Comparison of Automated Watershed Delineations: Effects on Land Cover Areas, Percentages, and Relationships to Nutrient Discharge

Matthew E. Baker, Donald E. Weller, and Thomas E. Jordan

Abstract

We compared manual delineations with those derived from ten automated delineations of 420 watersheds in four physiographic provinces of the Chesapeake Basin. Automated methods included commercial DEM*-based routines and different parameterizations of four enhanced methods: stream burning, normalized excavation, surface reconditioning, and normalized reconditioning. Un-enhanced methods resulted in individual watershed boundaries with some gross discrepancies in watershed size relative to manual delineations (error rate of 0.22 > 25 percent difference compared to manual) and significantly different watershed size distributions (Mann-Whitney U p = 0.012). Integrating mapped streams through enhanced methods substantially improved correspondence with manual watersheds (error rates of only 0.08–0.02 > 25 percent difference). Analysis of cropland area among methods showed a significant difference between manual estimates and un-enhanced estimates (p = 0.049) that was corrected using enhanced algorithms. Subsequent analysis of percent cropland revealed that measurements of land cover proportions were not always affected by delineation errors. However, differences were large enough to influence regressions with stream nitrate-N at the 90 percent confidence level within one physiographic province. Enhanced delineations produced statistical relationships between percent cropland and nitrate-N concentrations consistent with manual delineations. The results provide support for enhanced automated watershed delineation within the Chesapeake Basin and suggest that normalized excavation can be an effective augmentation of existing stream burning and reconditioning procedures.*

Introduction

As watershed analyses have become more pervasive in both landscape and aquatic ecology, watershed perspectives have become a prevailing paradigm in environmental management (Allan and Johnson, 1997, USEPA, 2001). This is particularly true in investigations of anthropogenic influences on aquatic condition (Hunsaker and Levine, 1995; O'Neill *et al.*, 1997). Watershed land-cover proportions and other landscape metrics have proven to be powerful predictors in statistical models of nutrient discharge (Omernik *et al.*, 1981; Johnson *et al.*, 1997; Jordan *et al.*, 1997a, 1997b; Herlihy *et al.*, 1998; Liu *et al.*, 2000; Castillo *et al.*, 2000; Jones *et al.*, 2001; Griffith, 2002; Griffith *et al.*, 2002; Reed and Carpenter, 2002; Weller *et al.*, 2003), and such landscape metrics are also relatively common geographic predictors of aquatic condition (Roth *et al.*, 1996; Richards *et al.*, 1997; Gergel *et al.*, 2002; Strayer *et al.*, 2003; Weigel, 2003). As aquatic assessment at specific sampling sites has become a primary tool for directing resource policy and management, identifying and characterizing upstream contributing areas for sampling sites across large regions is an important analytical step. Land-cover characterizations are usually based on the assumption that watershed boundaries are accurate, yet there are several methods for watershed delineation and differences among these methods are rarely, if ever, considered in landscape analyses.

In addition to manual interpretation of topographic maps (arguably the most accurate method) there are several readily available methods for automated watershed delineation using digital data. In the United States, automated delineation is facilitated by the availability of digital elevation models (DEMs) rasterized from 7.5-minute topographic quadrangles (**www.usgs.gov**). Advances in the analysis of flow direction and flow networks from DEMs have led to several automated methods for watershed and stream delineation (e.g., O'Callaghan and Mark, 1984; Jenson and Domingue, 1988; Tarboton, 1997). The methods generally follow three-steps: (a) modification of the DEM surface to correct for spurious sinks and indeterminate flow paths, (b) assignment and construction of a flow direction grid that allocates potential flux between adjacent raster cells, and (c) accumulation of cell contributions across a landscape. A flow direction surface can identify the contributing cells upslope of a user-defined outlet, whereas a flow accumulation surface may be employed to ensure that outlet locations match flow pathways. Because of the large samples typically utilized in regional aquatic assessment, automated watershed delineation represents an attractive alternative

Matthew E. Baker is with the Department of Aquatic, Watershed, and Earth Resources, Utah State University, 5210 Old Main Hill, Logan, UT 84322 (matt.baker@usu.edu).

Donald E. Weller, and Thomas E. Jordan are with the Smithsonian Environmental Research Center, 647 Contees Wharf Road, P. O. Box 28, Edgewater, MD 21037.

