

# Statistical Identification of Hydrochemical Response Units for Hydrologic Monitoring and Modeling in Maryland

by Stephen D. Preston

In cooperation with the



Maryland Department of Natural Resources



## Abstract

In support of Maryland's efforts to develop statewide water-quality management plans, a statistical analysis was performed to identify a set of representative and relatively homogeneous areas referred to as **Hydrochemical Response Units (HRUs)**. The State intends to select representative areas within each hydrochemical response unit for monitoring and model development, and then apply the model to the entire unit. To identify hydrochemical response units, cluster analysis was applied to 1,136 digitally defined watershed units. Basin characteristics including land cover, soil type, slope, and geology were determined for each watershed unit and a clustering algorithm was applied to the data sets. A total of 11 hydrochemical response units were identified by the analysis. Major features that were important in distinguishing different areas of the State include: 1) upland and lowland Coastal Plain settings; 2) igneous, shale, and carbonate geology; and 3) urban land cover. The hydrochemical response units described in this report are considered to be an initial classification of watersheds in Maryland that can be refined as geographic data sets are improved and additional hydrologic data are collected.

## Introduction

Regulation of point-source discharges and improvement of treatment processes have reduced the quantities of contaminants discharged directly to streams. As a consequence, nonpoint sources now represent the largest source of contaminants in many areas. To develop management plans for dispersed sources of contaminants in large areas, a scientific framework is necessary for integrating information on factors that affect water quality. Water-quality models can provide such a framework.

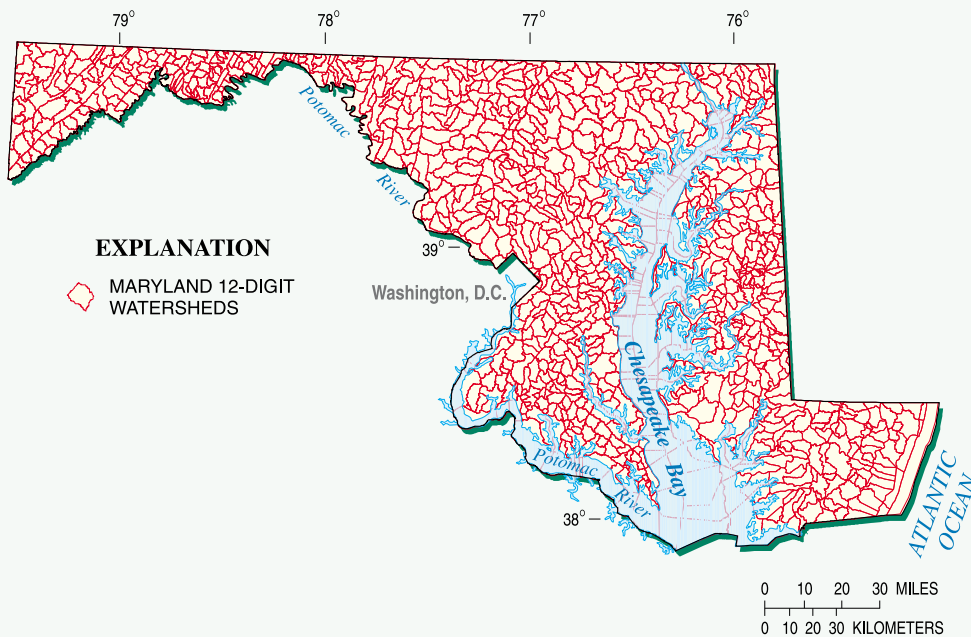
A water-quality model serves many purposes in the development of regional water-quality management plans. A model provides a framework for relating sources to downstream loads and for accounting for processes that affect contaminants as they are

transported by streams and other water bodies. In this way, the model provides a basis for determining which sources are most important to loads at specific stream locations in the watershed. Such information can be used to target management resources where they are most likely to benefit water quality, and to design management actions for the specific type of source that is present in a targeted area.

Another purpose of a model is to provide a means of spatially extrapolating information related to water quality. Water-quality and related types of data are commonly available in limited numbers and are often distributed inconsistently in space and time. Water-quality measurements are expensive and there is incentive for government agencies to minimize the amount of data that is collected. Sparse or poorly distributed water-

quality data complicate the development of regional management plans. If the data available in a single area are adequate for model calibration and validation, however, then the model can often be used to extrapolate water-quality information to similar areas and thus aid in the development of water-quality management plans for an entire region. Ideally, water-quality data collection and modeling programs should be designed simultaneously to maximize the value of the data and ensure that collected data are adequate for development of the model.

It is important to identify areas with similar geographic characteristics when a model is used to extrapolate water-quality information as part of the effort to develop regional management plans. The hydrology and water quality of such areas would be expected to respond similarly to forcing



**Figure 1.** Twelve-digit watershed delineations for Maryland.

functions like precipitation and contaminant discharge, and a model calibrated in one part of the area would likely be appropriate for predicting water-quality response in the entire area. The appropriate bases for identifying areas of similar response are not well defined. Basin characteristics such as land cover, soil type, and geology are known to be important in defining hydrologic and water-quality response. Areas with similar land cover, soil, and geology can be defined as “Hydrochemical Response Units” (HRUs) for which models can be applied for developing water-quality management plans.

