number of reversals. These metrics each fail one of the comparison tests but their responses to percent impervious surface area successfully parallel those found in the corresponding observed values.

Gage data were not available to test how well the simulated flow metrics represent flows in karst areas or flows in headwater streams. None of the Potomac-Susquehanna gaged watersheds with substantive amounts of karst meet the Reference criteria, and very few USGS stream gages are located in watersheds less than 10 mi^2. The watershed model appears to alter flow time series to reflect the effects of karst geology since 10 of the 15 tested flow metrics responded significantly and in a logical manner to karst in the baseline scenario ($p<0.05$). Due to the sampling designs of state macroinvertebrate monitoring programs, more than half of the final ELOHA watershed dataset (56 percent) are less than or equal to 10 mi^2. There was no real justification for taking small or highly karsted watersheds out of the final ELOHA dataset used to develop FA-E relationships, however. After some consideration, the decision was made to keep the untested simulated flow metrics from these watersheds in the ELOHA dataset. This maintained a larger sample size, which increases statistical confidence in the conditional probability curves predicting flow alteration impacts.

In summary, the modeling tools used in this study successfully account for the influences of several important natural factors on flow regimes and respond similarly to the dominant anthropogenic factors disturbing flow regimes in the ELOHA study area. The model provides a solid hydrologic foundation upon which to explore flow interactions with both water and land uses in the watershed.

Table 2. Results of simulated and observed flow metric comparisons.
See Supplemental Table A at the end of the main report for flow metric definitions and Appendix E for details about the comparison tests.

Flow Metric	Metric Type	Comparison of Simulated and Observed
3-day maximum (cfs/mi^2)	magnitude	poor
annual mean (cfs/mi^2)	magnitude	fair
median (cfs/mi^2)	magnitude	excellent
August median (cfs/mi^2)	magnitude	fair
3-day minimum (cfs/mi^2)	magnitude	good
flashiness (ratio)	rate of change	excellent
rise rate (cfs/mi^2)	rate of change	fair
fall rate (cfs/mi^2)	rate of change	fair
number of reversals (#/year)	frequency	fair
high pulse count (#/year)	frequency	excellent
ext. low flow frequency (#/year)	frequency	good
high flow index MH21 (days)	duration	good
high pulse duration (days)	duration	poor
high flow duration DH17 (days)	duration	excellent
low pulse duration (days)	duration	excellent

4. Stream Classification and Biometric Scoring

Poff et al. (2010) and Konrad (2011) suggest classifying streams and rivers in ELOHA studies if ecological responses to flow alteration are expected to vary by stream type. Stream classes are then analyzed by class. Classification serves two purposes in the ELOHA framework: (1) it allows extrapolation of FA-E relationships from gaged streams (where FA-E relationships are derived) to ungaged streams of the same type, and (2) classification informs the selection of future biological monitoring sites in regions with sparse pre-existing biological data or limited monitoring and research resources (Poff et al. 2010). Classification also increases the statistical significance of FA-E relationships by reducing natural variability in the biological variables.

Stream classification organizes water bodies into types, or classes, with similar attributes. It is a "subjective procedure, dependent upon its purpose and the type of data available" and numerous stream classification systems have been developed (Gordon et al. 1992). Poff et al. (2010) recommend classification to stratify "natural variation in measured characteristics among a population of streams and rivers to delineate river types that are similar in terms of [baseline] hydrologic and other environmental features." Baseline hydrologies represent undisturbed flow regimes for the streams under consideration. They are normally created with watershed models. Another approach developed by Olivero and Anderson (2008) uses attributes of existing conditions in the stream and landscape, such as gradient, temperature, stream size, geology, and baseflow, to create distinct stream classes. Both approaches use classification of the environment to reduce natural variability in biological communities, thus increasing confidence in the FA-E relationships. Both approaches assume each stream class has biological communities that (1) are similar to each other and distinctly different from communities in other classes, and (2) respond similarly to hydrologic alteration. A corollary to the baseline hydrology approach is: biological communities that are not otherwise influenced by anthropogenic, non-flow stressors and are very similar can be assumed to belong to the same stream class and will respond similarly to hydrologic alteration.

Several features in the Middle Potomac study area are well recognized as important natural factors governing aquatic communities. The River Continuum Concept (Vannote et al. 1980) describes the longitudinal gradient of physical and biological changes in free-flowing waters as headwater streams merge into small rivers and eventually large rivers. Watershed area and the closely related Strahler stream order are the physical framework on which the concept is based. Karst geology dominates portions of the Middle Potomac study area and, with its many springs and close connections to groundwater, is an important factor affecting stream density and flow volume. The elevated and highly variable topography of the Potomac River basin west of the Piedmont fall-line contrasts sharply with the flat, low-lying Coastal Plain in the east, suggesting fundamental differences in the stream hydrology and biological community structure and function by region.

Two approaches to stream classification for the MPRWA were attempted. The first method, based on the "attribute-based" approach of Olivero and Anderson (2008) proved only somewhat successful because it did not account for the confounding influences of other anthropogenic stressors. The second classification method is a modification of the approach described by Poff et al. (2010). In this study, stream classes were established based on similarities in the macroinvertebrate communities of reference stream sites, rather than their hydrologies. Bioregion was found to best explain the natural variability in macroinvertebrate communities of streams and small rivers. A scoring protocol was then developed, based on each bioregion's reference macroinvertebrate communities, which allows evaluation of the communities anywhere in basin, irrespective of bioregion. All the biological data in the study area can be merged into a single data pool which strengthens confidence in the resulting FA-E relationships. In summary, streams were classed by bioregion and then a biometric score was developed so that the biological data could be evaluated in a single set and FA-E relationships are not tied to bioregion.

Preliminary analyses of stream classification

Stream classification was initially tested using an existing Susquehanna River basin dataset with the purpose of finding a classification system that could potentially be applied to Middle Potomac streams and rivers (Appendix G). Applying the "attribute-based" approach of Olivero and Anderson (2008), statistical relationships between hydrologic alteration and biological responses were developed for the Susquehanna River, which was without a stream classification system (Apse et al. 2008, Case study 5). A single measure of anthropogenic impacts on flow—an index of cumulative water use relative to the 7Q10—was paired with macroinvertebrate and water quality data collected at the same locations and their relationships tested in stream classes defined by various attribute combinations. Attributes included watershed size, gradient, geology, temperature, baseflow, and the CBP bioregions. Nine stream classification systems were examined. The impacts of watershed land uses, impoundments, and discharges were not considered. Systems representing flow volume (watershed size) and groundwater input (baseflow index, karst) had relatively strong relationships, but only in medium sized Susquehanna rivers with low to moderate baseflow contribution or karst presence. Relationships in other stream classes were for the most part insignificant or weak. A full description of the analysis is provided in Appendix G.

A similar watershed area-karst geology classification system was tested using the Middle Potomac ELOHA dataset of delineated watersheds draining to biological sampling sites. Again, the only cause of flow alteration considered was withdrawals and biological data from all watersheds, disturbed and undisturbed, were used in the analysis. Watersheds were grouped according to the NEAHCS size categories (Table 3) and, in smaller watersheds, by whether or not they had relatively high karst geology (greater than 44 percent). The classification resulted in a preponderance of smaller NEAHCS streams because most macroinvertebrate monitoring programs in the study area focus on streams and small rivers. In NEAHCS classes 1a, 1b, and 2, relationships between surface withdrawals expressed as a percent of median stream flow and a suite of 20 biological metrics were weak at best (Table 4). Of the 120 size-karst-biometric regressions, only 10 (8 percent) were significant at $p<0.05$, suggesting most relationships are a result of Type I or II error. None of the significant relationships explained more than 11.2 percent ($r^2=0.112$) of biological variability.

In both the Susquehanna and Middle Potomac dataset with varying degrees of anthropogenic, non-flow disturbances, classification based solely on watershed size and karst did not remove much of the variability in the biological response to surface water withdrawals. Surface withdrawals are only one of several anthropogenic factors altering flow in these watersheds, and flow alteration is only one of several anthropogenic factors impacting macroinvertebrate communities. The attribute-based approach to stream classification was not very successful in the Susquehanna and Middle Potomac watersheds, at least as it relates to surface water withdrawals.

Table 3. Watershed size categories of the Northeast Aquatic Habitat Classification System (NEAHCS). From Olivero and Anderson (2008).
#, number of sampling events per watershed size category in the Middle Potomac study.

Category	Class 1a	Class 1b	Class 2	Class 3a	Class 3b	Class 4	Class 5
Name	Head-water	Creek	Small River	Medium Tributary River	Medium Mainstem River	Large River	Great River
Area (mi^2)	<3.86	3.86 - <38.6	38.6 - <200	200 - <1,000	1,000 - <3,861	3,861 - <9,653	$\geq$ 9,653
Area (km^2)	<10	10 - <100	100 - <518	581 – <2,590	2,590 - <10,000	10,000 - <25,000	$\geq$ 25,000
#	607	407	215	74	10	0	0

Table 4. Linear regression coefficients (r^2) for biometrics versus surface withdrawals (expressed as a percent of annual median stream flow) in six size-karst classes, Middle Potomac study area.
Significance: * = $0.01<p\leq0.05$; ** = $p\leq0.01$; ns = not significant. See Supplemental Tables for explanation of biometrics.

NEAHCS	Class 1a				Class 1b				Class 2			
Karst is:	Low		High		Low		High		Low		High	
Biometric / n is:	(593)		(14)		(325)		(82)		(180)		(35)	
Chessie BIBI	0.001	ns	0.024	ns	0.005	ns	0.032	ns	0.031	*	0.024	ns
ASPT modified index	0.000	ns	0.006	ns	0.013	*	0.028	ns	0.010	ns	0.003	ns
Beck's Index	0.000	ns	0.007	ns	0.008	ns	0.112	**	0.017	ns	0.014	ns
Number of Ephemeroptera	0.000	ns	0.072	ns	0.000	ns	0.086	**	0.003	ns	0.047	ns
%EPT	0.001	ns	0.042	ns	0.001	ns	0.020	ns	0.000	ns	0.024	ns
%Ephemeroptera	0.000	ns	0.014	ns	0.004	ns	0.024	ns	0.012	ns	0.042	ns
Hilsenhoff Family-Level Biotic Index	0.001	ns	0.024	ns	0.000	ns	0.055	*	0.001	ns	0.060	ns
GOLD Index	0.002	ns	0.056	ns	0.001	ns	0.021	ns	0.001	ns	0.019	ns
%Gatherers	0.000	ns	0.043	ns	0.003	ns	0.018	ns	0.005	ns	0.006	ns
%Scrapers	0.001	ns	0.007	ns	0.005	ns	0.052	*	0.001	ns	0.019	ns
%Dominant3	0.000	ns	0.028	ns	0.001	ns	0.015	ns	0.012	ns	0.019	ns
%Filterers	0.001	ns	0.020	ns	0.000	ns	0.006	ns	0.009	ns	0.001	ns
%Swimmers	0.000	ns	0.011	ns	0.002	ns	0.005	ns	0.018	ns	0.000	ns
%Tolerants	0.001	ns	0.010	ns	0.000	ns	0.017	ns	0.000	ns	0.013	ns
%Chironomids	0.001	ns	0.065	ns	0.001	ns	0.024	ns	0.000	ns	0.008	ns
%Clingers	0.002	ns	0.013	ns	0.001	ns	0.023	ns	0.004	ns	0.026	ns
Number of Sensitive Families	0.000	ns	0.005	ns	0.007	ns	0.116	**	0.018	ns	0.000	ns
Shannon-Weiner Index	0.000	ns	0.056	ns	0.000	ns	0.022	ns	0.016	ns	0.014	ns
Family-level Taxa Richness	0.000	ns	0.169	ns	0.000	ns	0.100	**	0.008	ns	0.016	ns
%Collectors	0.001	ns	0.044	ns	0.008	ns	0.065	*	0.056	**	0.004	ns