Photogrammetric Engineering & Remote Sensing
Vol. 72, No. 2, February 2006, pp. 159–168.

0099-1112/06/7202–0159/$3.00/0

that is potentially more objective, repeatable, cost-effective, and consistent with other digital data sets than manual delineation.

Despite its potential advantages, automated generation of upslope watershed boundaries in applied contexts involves several practical challenges and sources of error. First, stream sampling points, as watershed outlets, must be accurately located in proper relationship to streams and a flow direction grid. These sampling points can either be digitized manually from maps or captured by a global positioning system (GPS). Either method can produce errors that prevent sampling points from aligning with maps of streams and/or agreeing with local topography. Further errors arise because hydrographic and topographic maps rarely align perfectly, especially in moderate to low-relief landscapes. Such misalignment is more than an aesthetic issue because it can lead to very different interpretations of within-watershed flow pathways and resulting drainage boundaries.

Several enhanced algorithms attempt to resolve such discrepancies by combining topographic and hydrographic data. "Stream burning" uses a rasterized version of a digital vector hydrography map (also available at **www.usgs.gov**) to lower the relative elevations of stream pixels by a uniform depth. This excavates or 'burns' new channels into the DEM in an attempt to force alignment between topographically-derived flow pathways and independently-mapped hydrography. However, very deep channel excavations (e.g., 1,000 m) can cause undesirable distortion of watershed boundaries, and any excavation may result in unrealistic flow accumulation artifacts when neighboring channels, or *parallel streams* (one original and one excavated), compete for upslope flow (Hellweger, 1997; Saunders, 2000). An alternative to stream burning is "normalized excavation" which uses the topographic minimum from a specified local area to set a customized excavation depth for each stream pixel. This method helps resolve boundary distortions, but does not solve the parallel stream problem.

A third automated method, combining channel excavation with "surface reconditioning" using the AGREE algorithm, seeks to resolve the problem of undesirable competing flow paths within watersheds (Hellweger, 1997). AGREE involves the initial step of a uniform-depth excavation of channels from a vector stream map. The excavation is augmented by *reconditioning* within a specified distance of mapped streams, which creates a monotonic descent from the surrounding landscape to the excavated channel. Initial evaluations of AGREE seem promising (e.g., Saunders, 2000), but it remains unclear what effect such topographic modifications might have across a range of physiographic contexts, how they influence the results of automated delineation efforts, or what effect such modifications have on resulting land-cover analyses.

This paper seeks to evaluate different methods for automated watershed delineation by comparison of manual boundaries delineated from interpretation of 7.5-minute topographic maps with results from (a) un-enhanced delineation using standard DEM techniques, (b) delineation enhanced by uniform stream burning, (c) delineation enhanced by excavation normalized to surrounding topography, and (d) delineation enhanced by surface reconditioning using several different parameterizations of AGREE. A secondary objective addresses the sensitivity of delineation results to the relatively arbitrary parameters embedded in the automated methods. Our approach assumes that manual delineations are accurate and seeks to evaluate various automated methods relative to the manual standard. To better understand the potential importance of our findings, we also explore how the choice of delineation method affects the outcome of two common watershed analyses in aquatic resource management: estimates of cropland area and statistical models predicting in-stream nutrient concentration from cropland proportion. These analyses are used to place our results in a practical context rather than to provide a definitive investigation of factors controlling water quality.

Methods

Study Watersheds

We compared watershed delineations of 420 rural basins selected to study linkages between landscape factors and stream discharge (Liu *et al.*, 2000). The watersheds were located in 14 clusters distributed across four major physiographic provinces (Langland *et al.*, 1995) of the 166,000 km^2 Chesapeake Bay drainage (Figure 1): Coastal Plain (n = 120), Piedmont (n = 127), Appalachian Mountain (including Blue Ridge and Great Valley; n = 106), and Appalachian Plateau (n = 67). Watersheds were sampled in clusters to reduce variation due to sub-regional patterns of soil, geology, climate, and land management. All selected watersheds were characterized by a lack of sewage outfalls or other known point sources. Watershed sizes in our sample ranged from 5.5 to 48,010 ha. Liu *et al.* (2000) provide detailed descriptions of land-cover, physiography, and water chemistry.