The State of Maryland is developing regional water-quality management plans to address nonpoint-source nutrient and sediment loading of streams. However, available water-quality data are inadequate for detailed model calibration for many areas of the State. In response to the need for water-quality data, the Maryland Department of Natural Resources (MDDNR), in cooperation with the U.S. Geological Survey (USGS), has implemented a water-quality data collection and water-quality modeling program for the entire State. To limit the size (and the cost) of the monitoring and modeling program, MDDNR and USGS have decided to divide the

State into a relatively small number of units (approximately 8-12) on the basis of geographic characteristics. One or more subbasins in each unit would be selected for monitoring (in the absence of adequate existing data) and the developed models would then be used for evaluations across the entire geographic unit represented by the subbasin. In addition, subsequent strategic data collection could be used to improve the designation of units or to improve the scientific understanding of the processes controlling nonpoint-source loading. It is intended that this process will continue until the entire State has been evaluated.

### Purpose and Scope

This report describes the initial results of a method developed to identify areas in Maryland with consistent basin characteristics that can be considered HRUs for water-quality monitoring and modeling purposes. The general approach for identifying an initial set of HRUs includes statistically classifying watershed units in the State based on their land cover, soil type, and geology. This report describes the data and statistical methods used, and the initial results of a classification of watershed units in Maryland.

## Study Methods

The methods described in this report involve statistically classifying watershed units into HRUs on the basis of their geographic characteristics. A watershed segmentation scheme has already been developed for Maryland (Maryland Department of Natural Resources, 1997) and was used to define watershed units for the State. Basin characteristics are quantified for each watershed unit by overlaying land cover, soil type, and geology spatial coverages with the digital watershed boundaries that define the watershed units. Statistical clustering algorithms are then used to group the watershed units based on their characteristics, which in turn allows separation of the population of watershed units into a smaller number of groups that are assumed to be relatively homogeneous HRUs.

### Geographically Defined Data Bases

Maryland’s statewide digital watershed file (SHED1997) (Maryland Department of Natural Resources, 1997) was used to define watershed units that could be aggregated to form HRUs. The State watershed boundary data set originated from nondigital delineations that were developed for planning purposes of the Maryland Department of the Environment (MDE) and the Maryland Office of Planning (MOP). Watersheds were initially delineated for all third-order streams in Maryland as defined on 7.5-minute USGS topographic quadrangle maps, resulting in a set of 138 watershed units for the State. The data set was later refined by MDDNR by subdividing the 138 watershed units to form a new data set that contains 1,136 watershed units. Each watershed in the initial set of 138 units was assigned an 8-digit identification code, and the data set is referred to as the “8-digit watershed file.” Similarly, the refined set of 1,136 watershed units are identified by a 12-digit code and the file is referred to as the “12-digit watershed file.” The data base SHED1997 contains both sets of identifiers as attributes so that watershed

units can be accessed for either scale by aggregating the 12-digit units to the 8-digit units. The 12-digit watersheds were used in this study to maximize the level of spatial detail (fig. 1).

To account for the potential effects of land cover in determining hydrochemical response, land-cover data were obtained for the State from MOP for the year 1994 (Maryland Office of Planning, 1996). The 1994 Maryland land-cover data base includes a total of 27 land-cover classes. Such level of detail in land-cover classification is beyond the scope of this analysis, however, and the number of classes was reduced by aggregating the 27 land-cover classes to a total of 6 classes (table 1). The aggregated land-cover classes include urban, residential, agriculture, forest, wetlands, and barren land (fig. 2). These classes of land cover are included in this analysis because they are considered to be the most important factors in distinguishing land-cover effects on hydrology and water quality in Maryland.