Reference streams

An alternative to the Poff et al. (2010) recommendation to classify streams based on modeled baseline hydrologies is to classify streams based on the biota found in actual high quality, or "reference," stream sites. This approach requires sufficient numbers of these reference sites in the study area. When stream classification analysis is based on the highest quality stream sites, the confounding influences of anthropogenic activities on the biota are minimized and the watershed's natural features govern its macroinvertebrate communities. Highest quality sites are defined here as having excellent habitat conditions (habitat metric scores 16 – 20) and non-stressful levels of pH (6 – 9), conductivity (<500 μS/cm), and dissolved oxygen (>5 mg/liter). It is assumed that these sites have relatively undisturbed flow regimes since scores for habitat features such as bank stability, bank vegetation, embeddedness, channel alteration, epifaunal substrate, and riffle/run/pool ratios are all excellent. An examination of the simulated current hydrologies at the 78 reference sites found in the ELOHA dataset supports this assumption. Twenty-three of the sites (29 percent) were in watersheds that fully met the criteria for baseline. Fifty-five of the sites (71 percent) were in watersheds with more than 52 percent forest cover, a threshold identified in the Category and Regression Tree (CART) analysis of the Middle Potomac large river flow needs assessment as low risk of flow alteration (Appendix A). Average impervious surface cover was 0.34 percent; none of the watersheds were associated with more than 1.3 percent imperviousness. Most flow metrics at the reference sites showed less than +/-20 percent alteration from modeled baseline values. The exceptions typically occur in the low magnitude flow metrics, with alteration greater than 20 percent, and are associated with karst geology, discharges, and/or land use dominated by agriculture.
The "reference site" approach successfully identified distinct macroinvertebrate communities in the Potomac River basin (Astin 2006) and across the Chesapeake Bay watershed (Buchanan et al. 2011). A CART analysis of the Chesapeake data showed the strongest classification factor for a range of family-level biometrics was most often Level 4 ecoregion followed by elevation, latitude, and hydrogeomorphic region. The latter three factors are elements used to define ecoregions. Aggregation of the Level 4 ecoregions into "bioregion" classes maintained the ecoregion classification efficiency while increasing the number of reference samples in each class. Four bioregions underlay the Middle Potomac study area: Ridges, Valleys, Piedmont, and Coastal Plain (Figure 2, Main Report). More than 80 percent of streams in the ELOHA dataset flow across two bioregions by the time they drain out of NEAHCS Class 2 watersheds. Nevertheless, a Kruskal-Wallis one-way analysis of variance found significant differences between the Ridges, Valleys, and Piedmont bioregions in 38 of the 42 biometrics tested ($p<0.05$).[4] Reference macroinvertebrate communities are similar within bioregion and show significant differences with communities outside their bioregion. Recall that this is a desired outcome of the stream classification step recommended by Poff et al. (2010).

Three other classification factors were investigated by Buchanan et al. (2011) because of their known or suspected influence on macroinvertebrate communities. Under the River Continuum Concept, stream size imposes a longitudinal gradient on community composition and trophic relationships (Vannote 1980). Season governs macroinvertebrate growth rates, life cycles, and behaviors through its control of temperature and sunlight. Karst geology could potentially influence macroinvertebrate communities through its distinct influence on stream water flow, physical habitat, and chemistry, particularly conductivity.

---

[4] Habitat and water chemistry data needed to precisely identify reference stream sites in the Coastal Plain were not available at the time of this analysis, so the Kruskal-Wallis test could not be applied to the Coastal Plain data.

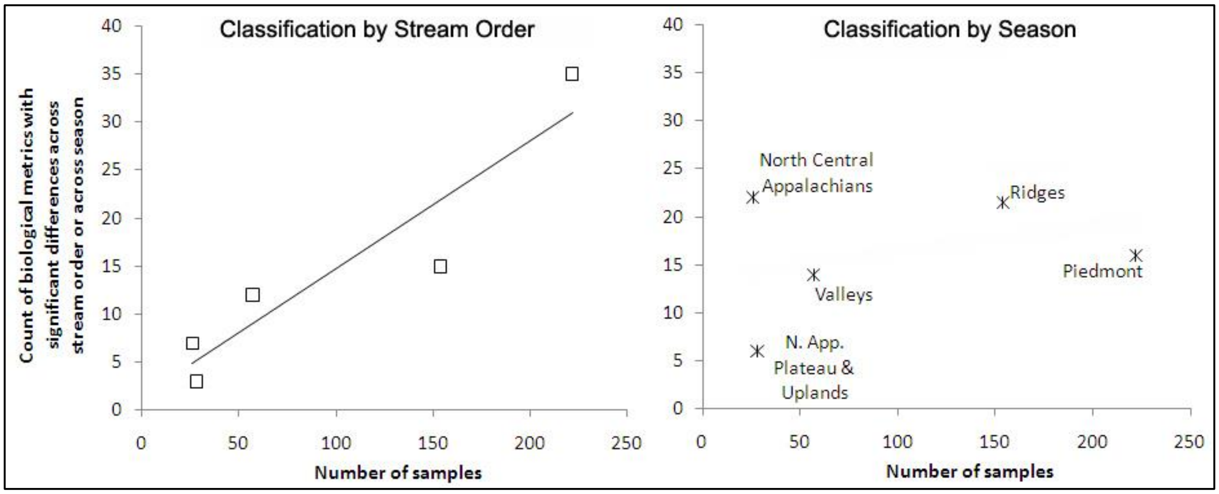

Figure 8. Number of biological metrics in each bioregion that show significant differences across first to fourth Strahler stream order (left panel) and across season (right panel), versus total number of samples from reference-quality sites.
Total number of family-level macroinvertebrate metrics tested is 42. (Figure 4 in Buchanan et al. 2011)

Within bioregion, the number of macroinvertebrate metrics responding to stream size expressed as Strahler order is directly related to sample size for Strahler orders 1-4 (Figure 8, left panel). In Chesapeake bioregions with the fewest reference samples (North Central Appalachian, n=26; Northern Appalachian Plateau and Uplands, n=28) only 6 or 7 of the 42 biometrics examined showed significant differences across Strahler order; in the bioregion with the most reference sites (Piedmont, n=222) 35 of the 42 metrics showed significant differences across Strahler order. Just one metric (%Shredders) showed a consistent change across Strahler order in all bioregions. Inherent variability in the biometrics, even under reference conditions, appears to confound macroinvertebrate relationships with stream order until sample sizes are large enough to overcome this variability and show statistically significant differences. The sample size effect indicates macroinvertebrate relationships to Strahler stream order are weak relative to other natural controls in first to fourth order streams, at least in the Chesapeake Bay drainage bioregions.

Evidence of seasonal controls on macroinvertebrate populations within bioregion is more pronounced and not as dependent on reference sample size (Figure 8, right panel). Of the three Middle Potomac bioregions, Ridges with its higher elevations and shorter growing season exhibited significant differences across season in 21 of 42 biometrics, whereas Valleys and Piedmont exhibited seasonal differences in only 14 and 16 of the 42 biometrics, respectively. Seven biometrics representing highly seasonal taxa (e.g., Diptera, Trichoptera) or taxa dependent on seasonal food sources (e.g., filterers, gatherers, net caddisfly) responded to season in all three bioregions. Twelve biometrics representing the range of macroinvertebrate community features show no season effect in any bioregion, including five taxa richness and diversity metrics, four pollution sensitivity metrics, two composition metrics, and one habit metric (%Clinger).

Karst geology is found in a broad swath through the Valleys bioregion in the Middle Potomac study area, and in a few areas of the Piedmont (Figure 4 in Appendix C). Comparisons of reference sample sites with and without karst geology in the Valleys bioregion indicated only two biometrics respond significantly to karst ($p<0.01$). Proportions of clingers and Ephemeroptera are significantly higher in karst areas; the remaining 40 biometrics are not significantly different.

In summary, for the Chesapeake Bay watershed which includes the Middle Potomac study area, stream size, season, and karst are less important than bioregion as classification factors explaining natural variability in macroinvertebrate communities of streams and small rivers. Stream size and season have some influence, however, it is impractical to minimize natural variability in macroinvertebrate communities by splitting the ELOHA dataset into 64 stream classes (4 bioregions x 4 Strahler stream orders x 4 seasons), or even more classes if karst is also considered. Macroinvertebrate metrics that are not responsive to one or more of the natural environmental factors would be needlessly split and the number of all samples in many stream classes would be very low. Even if similar classes are merged, only a few would have sample sizes adequate for developing FA-E relationships. Only bioregion was used to classify streams in the ELOHA dataset.

Biometric scoring

To avoid using stream classification explicitly, one can employ a scoring technique that takes into consideration each sample location's bioregion, stream order, season, and karst geology and converts the sample's biometric values to a common scale of low to high status. Such a scoring approach was developed in Buchanan et al. (2011) for the Chesapeake Bay bioregions. Percentiles of the distribution of each macroinvertebrate metric's values at reference sites were used to create bioregion-specific scoring thresholds for each metric. Within each bioregion, careful selection of the scoring thresholds avoided or minimized the effects, if any, of stream order, season, and karst. Metric scores could then be directly compared across the entire Chesapeake region.

The gradient of possible values for each biometric ranges widely in the study area, from least-like reference (degraded) to most-like reference. The narrower distributions of biometric values at independently identified reference and degraded sites typically overlap, creating a middle range of indeterminate values. A percentile in the tail of the reference distribution of biometric values (T percentile)—often the 25th percentile for biometrics that decrease with disturbance and the 75th percentile for biometrics that increase— is a threshold that most effectively separates the indeterminate values into "more like reference" and "more like degraded" groups. Values in the latter group, as well as those with clearly degraded values, receive the lowest score on the scoring scale. The remaining biometric values are further divided with the 50th percentile of the biometric's reference distribution (M percentile). Values above the 50th percentile for biometrics that decrease with disturbance (and below the 50th percentile for biometrics that increase with disturbance) are most different from biometric values found at degraded sites and receive the highest score on the scoring scale, along with a status of "good." (To increase resolution, this highest scoring range can be further divided to create a "good" and an "excellent" status.) Values between the T and M percentiles receive an intermediate score and a status equivalent to "fair." The T and M scoring thresholds for the biometrics included in this study are presented in Table 5.

In summary, the use of biometric scores rather than values in the MPRWA obviates the need to split the biological data into explicit stream classes. All the biological data associated with the ELOHA watersheds spread across the study area can be merged into a single data pool once the metrics have been scored. The large size of this pool strengthens confidence in the resulting FA-E relationships.

Table 5. Bioregion-specific thresholds used to score macroinvertebrate family-level metrics. M%ile is the 50th percentile of the biometric distribution at independently identified reference quality stream sites; T%ile is a percentile in the tail of the reference distribution closest to values associated with degraded quality stream sites. For biometrics that respond to degradation by increasing (pos), values greater than T%ile score 1, between T%ile and M%ile score 3, and less than M%ile score 5. For biometrics that respond to degradation by decreasing (neg), values less than T%ile score 1, between T%ile and M%ile score 3, and greater than M%ile score 5. See text for more detail.