Geographic Analyses

We analyzed publicly available geographic data sets for elevation, stream channels, and land-cover using the ArcInfo® 8.3 (ESRI, Inc) geographic information system (GIS). We used paper and digital versions of USGS 7.5-minute topographic quadrangles, 1:24 000 DEM data (approximately 30 m cells), and vector stream maps derived from the same 7.5 quadrangles. Land-cover information was calculated using the National Land Cover Dataset (NLCD), which was developed

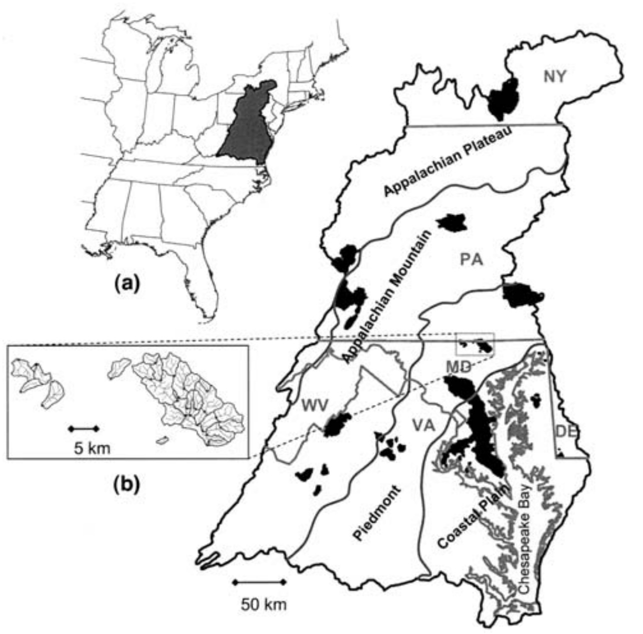

Figure 1. Location of study watershed clusters (black shading) in the physiographic provinces (after Langland *et al.*, 1995). Insets illustrate the location of the Chesapeake drainage in the Eastern U.S. (a) and the clustering of individual watersheds (b).

Table 1. Automated Watershed Delineation Methods and Parameters

Method	Reconditioning Width (# 30-m cells)	Normalization Distance (m)	Excavation (Smooth Drop) Depth (m)	Code
Un-enhanced	•	•	•	UNE
Stream Burning	•	•	2	SB2
Normalized Excavation	•	150	(original elev-local min) + 2	NX2
Reconditioning (AGREE)	2	•	2	R2_2
Reconditioning (AGREE)	5	•	2	R5_2
Reconditioning (AGREE)	10	•	2	R10_2
Stream Burning	•	•	10	SB10
Normalized Excavation	•	150	(original elev-local min) + 10	NX10
Reconditioning	5	•	10	R5_10
Normalized Reconditioning	5	150	(original elev-local min)	NR5

from 30 m Landsat Thematic Mapper images taken in 1992 (Vogelmann *et al.*, 1998a; 1998b, USEPA, 2000). Watershed boundaries were delineated manually (e.g., Dunne and Leopold, 1978) and using ten different parameterizations of four automated algorithms (Table 1). Manual delineations were interpreted from contour lines and streams on paper topographic maps, and then digitized into a GIS dataset (Liu *et al.*, 2000). In very low-relief landscapes of the outer Coastal Plain, county ditch maps and site visits were occasionally used to supplement topographic inferences about drainage divides (Jordan *et al.*, 1997a). Watershed outlets were located and digitized in the GIS from the paper maps.

Five methods of automated delineation were implemented using ArcInfo® macro (AML) programs (Table 1). In the un-enhanced delineation method (UNE), we did not modify the DEM and used standard ArcInfo® GRID commands (e.g., fill, flowdirection, flowaccumulation, snappour, watershed). This procedure generated a watershed boundary using information from surface topography alone.