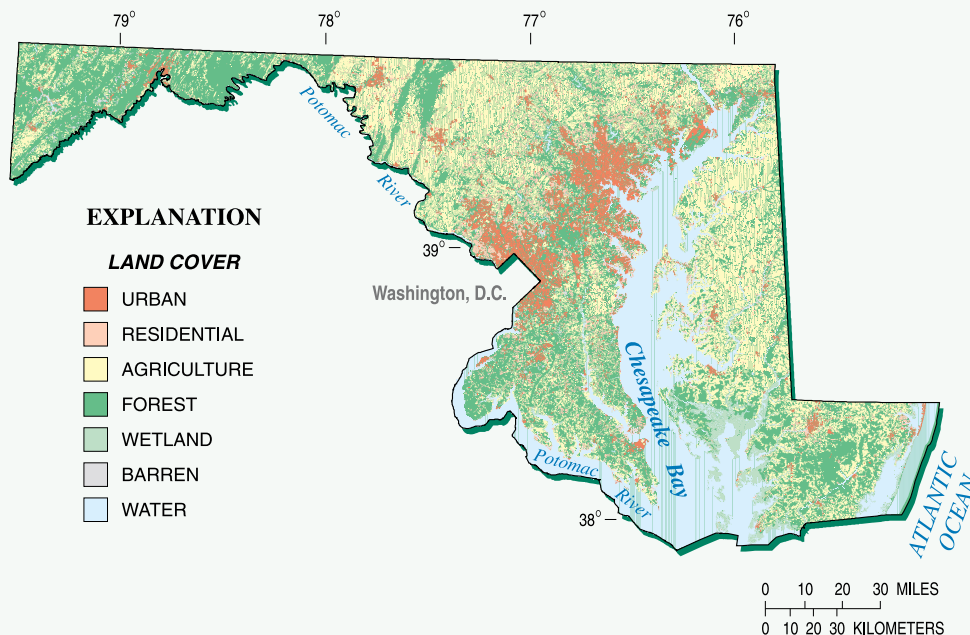


Figure 2. Generalized land cover in Maryland, 1994.

Soil characteristics are known to affect hydrologic response by determining infiltration rates, and, consequently, potential for runoff. A soils coverage developed by MOP

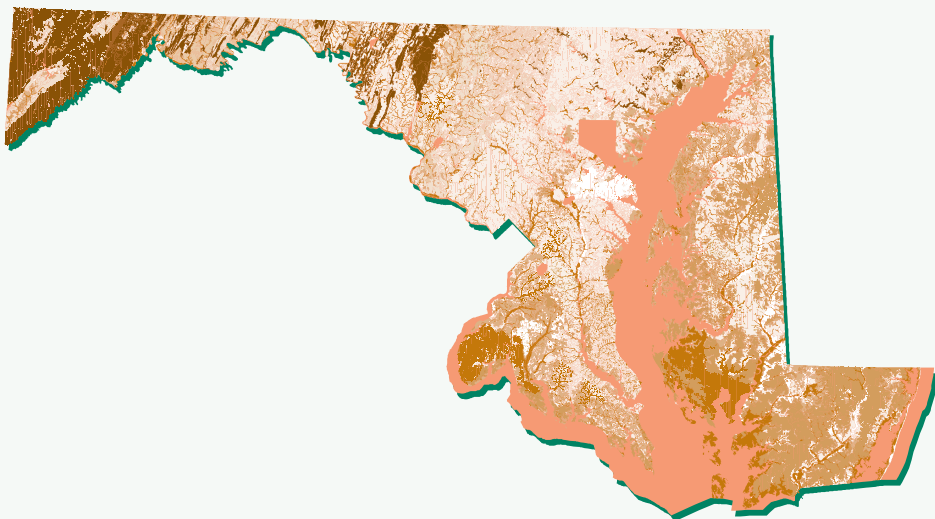
(Maryland Office of Planning, 1973) was used in the development of HRUs for Maryland. The Maryland soils data base is a coverage that classifies soil type according to Natural Soils Groups, in which soils are classified primarily on the basis of drainage potential, permeability, depth, rockiness, and slope. Each soil group is assigned a capital letter from A to H, and generally, soils are progressively more poorly drained from group A to group G; group H comprises rocky soils (fig. 3). Each group is further subdivided, but for the purposes of this analysis, only the percentages of the major soil groups (A-H) were included as variables in determining watershed unit similarities. Slope is indicated in the Natural Soils Groups through the use of small letters a, b, or c, which designate slope classes of 0-8 percent, 8-15 percent, and greater than 15 percent, respectively. To include the effects of slope in this analysis, a slope index was calculated by specifying the center value of each slope class (such as 4 percent, 12 percent, and 20 percent) and calculating an areally weighted average slope value for the watershed unit. This is not intended to be a precise measure of slope, but is assumed to provide a relative index of slope among watershed units.

Like soils, geology is an impor-

Table 1. Summary of aggregated and original land-cover classes for Maryland (Maryland Office of Planning, 1996)

AGGREGATED LAND-COVER CLASS	ORIGINAL LAND-COVER CLASS
URBAN	MEDIUM DENSITY RESIDENTIAL HIGH DENSITY RESIDENTIAL COMMERCIAL INDUSTRIAL INSTITUTIONAL
RESIDENTIAL	RESIDENTIAL LOW DENSITY RESIDENTIAL OPEN LAND LARGE LOT SUBDIVISION - AGRICULTURE LARGE LOT SUBDIVISION - FOREST
AGRICULTURE	AGRICULTURE CROPLAND PASTURE ORCHARDS, VINEYARDS ROW AND GARDEN CROPS FEEDING OPERATIONS AGRICULTURAL BUILDINGS, STORAGE FACILITIES
FOREST	FOREST DECIDUOUS FOREST EVERGREEN FOREST MIXED DECIDUOUS AND EVERGREEN FOREST BRUSH, CUT-OVER TIMBER, ABANDONED FIELDS
WETLANDS	WETLANDS
BARREN	EXTRACTIVE LAND BARREN LAND BEACHES BARE LAND