		Piedmont		Ridges		Valleys	
Biological Metric Name	**Resp.**	**M%ile**	**T%ile**	**M%ile**	**T%ile**	**M%ile**	**T%ile**
ASPT modified index	pos	4.2	4.6	3.5	3.7	4.1	4.4
Beck's Index	neg	9.0	7.0	13.0	10.0	9.0	6.9
Number of Ephemeroptera Families	neg	3.0	3.0	4.0	3.0	4.0	3.0
Hilsenhoff Family-Level Biotic Index	pos	3.63	4.54	3.64	4.05	4.10	4.70
GOLD Index	neg	0.919	0.804	0.827	0.793	0.827	0.675
%Chironomids	pos	3.25	6.31	9.29	11.80	7.53	10.78
%Clingers	neg	85.64	83.83	70.82	62.21	67.8	60.08
%Collectors	pos	52.71	71.02	54.93	54.93	66.67	74.18
%Dominant3	pos	63.14	69.89	60.47	69.13	65.45	70.50
%Ephemeroptera	neg	35.10	21.94	26.83	13.51	29.33	18.97
%EPT	neg	72.24	48.12	71.94	68.42	64.94	50.94
%Filterers	neg	30.14	18.62	17.25	17.25	23.20	20.00
%Gatherers	pos	16.92	39.20	34.44	34.44	37.72	45.52
%Scrapers	neg	19.16	10.62	11.06	3.37	14.41	7.02
%Swimmers	neg	10.45	4.36	10.73	4.59	10.74	6.19
%Tolerants	pos	3.23	12.30	10.77	12.48	9.14	17.60
Number of Sensitive Families	neg	7.0	6.0	9.0	7.0	6.0	5.4
Shannon-Weiner Index	neg	2.170	1.920	2.261	1.990	2.086	1.878
Family-level Taxa Richness	neg	15.0	14.0	16.0	13.25	15.0	13.0

5. Flow Alteration Assessment

The purposes of the basin-wide hydrologic alteration assessment were to identify where flow alteration has occurred in the basin and the physical stressors associated with the alteration. The drivers of hydrologic alteration evaluated as part of the MPRWA include land uses, impounded waters, withdrawals, and discharges.

Flow alteration is the percent difference in a particular flow metric between baseline and current conditions in a given watershed, as simulated by the hydrologic model for the ELOHA watersheds. This report section describes the spatial distribution of alteration found in the Potomac basin flow metrics and correlates them with the watershed characteristics driving the changes. A discussion on limitations is also presented.

Watershed characteristics

Watershed characteristics that are adjusted between the baseline and current scenarios to simulate hydrologic alteration include withdrawals, discharges, impoundments, and land uses. A brief description of each is provided below; however, a complete description of the ELOHA watershed characteristics, including maps of their spatial distribution, is provided in Appendix H. Details on the associated data sources are available in Appendix C.

The current model scenario includes withdrawals and discharges from the year 2005. Withdrawals and discharges are not simulated in the baseline scenario. Simulated withdrawals include only those taken from surface waters due to limitations in the groundwater withdrawal dataset and in the hydrologic model. There are 253 simulated surface withdrawals in the study area (Figure 9). The withdrawals are spatially distributed such that only 115 of the 747 ELOHA watersheds have any amount of modeled withdrawal. Further, a total of 143 point source discharges are simulated in the current scenario (Figure 10). These discharges are located in 82 of the ELOHA watersheds.

There are 16 modeled impoundments in the current scenario (five in Virginia, four in West Virginia, four in Maryland, one in Pennsylvania, and two that are located along the Maryland-West Virginia border). Only 15 of the ELOHA watersheds contain some amount of simulated impounded waters. Impoundments are not simulated in the baseline scenario.

Current scenario land uses are those represented in the model from the year 2000. Baseline land uses adjust the percent forest to be a minimum of 78 percent. If, under current conditions, a watershed has higher than 78 percent forest then the percent forest is not adjusted for baseline conditions. Percent impervious cover is also adjusted to a maximum of 0.35 percent.[5] All other land use categories were maintained in their original, current scenario proportions, but were adjusted to maintain the total watershed area. A complete description of the land use adjustment utilized to model baseline conditions is available in Appendix E. The land use changes resulted in a small increase in urban areas across much of the study area from baseline to current conditions, with larger increases in the WMA. Agriculture increased from the baseline to current scenario across most of the basin, with the largest increases occurring in the Monocacy, Antietam, Conococheague, and Shenandoah watersheds. Corresponding to this increase in urban and agricultural areas, much of the basin had a decrease in the amount of forest cover between baseline and current conditions.

Spatial distribution of flow alteration

The spatial distribution of alteration in selected flow metrics between the baseline and current scenarios can be found in Figures 11 – 14. The figures are grouped by the flow range of the metrics. A complete definition of each flow metric is available in the supplemental tables. A discussion of the spatial distribution of hydrologic alteration found in these metrics follows.

High flow metrics include high flow index MH21, high flow duration DH17, high pulse count, and 3-day maximum (Figure 11). High flow index MH21 decreases over much of the basin, with focused areas of increasing values. The most extreme decreases in the high flow index MH21 are located in the WMA. A similar spatial pattern is found in the high flow duration DH17 metric.

[5] The thresholds for baseline forests and impervious cover were determined utilizing a CART analysis, documented in Appendix D.

Both of these metrics measure duration of high flow events. High pulse count, driven to a large extent by the number of rainfall events, was somewhat less sensitive to landscape changes. Much of the basin exhibits zero percent alteration in this metric and only those watersheds with greater than approximately 1-2 percent impervious surface area, large withdrawals, and large discharges deviated from baseline levels. The 3-day maximum, representing the highest annual magnitude of flow, decreases relative to baseline levels through most of the South Fork of the Shenandoah watershed, the Conococheague, and through most of the central portion of the basin – regions which tend to be agricultural and/or underlain by karst geology but which also have increasing development. The metric increases moderately in the North Fork of the Shenandoah watershed, the Monocacy, and the North and South branches of the Potomac River, and increases steeply in the urbanized WMA and other urban areas. Multiple land and water uses appear to alter these high range flow metrics, sometimes in opposing directions, making it hard to generalize about net anthropogenic impacts across the Middle Potomac region.

The two mid-range flow metrics (median annual flow and August median flow) are similar in their overall distribution of alteration (Figure 12); however, there are some distinctions. For example, August median flow decreases, sometimes as much as 100 percent, in many of the WMA watersheds while a few of these same watersheds show slight increases in their overall median flows. August median flow decreases in a large section of the North Branch of the Potomac River while the overall median flow shows a slight increase. The competing influences of various land uses and withdrawals and discharges again make it difficult to generalize across the Middle Potomac region.

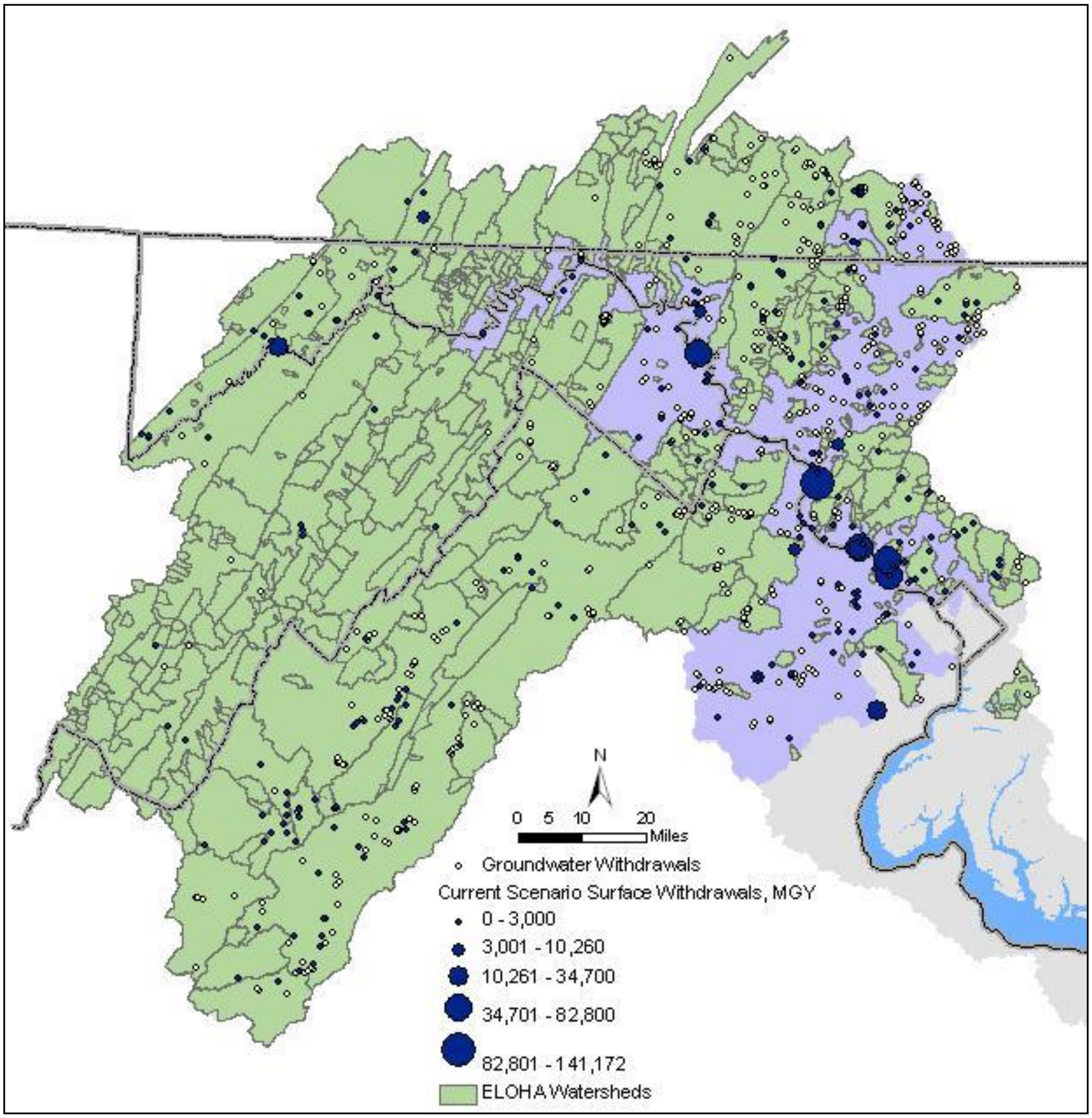

Figure 9. Withdrawal locations in the model simulated area.