For enhanced analyses, we used digital stream maps to modify the topographic surface represented by the DEM. Prior to analysis, we selected mapped hydrographic features corresponding to hydrologically connected stream channels, in-line lake shorelines, and reservoir shorelines. To implement the stream burning method (SB), stream maps were converted to single-cell strings in raster format using the cell size and extent of the DEM. As recommended by Saunders (1999), the hydrographic and topographic information came from the same 1:24 000 scale maps, and therefore had the same level of resolution. Stream pixel elevations were lowered by 2 m (SB2) and by 10 m (SB10) to compare the effects of using different uniform burn depths. In the normalized excavation method (NX), stream pixels were first lowered to the minimum elevation within 150 m, and then lowered an additional 2 m (NX2) or 10 m (NX10). All of the parameters were chosen to represent a range of reasonable modifications of the DEM based on published work (e.g., Saunders, 2000) and our own preliminary evaluation of errors from misalignment between stream maps and DEMs.

The remaining five delineations utilized surface reconditioning (R) and different parameterizations of AGREE (Table 1). The AGREE method involves three steps: (a) lowering stream pixels by an initial depth (smooth drop), (b) raising or lowering (reconditioning) DEM values within a specified distance from the mapped stream to ensure a smooth path of descent from the surrounding landscape to the stream channel, and (c) lowering stream pixels an additional depth (sharp drop) to create a trough-like channel in the already modified surface (Hellweger, 1997). We used smooth drop depths of 2 m and 10 m (e.g., R5_2 and R5_10) for comparison with stream burning and normalized excavation. In addition, we used reconditioning widths of 2, 5, and 10 pixels (e.g., R2_2, R5_2, R10_2; approximately 60 m, 150 m, and 300 m on either side of the stream, respectively) to explore the effect of different modification extents on delineation performance. Our rationale was that reconditioning widths less than 2 pixels would be methodologically indistinguishable from the simpler stream burning, whereas 5 pixels was similar to the area used by normalized excavation, and 10 pixels was illustrative of a larger modification. Thus, our design allowed comparison across methods at 2 m and 10 m excavation depths, as well as among different reconditioning widths. In addition, we explored the impact of adding "normalization" to AGREE by setting the smooth drop depth to the local elevation minimum within 150 m of each stream pixel in a fifth automated delineation method (NR5). A final sharp-drop depth of 1 m was employed in all parameterizations of AGREE.

In all five automated methods (un-enhanced DEM (a), stream burning (b), normalized excavation (c), surface reconditioning (d), and normalized reconditioning (e); Table 1), we followed a well-established, standard procedure of sink filling, assessing flow direction, and computing flow accumulation (e.g., Jenson and Domingue, 1988) following modification of the DEM. Watershed outlet locations were then adjusted to align with the greatest flow accumulation value (derived from each respective DEM) that occurred within the original, manually-digitized watershed boundary and within 1,000 m of the original sampling location. This adjustment ensured that each automated delineation procedure utilized the outlet location most likely to deliver a watershed similar to the manual boundary. Flow direction surfaces (also derived from each DEM) were then employed to delineate watershed boundaries upstream of the modified outlet locations.

We used the NLCD and delineation results to compare land-cover estimates using the various delineation methods. We focused specifically on row crop agriculture because percent cropland is the landscape metric most strongly correlated with nitrate-N concentrations in tributaries of Chesapeake Bay (Jordan *et al.*, 1997a; 1997b; Liu *et al.*, 2000; King *et al.*, 2005). Counts of cropland pixels were multiplied by the squared pixel size to estimate cropland area for each watershed. Watershed cropland area was divided by watershed area and multiplied by 100 to calculate percent coverage.

Water Samples

Water samples were collected from each watershed outlet quarterly for four seasons or longer, filtered, refrigerated, and later analyzed by a Dionex (Sunnyvale, California) Ion Chomatograph for nitrate-N (see Liu *et al.*, 2000 for details). To compare the differences among sites and minimize differences due to seasonal climate factors, we used mean nitrate-N concentration averaged across all seasons in our analyses.

Quantitative Analyses

We evaluated the relative accuracy of automated delineation methods by calculating the percent difference in watershed area between each automated result and the manually delineated watershed. We expected many small differences between manual and automated delineations to result from the interpretation of contour maps versus DEMs, but large discrepancies were interpreted as errors, particularly if they occurred in one but not all of the automated methods. By representing discrepancies as a percentage, we implicitly allowed for systematic delineation differences that might be expected to increase with watershed size. We also examined scatter plots of $\log_{10}$-transformed watershed area derived from each automated method against manually-derived areas to identify and track individual watersheds across delineation methods. There was no *a priori* criterion for identifying important errors in estimated watershed area, so we selected percent differences in excess of a 25 percent threshold to quantify the fraction of larger errors in each physiographic province for illustrative and comparative purposes. These error rates allowed us to assess which automated delineations were the most robust in approximating manual delineations both within a province and combined across all physiographic provinces. To better understand the magnitude of the observed differences, we compared the distribution of watershed areas from each automated delineation method to the distribution of watershed areas from manual delineations both across and within provinces by applying the nonparametric Mann-Whitney U two-sample test (Sokal and Rohlf, 1981).