tant factor in determining watershed hydrochemical response because it often determines the fraction of water that is transmitted through groundwater pathways to streams. For this reason, geologic data were used in this analysis to help distinguish watershed units on the basis of their potential to generate runoff or transmit ground water. Geologic data were obtained from the Maryland Geological Survey (MGS) surficial geology map (scale 1:250,000) for the State (Maryland Geological Survey, 1968). The map was converted to digital format by the USGS in 1995, and the coverage was used to define geologic characteristics for each watershed unit.

The geologic map created by MGS is quite detailed in the number of geologic formation classes delineated. Because the level of detail is beyond the scope of this analysis, the number of classes was reduced from 89 to 10 by aggregating to more general geologic classes (table 2). Aggregated classes (fig. 4) are based on major rock types such as igneous/metamorphic, metasedimentary, and carbonate. Aggregated classes also include a number of Coastal Plain unconsolidated sediment groups that may be useful in distinguishing different hydrologic characteristics of Coastal Plain watersheds.

### Statistical Analysis Methods

The statistical methodology used for aggregating watershed units into HRUs is referred to as cluster analysis. Cluster analysis is a multivariate technique that is designed to find “natural groupings” of objects by partitioning them on the basis of measures of their characteristics (Chatfield and Collins, 1980; SAS Institute Inc., 1985). In

this case, cluster analysis is being applied to find natural groupings (HRUs) of watershed units on the basis of geographic characteristics that may affect hydrochemical response.

Many types of cluster analysis have been developed to account for complications in analyzing various types of data. In general, cluster analysis techniques are either agglomerative or divisive. Agglomerative techniques begin by treating each observation as a single group and progressively combining groups based on the smallest Euclidean distance. Conversely, divisive techniques begin by treating all observations as a single group and progressively dividing groups by the greatest Euclidean distance. In either case, a hierarchical tree can be developed to illustrate the separation of groups from 1 to  $n$  (the number of observations) as a function of distance or dissimilarity. From the hierarchical tree, the appropriate number of groups must be selected on the basis of the purpose of the analysis and statistical considerations.

Many variations in techniques are available within agglomerative and divisive types of cluster analysis. Most differ in the way that distance is