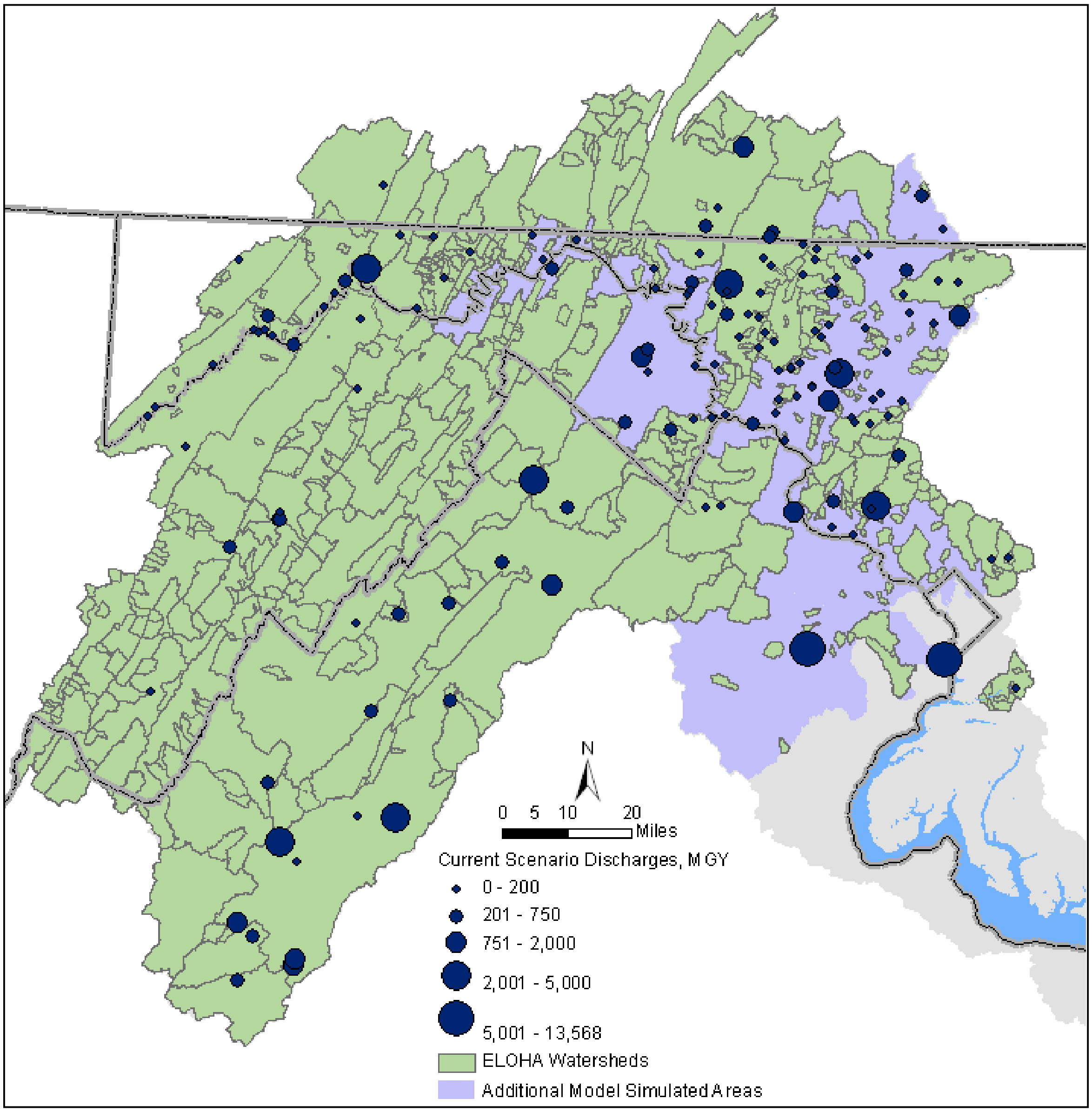

Figure 10. Location of current scenario discharges.

Flow metrics representing low flow conditions include low pulse duration, 3-day minimum, 7Q10, seasonal Q85, and low pulse count (Figure 13). Seasonal Q85, 3-day minimum, and 7Q10 represent low magnitude and for the most part show similar spatial patterns of decreasing levels over much of the basin. A large number of watersheds in the upper portion of the basin show no alteration in the duration of low pulses, while the majority of the basin experiences shorter periods of low flows. Some exceptions include (but are not limited to) the Conococheague, portions of the North Fork of the Shenandoah River, and Goose Creek – all showing increases in the duration of low pulses between the current and baseline scenarios. Similarly, the count of low pulses shows zero alteration in many watersheds in the upper portions of the basin, but in the swath of the basin from the Shenandoah through the Conococheague, an area underlain by karst geology, extreme amounts of negative alteration in the count of low pulses are noted. The WMA and scattered upstream watersheds, on the other hand, show increases up to 1900 percent in the count of low pulses.

Rate of change metrics include flashiness, number of reversals, rise rate, and fall rate (Figure 14). Overall, flashiness and the number of reversals increase with development over much of the basin. Flashiness is the frequency and rapidity of short-term changes in flow. Alteration in rise rate and fall rates show more spatial variation, with portions of the basin increasing while others decrease. It is evident in the comparisons that multiple anthropogenic factors are at play in altering the stream flow of each watershed in the Middle Potomac study area, sometimes in opposing directions. The net result of these factors leads to some sharp discrepancies between watersheds in how stream flow changes from baseline, and prevents making many generalizations for the entire study area. Watersheds need to be examined on an individual basis, to account for the combined impacts of different land and water uses.

Flow alteration associations with watershed characteristics

Understanding which anthropogenic factors drive the largest changes in stream flow can facilitate management efforts. An analysis was undertaken to evaluate how changes in watershed characteristics are correlated with the alteration in various flow metrics. A complete description of the results of this analysis is provided in Appendix H. Figure 10 in the appendix illustrates the relationships in the ELOHA watersheds between seventeen flow metrics and percent forest, percent agriculture, and percent urban land uses. Figure 11 in the appendix illustrates the relationships between the seventeen flow metrics and percent impervious surface cover. The findings are summarized below and in Table 4 of Appendix H.

Several watershed features are associated with the high range flow metrics. (1) Decreases in the duration of high flow events as measured by the high flow index MH21 and the high flow duration DH17 occur in many places in the Potomac basin and are correlated with urban areas. These changes were experienced across much of the basin as the region was converted from predominantly forested and agricultural to urbanized areas. (2) High pulse count shows relatively little alteration across the basin, however, increases are associated highly urban areas. This explains why the most extreme increases in high pulse count occur in the WMA. Decreases in high pulse count are associated with increasing withdrawals. (3) Increases in the 3-day maximum are associated with increasing urban area, again explaining why the most extreme positive alteration values are found in the WMA. Increasing agriculture is associated with decreases in the 3-day maximum, found in the Shenandoah watershed and other agricultural watersheds throughout the basin. Increasing withdrawals is another cause for decreases in the 3-day maximum. Keep in mind that multiple

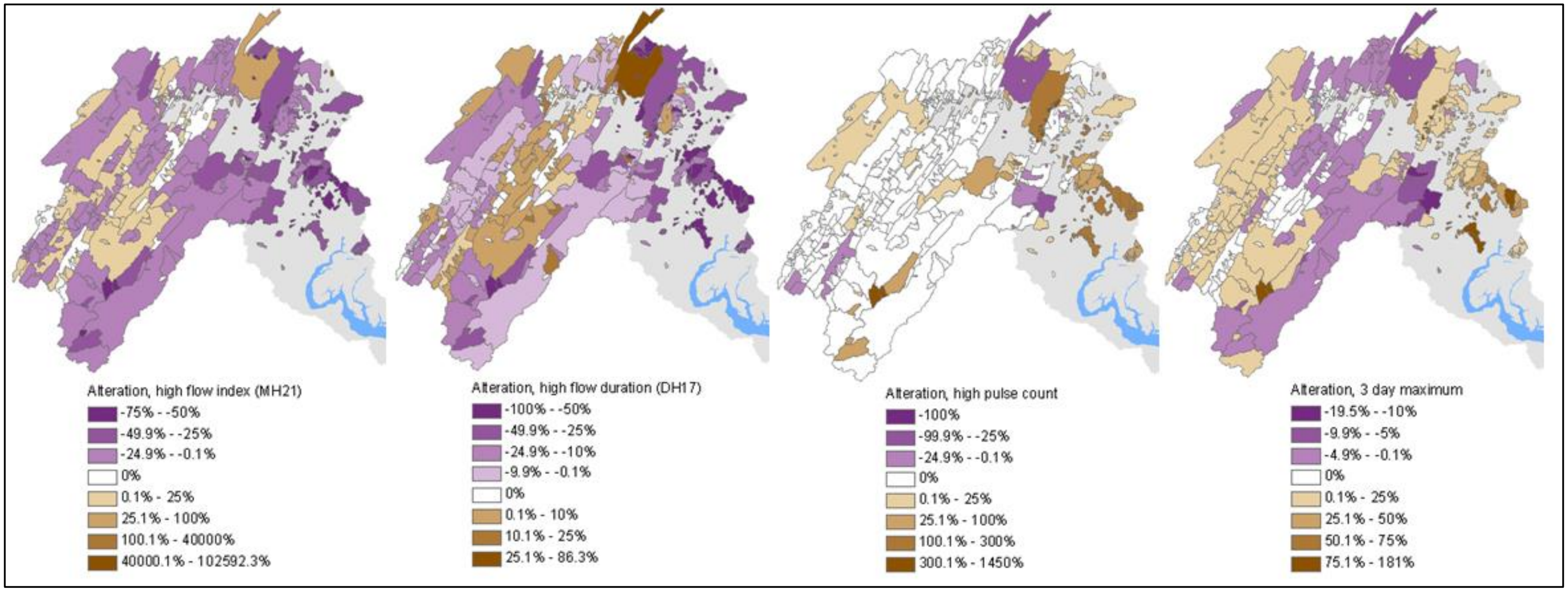

Figure 11. Alteration in select high-range flow metrics between the baseline and current scenarios.

Alteration in each metric is calculated as (current-baseline)/baseline and is expressed as a percent. Negative alteration is shown in purple and positive alteration is shown in brown. Negative and positive are not indicative of "good" and "bad", but rather indicate the direction of change from baseline conditions. More extreme alteration is shown in darker shades of either purple or brown. See Supplemental Table A at the end of the main report for flow metric definitions.

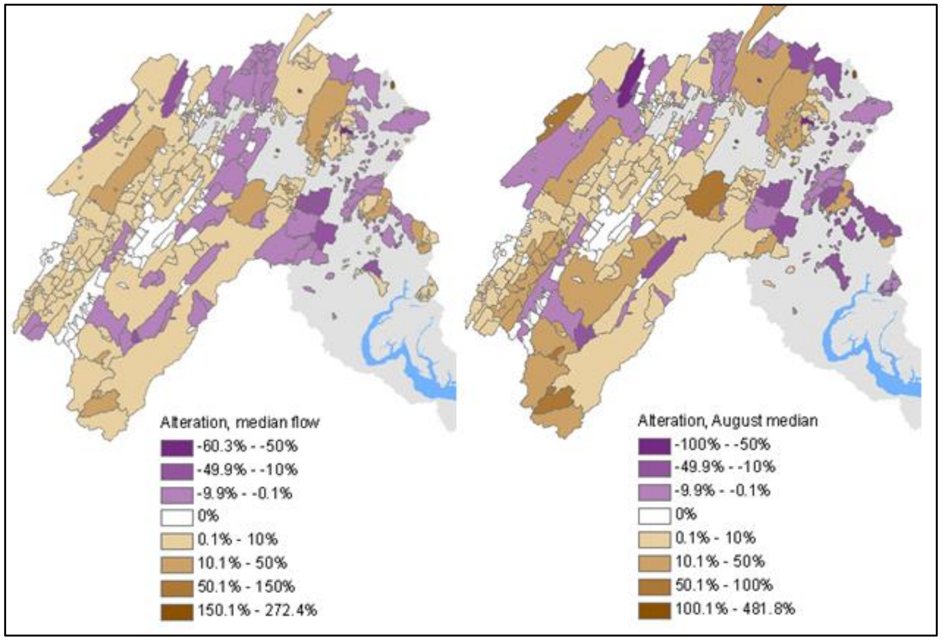

Figure 12. Alteration in select mid-range flow metrics between the baseline and current scenarios.

Alteration in each metric is calculated as (current-baseline)/baseline and is expressed as a percent. Negative alteration is shown in purple and positive alteration is shown in brown. Negative and positive are not indicative of "good" and "bad", but rather indicate the direction of change from baseline conditions. More extreme alteration is shown in darker shades of either purple or brown. See Supplemental Table A at the end of the main report for flow metric definitions.

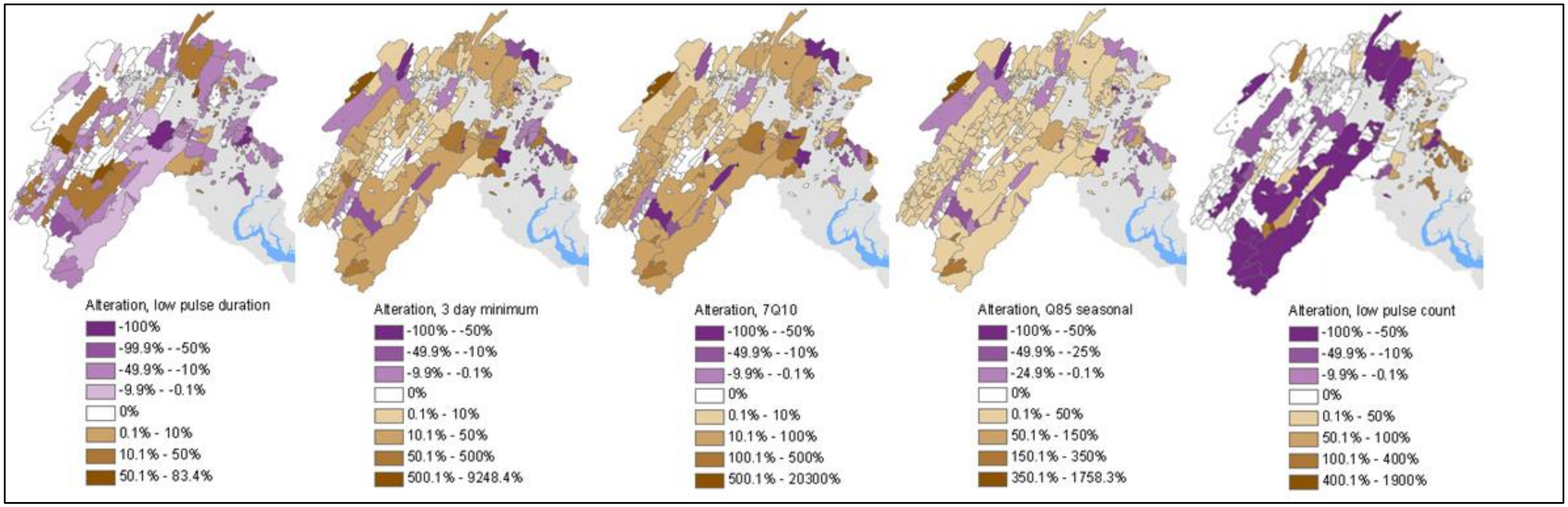

Figure 13. **Alteration between the baseline and current scenarios in select low-range flow metrics.**

Alteration in each metric is calculated as (current-baseline)/baseline and is expressed as a percent. Negative alteration is shown in purple and positive alteration is shown in brown. Negative and positive are not indicative of "good" and "bad", but rather indicate the direction of change from baseline conditions. More extreme alteration is shown in darker shades of either purple or brown. See Supplemental Table A at the end of the main report for flow metric definitions.

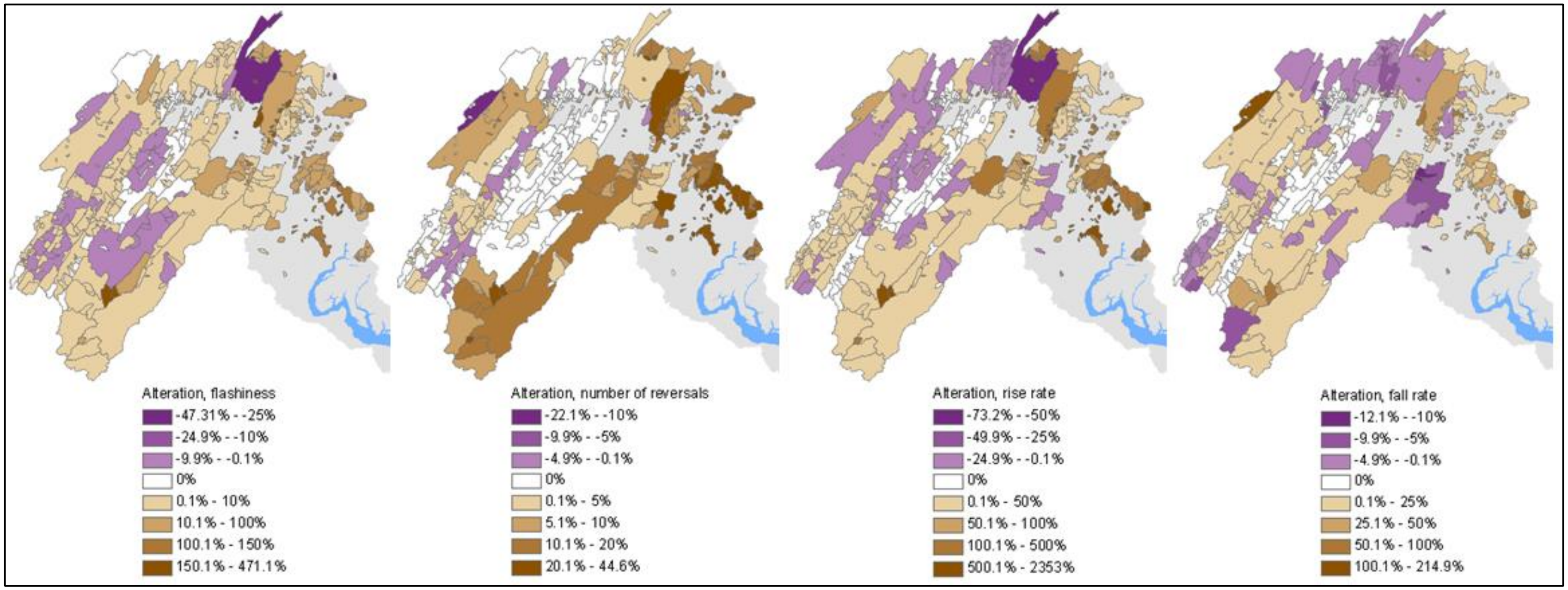

Figure 14. Alteration in select rate of change flow metrics between the baseline and current scenarios.

Alteration in each metric is calculated as (current-baseline)/baseline and is expressed as a percent. Negative alteration is shown in purple and positive alteration is shown in brown. Negative and positive are not indicative of "good" and "bad", but rather indicate the direction of change from baseline conditions. More extreme alteration is shown in darker shades of either purple or brown. See Supplemental Table A at the end of the main report for flow metric definitions.

watershed characteristics are changing simultaneously. A moderate change in one direction from a particular watershed characteristic may be overcome by an extreme influence of another watershed characteristic in the opposite direction.

Changes in watershed characteristics also correlate with changes in mid-range flow metrics—median flow and August median flow. Lower levels of these two flow metrics occur in highly urban areas with greater than 10 percent impervious surface however the strongest deviations from baseline are related to discharges and withdrawals. The largest increases are associated with proportionally large discharges in the watershed; the largest decreases are associated with proportionally large withdrawals. Impoundments may also change median levels, sometime increasing and other times decreasing the levels. August median also appears influenced by the percent agriculture in the watershed. A possible explanation is the drop in usual evapotranspiration when forest is converted to agriculture. Reducing forest cover decreases the amount of water lost to evapotranspiration, and makes more rainwater available for streamflow.

The low magnitude flow metrics 3-day minimum and 7Q10 show responses similar to the August median (above). Large withdrawals decrease levels because less water is available in-stream; large discharges increase levels by adding water, albeit with different water quality. Levels of both flow metrics are lower in areas with greater than roughly 25-30 percent urban cover although this threshold is often confounded by withdrawals and discharges. On the other hand, increasing agricultural area increases low flows, particularly 3-day minimum because of lower evapotranspiration rates under crop cover (current scenario) when compared to forest (baseline scenario). Decreasing evapotranspiration makes more water available for streamflow. Simulated low flows below the major impoundments in the Middle Potomac study area suggest that flow management may serve to increase minimum flows.

The number of low pulses is zero in watersheds with very high discharges, but low pulses become more frequent in watersheds with large withdrawals and in highly urban areas. They appear unaffected by agricultural land cover. The duration of low pulses is shorter in highly urban areas but longer in watersheds with large withdrawals. Increasing the duration of low pulses can be moderately associated with increasing withdrawals and impoundments. The number of low pulses is reduced significantly in areas with the greatest increases in agriculture, the Shenandoah and Conococheague watersheds, for reasons provided above. The number of low pulses is most dramatically increased in the WMA, where the largest urbanization occurred between baseline and current conditions.

Metrics representing the rate of hydrologic change, including flashiness, number of reversals, rise rate, and fall rate, all increase sharply in highly urban watersheds. This explains why the most populated watersheds in the basin show extreme positive alteration in rate of change flow metrics. Increases in withdrawals also increase overall flashiness and fall rate; however, they slightly decrease the number of reversals. Discharges tend to reduce flashiness.

Limitations

The spatial assessments of hydrologic alteration show where hydrologic alteration has occurred across the basin, and the associations between alteration in individual flow metrics and anthropogenic activities suggest which factors are driving those changes. There are several limitations to this analysis, however. First, this alteration assessment was conducted utilizing simulated flows and available observed data in the study area and is, therefore, subject to the limitations of the model as described in the discussion model uncertainty as well as the limitations in observed data described in the conclusions of this chapter. One limitation stems from the locations and sizes of the ELOHA watersheds which were originally selected based on the locations of biological monitoring stations. The specific characteristics of these ELOHA watersheds bound the degree of alteration that can be described through this work. For example, a large number of ELOHA watersheds have experienced urbanization and/or agriculture, so relationships between land uses and flow alteration are more readily discernible. A limited number of watersheds show the influence of large impoundments, withdrawals, and discharges, making it difficult to discern the associated flow impacts of these anthropogenic activities. Results of some statistical analysis on the impacts of withdrawals and discharges on select flow metrics are presented in Appendix H and indicate the tenuousness of their flow metric associations. Finally, a large number of the ELOHA watersheds are less than 10 mi^2 and their simulated flows are difficult to confirm since most gaged watersheds are larger than 10 mi^2.

Another limitation to this assessment of hydrologic alteration is that groundwater withdrawals were not included in the hydrologic model due to the limited availability of groundwater data and current limitations in the hydrologic model. Over 70 percent of the permitted withdrawal locations in the study area are from groundwater sources. (The volume of the groundwater withdrawals, and therefore its percent of total withdrawals, is not well documented.) Although many of them are relatively small, the cumulative impacts of these groundwater withdrawals may be considerable and are worth evaluating in future efforts.