**Table 2. Comparison of aggregated and original geologic classes for Maryland**  
(Maryland Geological Survey, 1968)

AGGREGATED GEOLOGIC CLASS	ORIGINAL GEOLOGIC CLASS
<b>IGNEOUS / METAMORPHIC</b>	ELLICOTT CITY GRANODIORITE (Pze) GUILFORD QUARTZ MONZONITE (Pzg) QUARTZ GABBRO AND QUARTZ DIORITE GNEISS (Pzdg) GUNPOWDER GRANITE (Pzgg) KENSINGTON QUARTZ DIORITE (Pzk) MUSCOVITE QUARTZ MONZONITE GNEISS (Pzmg) NORBECK QUARTZ DIORITE (Pzn) PEGMATITE DIKES (Pzp) PORT DEPOSIT GNEISS (Pzpd) RELAY QUARTZ DIORITE (Pzr) WOODSTOCK QUARTZ MONZONITE (Pzw) BALTIMORE GABBRO COMPLEX (bgb) GEORGETOWN MAFIC COMPLEX (gm) JAMES RUN GNEISS (jg) METAGABBRO AND AMPHIBOLITE (mgb) BALTIMORE GNEISS (pCgb) CATOCTIN METABASALT (pCc) GRANODIORITE AND BIOTITE GRANITE GNEISS (pCg) METARHYOLITE AND ASSOCIATED PYROCLASTIC SEDIMENTS (PCmv) SAMS CREEK METABASALT (scm) ULTRAMAFIC AND GABBROIC ROCKS (ug) ULTRAMAFIC ROCKS (um) VOLCANIC COMPLEX OF CECIL COUNTY (vc) DIABASE SILLS AND DIKES (Fd)
<b>METASEDIMENTARY</b>	CARDIFF METACONGLOMERATE (cc) JAMSVILLE FORMATION (jf) MARBURG SCHIST (ms) LIBERTYTOWN METARHYOLITE (lmr) PEACH BOTTOM SLATE (pbs) SETTERS FORMATION (sf) URBANA FORMATION (uf) SWIFT RUN FORMATION (pCsR) BOULDER GNEISS (wbg) LOWER PELITIC SCHIST (wlps) METAGRAYS WACKE (wmg) WISSAHICKON FORMATION (undivided) (wu) UPPER PELITIC SCHIST (wups) COCKEYSVILLE MARBLE (cm) SILVER RUN LIMESTONE (srl) WAKEFIELD MARBLE (wm)
<b>CHILHOWEE GROUP</b>	ANTIETAM FORMATION (Ca) HARPERS FORMATION (hf) LOUDOUN FORMATION (lf) SUGARLOAF MOUNTAIN QUARTZITE (sq) WEVERTON FORMATION (wf)
<b>CARBONATE</b>	ELBROOK LIMESTONE (Ce) FREDERICK LIMESTONE (Cf) TOMSTOWN DOLOMITE (Et) HELDERBERG FORMATION / KEYSER LIMESTONE (DShk) CONOCOHEAGUE LIMESTONE (OCc) GROVE LIMESTONE (OCg) CHAMBERSBURG LIMESTONE (Ocs) PINESBURG STATION DOLOMITE (Ops) ROCKDALE RUN FORMATION (Orr) STONEHENGE LIMESTONE (Os) TONOLOWAY LIMESTONE (Stl) GREENBRIER FORMATION (Mg)
<b>TRIASSIC LOWLANDS</b>	GETTYSBURG SHALE (Frg) NEW OXFORD FORMATION (Ferno), (Ftrlc), (Ftrqc)
<b>APPALACHIAN VALLEY AND RIDGE</b>	WAYNESBORO FORMATION (Ewb) CHEMUNG FORMATION, PARKHEAD SANDSTONE (Dch) HAMPSHIRE FORMATION (Dh) HAMILTON GROUP (Dhn) RIDGELEY SANDSTONE, SHRIVER CHERT (Do) JUNIATA FORMATION (Oj) MARTINSBURG FORMATION (Om) MAUCH CHUNK FORMATION (Mmc) POCONO GROUP (Mp) ALLEGHENY FORMATION, POTTSVILLE FORMATION (Ppap) CONEMAUGH FORMATION (Pc) DUNKARD GROUP (Pd) MONONGAHELA FORMATION (Pm) ROCHESTER SHALE, KEEFER SANDSTONE, ROSE HILL FORMATION (Sc) MCKENZIE FORMATION (Sm) TUSCARORA SANDSTONE (St) WILLS CREEK SHALE, BLOOMSBURG FORMATION (Swb)
<b>COASTAL PLAIN - MARINE UNITS</b>	MAGOTHY FORMATION (Km) MATAWAN FORMATION (Kma) MONMOUTH FORMATION (Kmo) POTOMAC GROUP (Kp) ST. MARYS FORMATION (Ts) CHOPTANK FORMATION (Tch) CALVERT FORMATION (Tc) AQUIA FORMATION (Ta) NANJEMOY FORMATION (Tn)
<b>COASTAL PLAIN - UPLAND DEPOSITS</b>	UPLAND DEPOSITS (WESTERN SHORE) (QTu) UPLAND DEPOSITS (EASTERN SHORE) (Qu)
<b>COASTAL PLAIN - LOWLAND DEPOSITS</b>	LOWLAND DEPOSITS (Ql)
<b>COASTAL PLAIN - UNDIVIDED</b>	QUATERNARY DEPOSITS (undivided) (Qdu)

defined, and most were developed to counteract specific problems that arise in cluster analysis. Each of the techniques has advantages and disadvantages and there is no well-defined basis for selecting one over another. For the purposes of this analysis, the centroid method was selected because of its robustness to outliers. The centroid method is an agglomerative hierarchical technique that separates clusters on the basis of the Euclidean distance between their centroids (SAS Institute Inc., 1985). The use of centroids minimizes the influence of any single observation and any potential outlier. A similar study designed to classify drainage areas in central Idaho also used the centroid clustering method (Lipscomb, 1998).

To perform cluster analysis to define HRUs within Maryland, the centroid clustering method was applied to the geographic data sets for the 12-digit watershed units. Areal percentages of each of the generalized geographic characteristics in each of the watershed units were calculated and used as input to the clustering algorithm. Prior to performing cluster analysis, it was decided to limit the number of HRUs to approximately 10 in order to allow cycling of monitoring and modeling programs across the State within a reasonable timeframe. Assuming that a new monitoring site will be established each year, every HRU would be monitored within a 10- to 15-year timeframe. Statistical diagnostics were used to verify that approximately 10 HRUs was a reasonable number to account for most of the variability in the data sets. Diagnostics such as the cubic clustering criterion, the pseudo  $F$  and the pseudo  $t^2$  statistics (SAS Institute Inc., 1985; Lipscomb, 1998) provide guidance for determining the optimal number of clusters to account for the variability in the data.