Utilizing the withdrawal and discharge data in combination, consumptive use estimates can be obtained for the ELOHA watersheds; however, the consumptive use values are potentially problematic for a number of reasons. As mentioned before, there are no groundwater withdrawals in the hydrologic model. Discharge amounts, however, include waters originating from ground and surface water sources. Due to limitations in the data, it is not possible to model only discharges associated with surface water withdrawals. In addition, the withdrawal and discharge datasets are not linked. That is, the dataset does not explicitly note which withdrawals are released to which discharge locations. A simple comparison of withdrawals and discharges for a particular watershed to estimate consumptive use could be erroneous without taking into account these linkages because withdrawals and associated discharges can be located in different watersheds.

Although there are numerous small impoundments in the study area, only 16 are included in the hydrologic model (Appendix E has a description of this methodology). This relatively small sample size limits the statistical ability to evaluate the hydrologic impacts of impoundments. Additional efforts to understand the impacts of impoundments could include incorporating additional dams in to the hydrologic model and/or selecting additional watersheds that are influenced by the 16 that are already included to increase the sample size.

Future flow assessment

The assessment of hydrologic alteration from baseline to current conditions demonstrated significant relationships between flow alteration and increases in agriculture and especially urban areas, which correspond to decreases in forest cover. The assessment made some limited conclusions about the influence of impoundments, withdrawals and discharges on flow statistics. The Middle Potomac watershed is projected to experience significant increases in human population between now and 2030, accompanied by land cover changes and increasing demands for water. Details are described in Chapter 4 and Appendix B.

Flow alterations projected for five different future scenarios were evaluated to identify areas of potential concern. In Appendix E, the details of developing the five future scenarios in the CBP HSPF modeling environment are described and the resulting hydrologic alteration in seven selected flow metrics are presented. The maps in Figure 15 illustrate the effects on four of the seven metrics of the projected changes from current conditions to the five future scenarios. Simulated flows for these future scenarios suggest that subwatersheds within the Potomac basin will be variously affected depending on the intensities of each natural or anthropogenic factor. The magnitude and direction of change of hydrology (future compared to current conditions) depends on the assumptions of each scenario, on changes in individual watershed characteristics (land and water use), and on what aspect of the flow regime (as measured by a specific flow statistic) is being considered. There was no regional pattern of flow alteration that applied to all scenarios and all flow metrics, and the impact on flow of each future scenario on different subwatersheds is subject to its own unique interpretation. The climate change scenario, for example, shows changes (from current conditions) in low pulse duration in both the positive and negative direction, but typically decreased high pulse count and median flow across the entire area, and decreased flashiness in most places except for areas projected to experience development. Portions of the basin are expected to experience increases in median flow under the DP1, DP2, and power scenarios resulting from a decrease in forest. Forest removal decreases the amount of water lost to evapotranspiration, making more water available for streamflow. Other portions of the basin under those same scenarios are expected to have a decrease in median daily flows due to urbanization, population growth, and an increase in net withdrawals. In general, the most extreme hydrologic alteration was found in the hot and dry scenario, followed by the climate change scenario.

The major conclusion of this analysis is that flow alteration projections for different future scenarios need to account for all the natural and anthropogenic factors at work in each watershed that significantly and often simultaneously alter flow, sometimes in competing directions. Figure 5 in the Main Report lists many of these factors. Despite similar intensities or spatial distributions in one factor, differences in other factors in otherwise comparable watersheds can change streamflow in divergent ways, eventually leading to differences in flow regimes. Drafts of the maps presented in Figure 15 were presented in the February 2012 webinar and are discussed in more detail in the notes associated with the presentation slides (Appendix J).

6. Development of Flow Alteration-Ecology Relationships

From the initial set of 256 flow metrics and 51 family-level macroinvertebrate metrics, 24 flow metrics and 20 macroinvertebrate metrics were selected as candidates for development of FA-E relationships. They are listed in two supplemental tables at the end of this report. An important criterion was representation of major components of both the flow regime and macroinvertebrate community without much redundancy. Other considerations included: (a) the model's accuracy in reproducing observed flow metric responses to natural and environmental factors, (b) flow metrics that show strong relationships to land/water uses, (c) professional judgment based on literature pertaining to flow alteration effects on biological metrics, some of which is cited in the upcoming section on possible mechanisms underlying FA-E relationships, and (d) flow and macroinvertebrate metrics commonly used by state agencies in the Potomac River watershed.

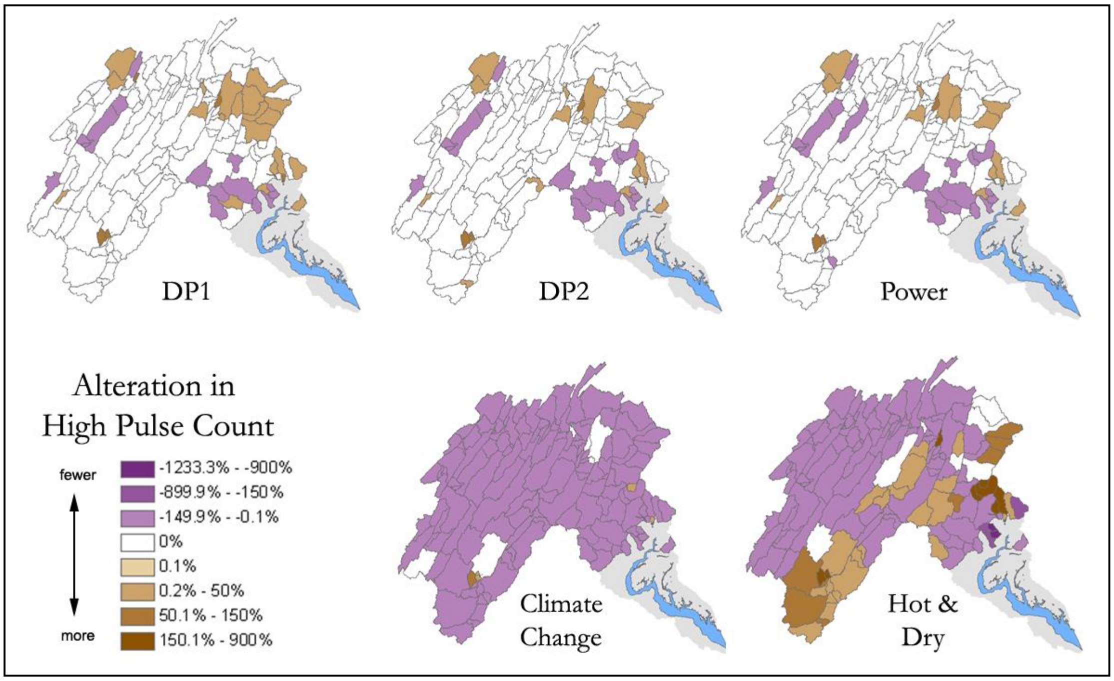

Figure 15a. Future alteration from current conditions in high pulse count under the five future scenarios, displayed spatially by HSPF model river segment.

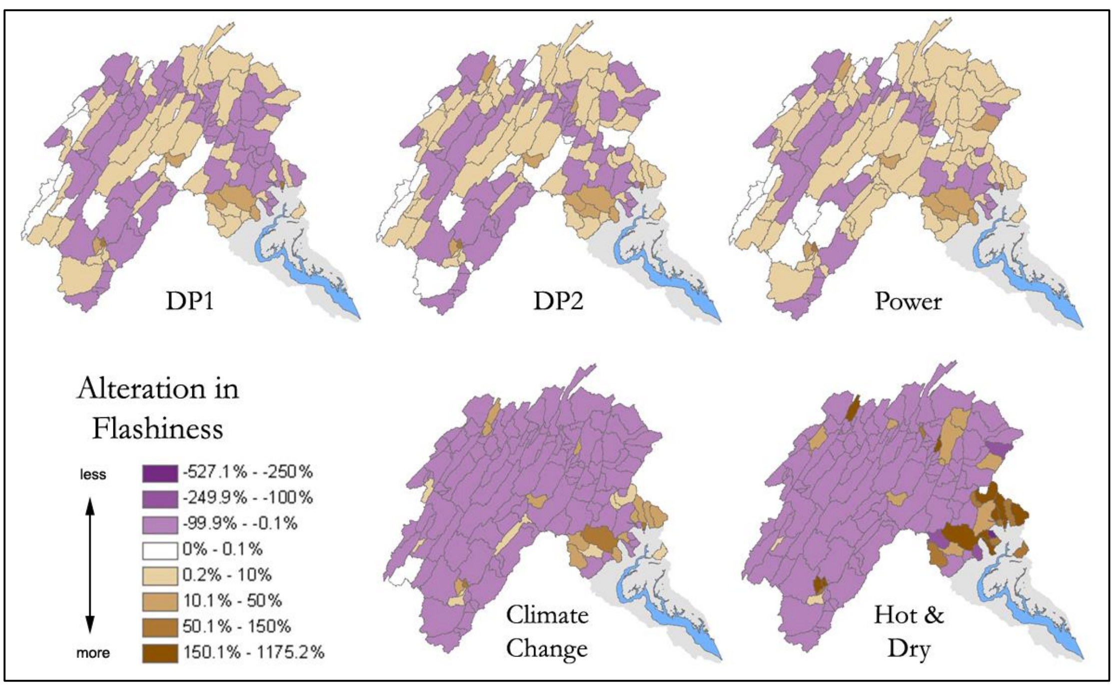

Figure 15b. Future alteration from current conditions in flashiness under the five future scenarios, displayed spatially by HSPF model river segment.

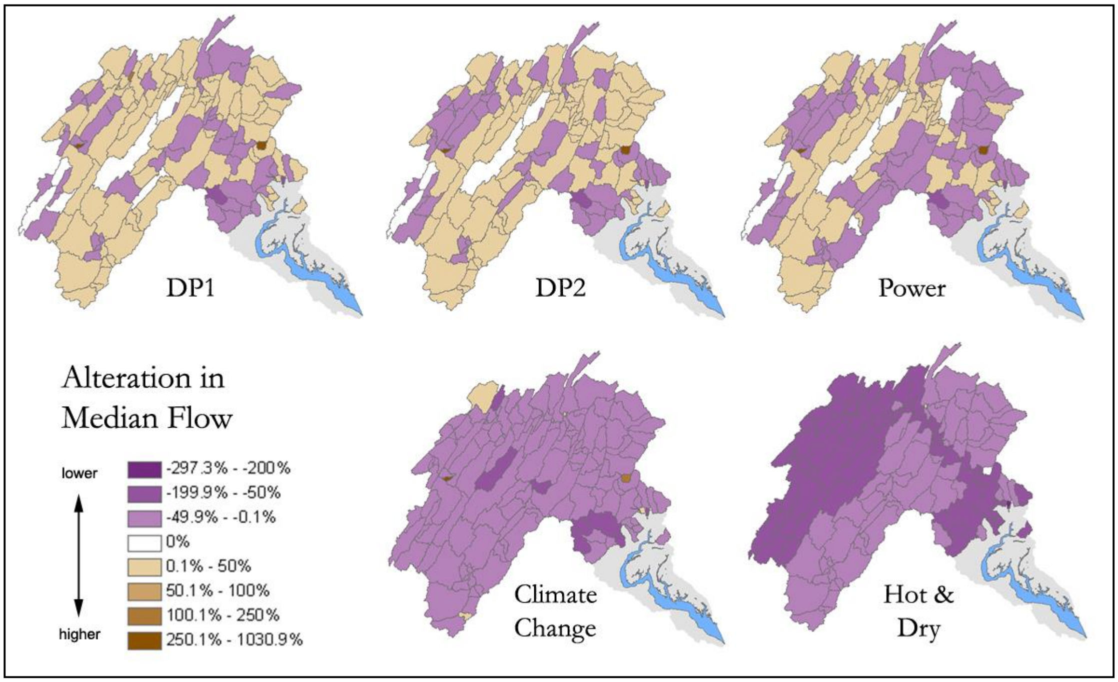

Figure 15c. Future alteration from current conditions in median flow under the five future scenarios, displayed spatially by HSPF model river segment.