### Results of Cluster Analysis

Results of the application of cluster analysis to the Maryland watershed data are shown in figure 5. In all cases, clusters appear in the order that

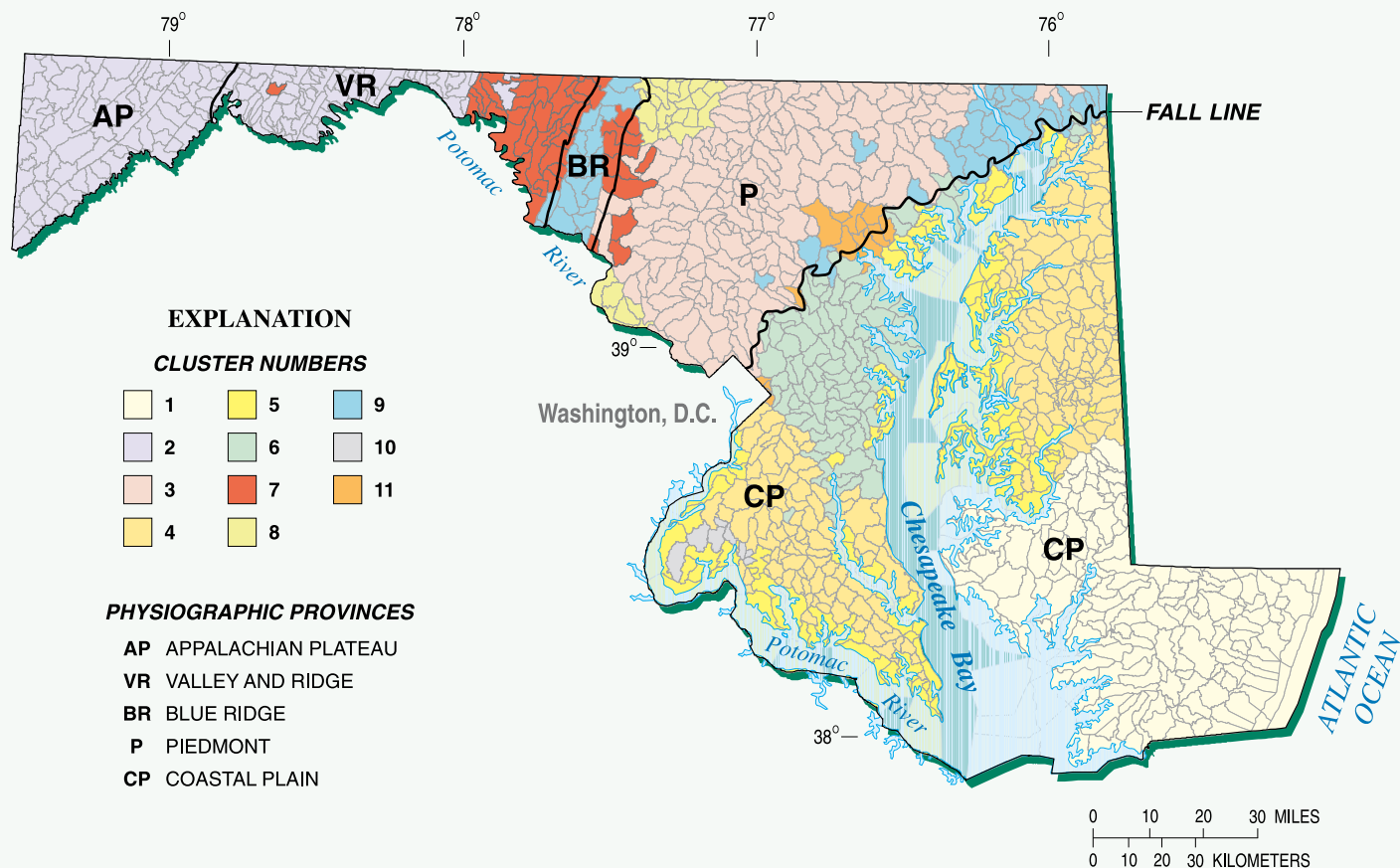


Figure 5. Results of cluster analysis as applied to geographic data sets for defining hydrochemical response units (HRUs).

they were distinguished so that the first cluster is the most discernible group and the last cluster is the least discernible. The dominant characteristics of the watersheds in each cluster are summarized in table 3. Listed in the table are the number of watershed units and mean slope (percent) for each cluster, followed by the dominant land cover, soil type, and geology, and the mean percentages of those geographic features for watersheds in each cluster.

Statistical diagnostics initially indicated that the optimal number of clusters for differentiating geographic features in Maryland was 13. The last two clusters (12 and 13), however, consisted of just one watershed unit each. These two watershed units are located along the coastal margins of the Chesapeake Bay and have unique combinations of geographic characteristics. Instead of considering these two small areas as distinct HRUs, they

were treated as outliers and eliminated from the analysis. Thus, a total of 11 clusters were identified as the final number of distinct HRUs.