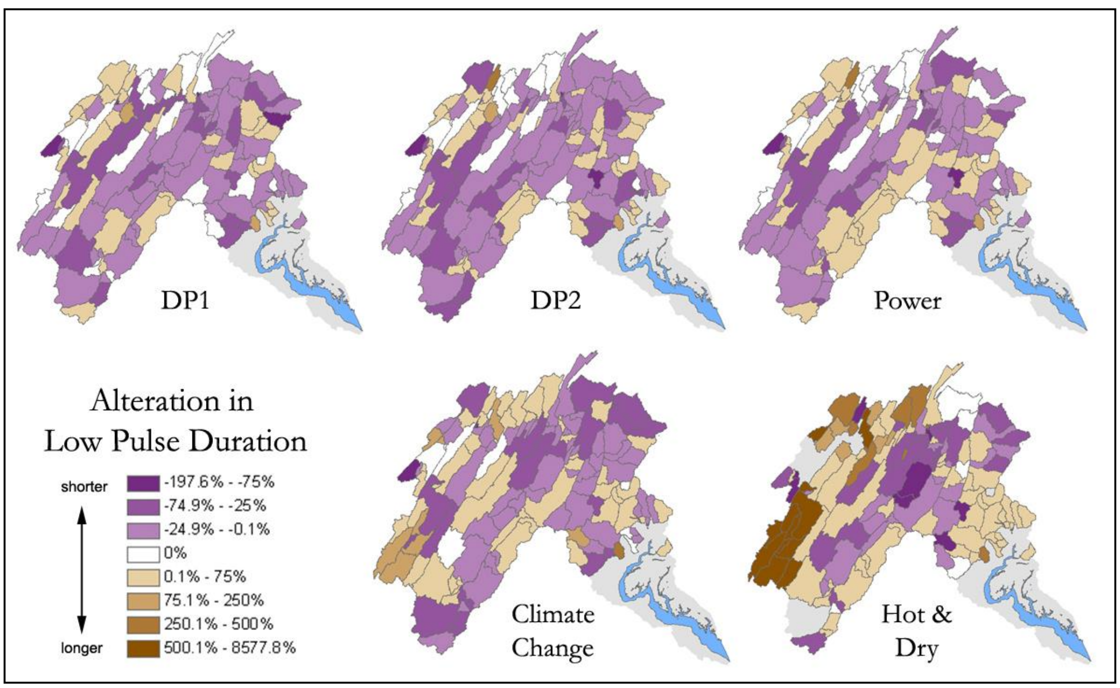

Figure 15d. Future alteration from current conditions in low pulse duration under the five future scenarios, displayed spatially by HSPF model river segment.

Three methods were used to examine macroinvertebrate responses to flow alteration: Pearson correlations, quantile regression plots, and conditional probability plots. Each has its strengths and weaknesses (Table 6). This process continued to narrow the selection of biological and flow metrics.

Pearson correlations

The Pearson correlation method measures strength of the linear relationship between macroinvertebrate metrics and percent alteration in each flow metric. The objective was to identify flow metrics that, when altered, associate with strong biological responses. The Pearson method returns a correlation coefficient (r) value between +1 and −1; 0 indicates no relationship and $\pm$1 indicates a perfect relationship. A positive sign signifies both variables increase together and a negative sign signifies one variable decreases as the other increases. In this analysis, biometric values and not scores were used, so the results do not account for biological differences relating to bioregion and other natural factors ("Stream Classification and Biometric Scoring", above).

Table 7 presents color-coded results of the Pearson correlations. For biological metrics that decrease with increasing stress ("neg"), flow alteration that tends to shift away from baseline values in either the positive (brown) or negative (purple) direction relates to a decrease in the biological metric. For biological metrics that increase with increasing stress ("pos"), flow alteration away from the baseline in either the positive (purple) or negative (brown) direction relates to an increase in the biological metric. The darkest colors indicate the strongest correlation coefficients ($r > 0.45$) and represent good possibilities for developing FA-E relationships. The lightest colors indicate weaker relationships.

Almost all macroinvertebrate metrics correlated strongly with alteration in high magnitude flows, many of the duration and frequency metrics for high and low flow events, and flashiness and rise rate. The exceptions are %Collectors, %Filterers, %Clingers, and taxa richness, each of which correlated weakly with alteration in all of the flow metrics. None of the 20 biometrics correlated strongly with alteration in the median or metrics representing low magnitude flows, frequency of extreme low flow events, or fall rates.

If metric values from the current flow scenario are used rather than flow alteration percentages, the overall strength of the correlation coefficients diminishes and some of the relationships even disappear. Figure 16 illustrates for one flow metric, flashiness, why this happens. This flow metric is significantly affected by both watershed size and %karst (Figure 1 in Appendix F). Current scenario values of the flow metric reflect the influences of these two natural factors and the anthropogenic factors in the watersheds (Figure 16 top panel). Baseline scenario values reflect only the natural factor influences. In some watersheds, values of the flow metric are strongly influenced by the natural factors while in others they are not. The calculation of percent alteration in the flow metric, when each watershed's baseline value is subtracted from its current value and the difference is expressed as a percent of baseline, effectively removes natural variation from the flow metric value and leaves the variation due to anthropogenic factors (Figure 16 bottom panel).

Quantile regressions

Scatter plots that relate biological metrics to alteration in flow metrics, such as the bottom panel in Figure 16, show a lot of biological variability, even in watersheds where current flow conditions are essentially the same as baseline conditions (i.e., percent alteration equals zero). This is because flow alteration is only one of several factors affecting macroinvertebrate communities. The broad

Table 6. Methods used to investigate flow alteration-ecology relationships. Some of their strengths and weaknesses are also listed.

	What the method does	Strength	Weakness
Pearson Correlations†	Quantifies strength of flow alteration impacts on biometrics in the context of all other environmental factors affecting biometrics	• Does not require large dataset	• No control of natural variability in biometrics (i.e., data not classified by bioregion) • Linear regression crosses both negative and positive flow alteration • Relationship affected by other anthropogenic factors impacting biometrics
Quantile Regressions†	Quantifies best possible biological status that can be achieved as flow alteration increases or decreases away from baseline condition	• Other anthropogenic impacts on biometrics do not affect quantile regression much, if at all • Responses to negative and positive flow alteration are separated	• No control of natural variability in biometrics (i.e., data not classified by bioregion) • Quantile regression line may shift if water or habitat quality changes significantly • Linear regression • Requires large dataset
Conditional Probability Plots‡	Quantifies probability of biometric attaining a specific condition (e.g., a status of "fair" or better) at a given level of flow alteration in the context of all other environmental factors affecting biometrics	• Natural variability in biometrics minimized by bioregion classification and biometric scoring • Non-linear relationship (LOESS regression) • Responses to negative and positive flow alteration are separated	• Probability at a given flow alteration level may change if water or habitat quality changes significantly • Requires large dataset

†Uses biometric values
‡Uses scoring

Table 7. Pearson correlations between percent alteration in flow metrics and macroinvertebrate metrics. See Supplemental Tables A and B for metric descriptions. Correlation coefficient values (r) are color coded. For biometric values that are *lower* in degraded conditions ("neg"): brown = as [flow metric] increases from baseline, [biometric value] decreases; purple = as [flow metric] decreases from baseline, [biometric value] decreases. For biometric values that are *higher* in degraded conditions ("pos"): brown = as [flow metric] decreases from baseline, [biometric value] increases; purple = as [flow metric] increases from baseline, [biometric value] increases.

Pearson correlation Dark purple: r > 0.45 White: r = -0.1 to 0.1 Dark brown: r < -0.45	Response to Degradation	1Day Maximum	3Day Maximum	Annual Mean	Median	Q85Seas	August median	Base Flow Index	1Day Minimum	3Day Minimum	7Q10	High Pulse Duration	High Flow Index MH21	High Flow Duration DH17	Flood Free Season	Low Pulse Duration	Extreme Low Flow Dur.	High Pulse Count	High Flow Frequency	Number Reversals	Low Pulse Count	Extreme Low Flow Freq.	Flashiness	Rise Rate	Fall Rate
BIBI	Neg																								
%EPT	Neg																								
%Ephemeroptera	Neg																								
# Ephemeroptera Families	Neg																								
#Sensitive Taxa	Neg																								
Family Hilsenhoff Index	Pos																								
ASPT Modified Index	Pos																								
%Tolerant	Pos																								
%Dominant3	Pos																								
%Chironomidae	Pos																								
Beck's Index	Neg																								
Gold Index	Neg																								
%Gatherer	Pos																								
%Collector	Pos																								
%Scraper	Neg																								
%Filterer	Neg																								
%Swimmer	Neg																								
%Clinger	Neg																								
Shannon Wiener Index	Neg																								
Taxa Richness	Neg																								

distribution of biometric values at a given level of flow alteration shows the combined impact of flow alteration and multiple other stressors. The quantile method applies a linear regression to the upper or lower boundary of a cloud of data points. The envelope created by the regression line represents the best possible biological status that can be achieved as flow alteration increases or decreases away from baseline condition**.**

Alteration in some flow metrics usually occurs in one direction in the ELOHA watersheds (Section 5). For example, flashiness usually changes in a positive direction and high flow duration DH17 in a negative direction, meaning the watershed factors in the study area that alter flow tend to make streams flashier with shorter high flow periods. In the few instances where flashiness and low pulse duration changed in opposite directions, becoming less flashy with longer periods of high flow, some of the sites are associated with sizeable discharges. Alteration in the middle and low magnitude flow metrics occurs about equally in the negative and positive directions. These include the annual mean, median, Q85Seas, August median, baseflow index, and the 1-day and 3-day minimum. The occurrence and extent of flow alteration in positive and negative directions in the other flow metrics is discussed below and depends largely on the combination of flow-altering factors that are present in the ELOHA watersheds.