The first three clusters are the most easily discernible and tend to follow physiographic province boundaries (fig. 5). Cluster 1 is the most distinct group and is made up of watershed units on the lower Eastern Shore. Cluster 1 watershed units have primarily forest and agriculture land cover, but have the largest percentage of wetlands of all groups (approximately 14 percent on average). Soils are poorly drained and are mostly in the F and G groups, which reflects the presence of wetlands. Surficial geology is entirely undivided Quaternary Deposits of sands and gravels. Cluster 2 is the second most distinct group and is made up of watershed units in the Appalachian Plateau and the Valley and Ridge Physiographic Provinces. Watershed units in this group are pri-

marily forested, have the greatest slopes of all groups, have well drained, rocky soils, and have geologic features that are composed primarily of sandstones and shales. Cluster 3 is the third most distinct group and is made up of watershed units in the Piedmont Physiographic Province. These watershed units have primarily agriculture and forest land covers, but have a significant amount of urban (11 percent) and residential (14 percent) areas. Soils are well drained (B and C) and the surficial geology of cluster 3 is primarily metasedimentary deposits mostly composed of micaceous schists related to the Wissahickon Formation.

The second three clusters (4, 5 and 6) are all in the Coastal Plain Physiographic Province. Cluster 4 is composed primarily of watersheds underlain by well drained upland deposits. Land cover is primarily agriculture and forest, soils are mostly well drained (B) and surficial geology

**Table 3. Summary of dominant watershed characteristics in clusters**  
[% , percent]

CLUSTER	NUMBER OF WATERSHED UNITS	SLOPE (%)	DOMINANT LAND COVER	DOMINANT SOIL TYPES	DOMINANT GEOLOGIC CLASSES
1	172	4	FOREST (43%) AGRICULTURE (36%)	F (40%) B (15%) G (15%)	COASTAL PLAIN - UNDIVIDED (99%)
2	170	16	FOREST (74%)	H (44%) C (25%) D (15%)	APPALACHIAN VALLEY AND RIDGE (94%)
3	171	8	AGRICULTURE (47%) FOREST (28%)	B (67%) C (23%)	METASEDIMENTARY (81%)
4	214	6	AGRICULTURE (45%) FOREST (43%)	B (48%) E (20%) F (13%)	COASTAL PLAIN - UPLAND DEPOSITS (69%) COASTAL PLAIN - MARINE UNITS (23%)
5	112	5	FOREST (39%) AGRICULTURE (34%)	B (20%) E (17%) F (16%)	COASTAL PLAIN - LOWLAND DEPOSITS (76%)
6	89	8	FOREST (44%) URBAN (21%) AGRICULTURE (21%)	B (62%) A (12%)	COASTAL PLAIN - MARINE UNITS (76%)
7	44	9	AGRICULTURE (55%) FOREST (29%)	B (51%) H (25%)	CARBONATE (65%) CHILHOWEE GROUP (18%)
8	26	6	AGRICULTURE (77%)	C (52%) D (18%) B (12%)	TRIASSIC LOWLANDS (93%)
9	46	9	AGRICULTURE (43%) FOREST (41%)	B (63%) H (15%)	IGNEOUS / METAMORPHIC (79%)
10	8	5	FOREST (71%) AGRICULTURE (18%)	G (85%)	COASTAL PLAIN - UPLAND DEPOSITS (65%) COASTAL PLAIN - MARINE UNITS (21%)
11	14	5	URBAN (77%)	B (27%) E (14%)	COASTAL PLAIN - MARINE UNITS (41%) IGNEOUS / METAMORPHIC (37%)

is primarily sand and gravel. Some marine deposits are also exposed in watersheds where streams have cut through the upland deposits to form ravines or valleys. Cluster 5 is composed primarily of watersheds with poorly drained lowland deposits. Land cover is primarily forest and agriculture, but a significant amount of urban (9 percent) and residential (9 percent) areas are also present. Soils range from well drained to poorly drained and a significant amount of wetland area (6 percent) is present. Cluster 6 is distinguished by a relatively high fraction of urban and residential area and by the dominance of geologic features defined by marine deposit outcrops. Forest is the dominant land cover for these watersheds, but the urban and residential areas of Prince Georges and Anne Arundel Counties contribute to a relatively high percentage of these land covers (21 percent and 11 percent, respectively). Soils

are well drained and consist primarily of types B (62 percent) and A (12 percent). The geology of the watersheds in cluster 6 is primarily marine deposit outcrops of glauconitic fine sands and clays.

The remaining clusters are representative of diverse watershed characteristics that occur throughout central Maryland. Cluster 7 is distinguished primarily by carbonate geology that occurs mostly in the Great Valley Physiographic Province. Land cover in cluster 7 is primarily agriculture (55 percent) and soils are well drained (50 percent type B) or rocky (25 percent type H). Cluster 8 is distinguished primarily by Triassic Lowlands geology, which is mostly shales that may affect runoff characteristics. Land cover is primarily agriculture (77 percent), and soils tend to be moderately to well drained (types C, D, and B). Cluster 9 is distinguished by a high percentage of area underlain by igneous and meta-

morphic rocks (79 percent) and includes watersheds in the Blue Ridge and the Piedmont Physiographic Provinces. Land cover is predominantly agriculture (43 percent) and forest (41 percent) and soils are well drained (63 percent type B) or rocky (15 percent type H). Cluster 10 is distinguished by a high percentage of poorly drained soils (85 percent type G) in watersheds in southwestern Maryland near the Potomac River. Poorly drained soils occur in these watersheds where streams have incised upland deposits and where forested wetlands are present. Land cover is predominantly forest (71 percent) and the surficial geology is predominantly upland (65 percent) and marine Coastal Plain deposits. Cluster 11 is distinguished by a high percentage of urban area (77 percent) that occurs in watersheds around Baltimore, Md., and Washington, D.C.



## Discussion

Cluster analysis has provided an objective basis for preliminary classification of watershed units in the State of Maryland. Many of the clusters identified are similar to what might be predicted based on knowledge of the geographic characteristics of the State. It would be expected that the Carbonate Valley and the Appalachian Plateau areas of Maryland would be separate units as compared with the rest of the State. One advantage of the statistical analysis described here is that other factors besides physiography and geology can be objectively included in identification of HRUs for Maryland. The inclusion of land cover is a significant factor for hydrochemical response, and in this analysis, land cover was an important factor in distinguishing the urbanized watershed units that occur along the Fall Line, which separates the Piedmont and Coastal Plain Physiographic Provinces. Another advantage of this type of analysis is that more information can be included for identifying more spatially detailed HRUs than have been identified previously on the basis of physiography and geology alone. Inclusion of more detail on Coastal Plain geologic deposits provided a basis for identifying multiple HRUs for the Coastal Plain, which is commonly considered to be one unit.

It is not yet clear if the HRUs identified here will be consistent with the actual hydrochemical variation in Maryland. This analysis is intended to provide an initial classification of watershed units that is based on factors that can affect hydrologic and water-quality response of streams to precipitation and contaminant sources. It is possible that some of the HRUs identified for this study will behave similarly so that there would not be a need for separation. It is also possible that factors not accounted for in this analysis will cause variations within a single HRU that may complicate monitoring and modeling. These potential problems can be identified only with the collection of additional data. The

purpose of this analysis is to provide an initial framework for monitoring and modeling that can be modified as additional data are collected.

In addition, digital geographic data sets are continuously being improved to make them more current, more detailed and more relevant to hydrologic response. These refinements need to be incorporated in this analysis to keep any HRU definition relevant to current conditions and consistent with current understanding of earth and environmental sciences. For these reasons, the HRU definition as described here should not be considered a final product, but should be treated as a continuously evolving and improving classification of watersheds in Maryland.

## Summary

Cluster analysis was applied in order to classify watershed units into a representative and relatively homogeneous set of "Hydrochemical Response Units" that will be used for water-quality monitoring and modeling to aid in the development of water-quality management plans for the State of Maryland. To identify hydrochemical response units, a clustering algorithm was applied to basin characteristics data sets for 1,136 watershed units defined for the entire State. Basin characteristics information includes land cover, soil type, and geology data, all of which are expected to be significant in determining hydrologic response, and which are available digitally for the entire State. Results of the analysis indicated that a set of 11 clusters was optimal for distinguishing the major land cover, soils, and geologic characteristics in the State. These 11 clusters provide an initial classification of Maryland watersheds that are consistent with expected hydrologic and hydrochemical variation. This initial classification can be continually updated as digital geographic information is improved and the relation between watershed characteristics and hydrochemical response is further defined.

## References Cited

- Chatfield, C. and Collins, A.J., 1980**, Introduction to Multivariate Analysis: New York, New York, Chapman and Hall, 246 p.
- Lipscomb, S.W., 1998**, Hydrologic classification and estimation of basin and hydrologic characteristics of subbasins in central Idaho: U.S. Geological Survey Professional Paper 1604, 49 p.
- Maryland Department of Natural Resources, 1997**, Maryland Twelve Digit Watershed File (SHED1997): Annapolis, Maryland, Maryland Department of Natural Resources.
- Maryland Geological Survey, 1968**, Geologic Map of Maryland: Scale 1:250,000: Baltimore, Maryland, Maryland Geological Survey, 1 sheet.
- Maryland Office of Planning, 1973**, Natural Soils Groups of Maryland: Baltimore, Maryland, Publication Number 199, Maryland Office of Planning, 153 p.
- \_\_\_\_\_, **1996**, Maryland 1994 Land Use/Land Cover Data Set: Baltimore, Maryland, Maryland Office of Planning.
- SAS Institute, Inc., 1985**, SAS User's Guide: Statistics, Version 5 Edition: Cary, North Carolina, SAS Institute Inc., 956 p.

---

### For further information contact:

**District Chief**  
**U.S. Geological Survey**  
**8987 Yellow Brick Road**  
**Baltimore, Maryland 21237**

or visit the Maryland-Delaware-D.C.  
District Homepage on the World Wide  
Web at: <http://md.usgs.gov>

WRIR 00-4232