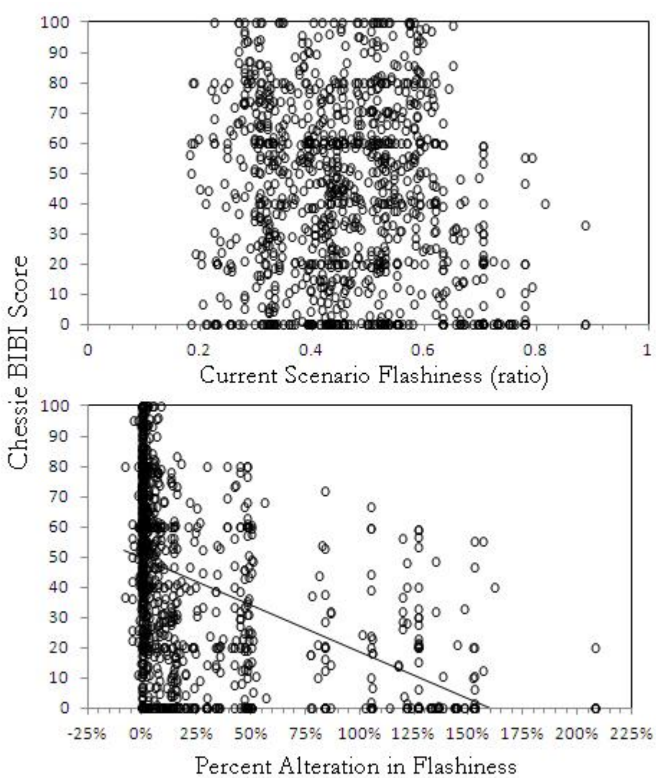

Figure 16. Chessie BIBI plotted against current scenario values of flashiness (upper panel) and against percent alteration in flashiness from baseline (lower panel). A significant relationship (solid line) is found only in the latter.

Quantile regression was used to determine if alteration in the positive or negative direction was statistically significant for combinations of representative flow metrics and macroinvertebrate metrics ($p<0.01$). Those with significant relationships were among the likely candidates for developing FA-E relationships. Again in these analyses, biometric values and not scores were used, so the results do not account for biological differences relating to bioregion and other natural factors. Details of how the quantile method was applied are described in Appendix I.

Macroinvertebrate metrics did not show strong, significant quantile regressions in either the positive or negative direction with the low magnitude flow metric (3-day minimum), frequency of extreme low flow, or median flow. Of the remaining flow metrics, significant quantile regressions occurred most often on the positive side for high pulse count and flashiness and on the negative side for the high flow index MH21, high flow duration DH17, and low pulse duration. The quantile regression results corroborate the Pearson correlation results above for the same metrics. They also indicate if there is an adequate and sufficient amount of data to discern a response to negative and positive alteration and justify development of FA-E relationships.

Conditional probability

The biological scoring approach (Section 4) configures each metric so that values most like those in each bioregion's reference population receive high (fair or better than fair) scores and those most like those in the degraded populations receive low ("poor") scores. Scores are thus directly comparable across bioregion. The probability of a biological metric receiving a fair or better than fair score is then calculated for increments of flow alteration moving either in the positive or negative direction. Details of how conditional probabilities were calculated are described in Appendix I.

Conditional probability curves were developed from the proportion of samples scoring fair or better than fair in each flow alteration increment. An example illustrating positive flow alteration away from baseline is shown in Figure 17. The considerable scatter in the conditional probability points reiterates that environmental factors other than flow alteration impact biological communities, including recent weather events. Recall that biological values are single observations while flow alteration is calculated from baseline and current model scenarios covering a 21 year period (water years 1984-2005). The regression line and its confidence interval nevertheless show a strong tendency to decrease as the amount of flow alteration increases.

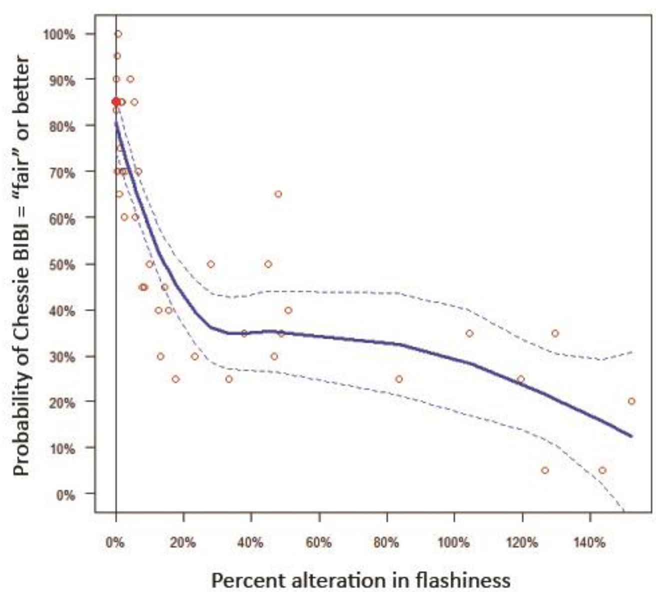

Figure 17. Illustration of a flow alteration-ecology (FA-E) plot.
Open red circles, conditional probabilities that biological status (Chessie BIBI) is "fair" or better for increments of alteration in Flashiness; solid blue line, Loess smoothed regression through the probabilities; dashed blue lines, 0.05 confidence interval around the regression line (the confidence interval around the data points is not shown); solid dark red circle on vertical line at 0 percent alteration, probability of "fair" or better status for just the watersheds meeting reference habitat and water quality criteria.

Some biometric – flow metric combinations show stronger relationships than others but there is a consistent pattern of decreasing probability of fair or better biological status as the amount of flow alteration increases. For many biometric – flow metric combinations, the decrease in the probability of fair or better status is quite steep at low levels of flow alteration. In the next section, the FA-E plots are individually examined and interpreted.

7. References

Apse, C., DePhilip, M., Zimmerman, J., and Smith, M. P. 2008. Developing instream flow criteria to support ecologically sustainable water resource planning and management. Final report to the Pennsylvania Instream Flow Technical Advisory Committee. The Nature Conservancy, Harrisburg, PA.

Astin, L. E. 2006. Data synthesis and bioindicator development for nontidal streams in the interstate Potomac River basin, USA. Ecological Indicators, 6: 664-685.

Archfield, S.A., R.M. Vogel, P.A. Steeves, S.L. Brandt, P.K. Weiskel, and S.P. Garabedian. 2010. The Massachusetts Sustainable-Yield Estimator: A Decision-support Tool to Assess Water Availability at Ungaged Stream Locations in Massachusetts. Scientific Investigations Report. USGS.

Buchanan, C., K. Foreman, J. Johnson, and A. Griggs. 2011. Development of a Basin-Wide Benthic Index of Biotic Integrity for Non-Tidal Streams and Wadeable Rivers in the Chesapeake Bay Watershed: Final Report to the Chesapeake Bay Program Non-Tidal Water Quality Workgroup. ICPRB Report 11-1. Prepared for the US Environmental Protection Agency, Chesapeake Bay Program. http://www.potomacriver.org/cms/publicationspdf/ICPRB11-01.pdf.

Chesapeake Bay Program. 2010. Chesapeake Bay Program Environmental Modeling - Backgrounder. Annapolis, MD. http://www.chesapeakebay.net/about/programs/modeling/53.

Donigian, A.S. 2002. Watershed model calibration and validation-the HSPF experience. In Specialty Conference Proceedings: Water Environment Foundation, 30. Phoenix, AZ.

Foreman, K., C. Buchanan, and A. Nagel. 2008. Development of ecosystem health indexes for non-tidal wadeable streams and rivers in the Chesapeake Bay basin. Unpublished report to the Chesapeake Bay Program Non-Tidal Water Quality Workgroup. December 5, 2008.

Gassman, P.W., M.R. Reyes, C.H. Green, and J.G. Arnold. 2007. The soil and water assessment tool: Historical development, applications, and future research directions. Ames, Iowa: Center for Agricultural and Rural Development, Iowa State University.

Gordon, N. D., T. A. McMahon, and B. L. Finlayson. 1992. Stream Hydrology: an Introduction for Ecologists. John Wiley & Sons, 542 pp.

Henriksen, J. A., Heasley, J., Kennen, J.G., and Niewsand, S. 2006. Users' manual for the hydroecological integrity assessment process software (including the New Jersey Assessment Tools): U.S. Geological Survey, Biological Resources Discipline, Open File Report 2006-1093, 71 p. http://www.fort.usgs.gov/Products/Software/NATHAT/

Konrad, C. 2011. Principal technical decision points for developing regional environmental flow standards. Unpublished draft, March 1, 2011.

Kudlas, S. 2009. Models, Uncertainty, and Cumulative Impacts on Water Resources: The Virginia DEQ Approach to Water Resource Modeling presented at the Southeast States Water Resources Management and Planning Forum, Richmond, VA. http://sogweb.sog.unc.edu/Water/images/7/72/Vadeq_approach_v02-Kudlas.pdf.

Linker, L. C., G. W. Shenk, R. L. Dennis, and J. S. Sweeney. 1999. Cross-media models of the Chesapeake Bay watershed and airshed. Annapolis, MD: Chesapeake Bay Program Office. http://archive.chesapeakebay.net/pubs/948.pdf.

Lumb, A.M., R.B. McCammon, and J.L. Kittle. 1994. Users manual for an Expert System (HSPEXP) for calibration of the Hydrological Simulation Program: Fortran. Water Resources Investigations Report. Reston, VA: USGS.

Nash, J.E., and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models, Part 1: A discussion of principles. Journal of Hydrology 10:282-290.

Olivero, A. P., and M. G. Anderson. 2008. Northeast aquatic habitat classification system. Report prepared by The Nature Conservancy Eastern Regional Office in collaboration with the Northeast Association of Fish and Wildlife Agencies. http://www.fws.gov/filedownloads/ftp_gis/R5/GOMP/Jed/Prints/6/3%20NE%20Habitat%20Classification%20-%20Aquatic.pdf

Poff, N. L., B. D. Richter, A. H. Arthington, S. E. Bunn, R. J. Naiman, E. Kendy, M. Acreman, C. Apse, B. P. Bledsoe, M. C. Freeman, J. Henriksen, R. B. Jacobson, J. G. Kennen, D. M. Merritt, J. H. O'Keeffe, J. D. Olden, K. Rogers, R. E. Tharme, and A. Warner. 2010. The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. Freshwater Biology 55(1):147–170.

Rode, M., and U., Suhr. 2006. Uncertainties in selected surface water quality data. Hydrology and Earth System Sciences 3:2991-3021.

Rossman, L.A. 1990. DFLOW User's Manual. Washington, DC: EPA.

Schultz, C.L., S.A. Ahmed, R. Mandel, and H. Moltz. In review. Implementation of a nonlinear groundwater storage-discharge algorithm in HSPF model of the upper Potomac River basin.

Semmens, D., M., Hernandez, D.C. Goodrich, and W.G. Kepner. 2006. A retrospective analysis of model uncertainty for forecasting hydrologic change. Proceedings of iEMSs 3rd Biennial

Meeting: Summit on Environmental Modeling and Software, International Environmental Modeling and Software Society, Burlington, VT.

The Nature Conservancy. 2007. Indicators of Hydrologic Alteration, Version 7 User's Manual. http://www.nature.org/initiatives/freshwater/conservationtools/.

USEPA (U.S. Environmental Protection Agency). 2010. Chesapeake Bay Phase 5.3 Community Watershed Model. EPA 903S10002 - CBP/TRS-303-10. U.S. Environmental Protection Agency, Chesapeake Bay Program Office, Annapolis MD. December 2010.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The River Continuum Concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130-137.

Waldron, M. C., and S. A. Archfield. 2006. Factors Affecting Firm Yield and the Estimation of Firm Yield for Selected Streamflow-dominated Drinking-water-supply Reservoirs in Massachusetts, US Scientific Investigations Report. USGS. http://pubs.usgs.gov/sir/2006/5044/pdf/SIR2006-5044.pdf.Ward, J. V. 1991. Aquatic Insect Ecology. Part I, Biology and Habitat. Wiley. 456 pgs.