Mann-Whitney U tests were also employed to compare the distribution of cropland area and percent cropland from manual delineations with estimates from each automated delineation method. Percent cropland was also compared to manual percentages in scatter plots across regions. In these comparisons, we focused on cases where large differences in cropland proportions occurred and examined whether the differences might be attributed to changes in watershed area. We were also interested in cases where large differences in watershed area did not alter cropland proportions. To evaluate the influence of delineation method on nutrient predictions, we regressed mean nitrate-N concentrations against percent cropland for watersheds within each physiographic province and compared the resulting regression slopes (the per-unit effect of cropland on nitrate-N concentration) using 90 percent confidence intervals. We did not analyze nutrient data across physiographic provinces because previous research has shown marked differences in nutrient concentrations among provinces at similar percent cropland (Jordan *et al.*, 1997a; 1997b; Liu *et al.*, 2000).

Results

Watershed Size

When the fractions of increases and decreases were summed, more than one-fifth of un-enhanced (UNE) watersheds exhibited areal discrepancies in excess of the 25 percent threshold when compared to manual areas across all physiographic provinces (Figure 2). The majority of discrepancies were errors of omission where un-enhanced watersheds were smaller than their manually-delineated counterparts. Among the physiographic provinces, watersheds in the Coastal Plain exhibited the greatest number of large areal discrepancies whereas watersheds in the Appalachian Plateau exhibited the fewest discrepancies.

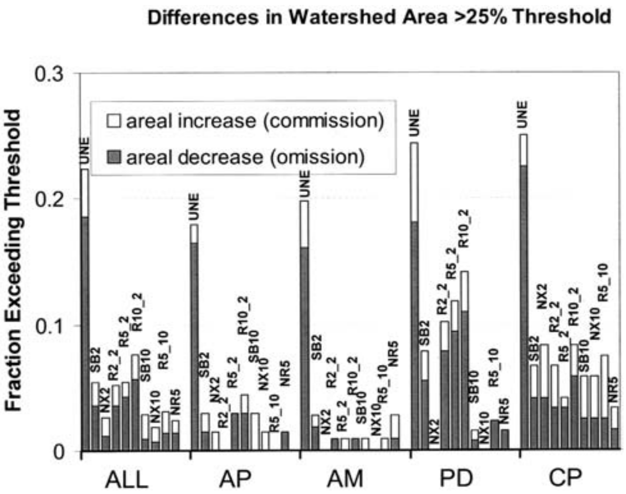

Figure 2. Comparison of delineation method error rates greater than a 25 percent discrepancy threshold as a fraction of all watersheds and those within Appalachian Plateau (AP), Appalachian Mountain (AM), Piedmont (PD), and Coastal Plain (CP) provinces. Errors are differences in watershed area from manual estimates as determined by un-enhanced DEM-based delineation (UNE), stream burning (SB), normalized excavation (NX), surface reconditioning (R), and normalized reconditioning (NR) methods (Table 1) relative to manual delineations.

Enhancing the automated delineations through stream burning (SB), normalized excavation (NX), or reconditioning (R) sharply decreased the number of discrepancies across all watersheds and within each physiographic province (Figure 2). For all enhanced methods, channel excavation resulted in far fewer errors of omission than the un-enhanced method. Errors of commission, where automatically delineated watersheds were larger than manual watersheds, were also reduced relative to un-enhanced delineations, except for delineations in the low-relief landscape of the Coastal Plain and the 2 m stream burn in the Appalachian Plateau. For all provinces and most automated methods, the largest errors tended to be those of omission. Large errors of commission were less frequent.

For all but one enhanced method, channel excavations of 10 m reduced the frequency of large errors relative to 2 m excavations. The sole exception was R5_10 in the Coastal Plain, which exhibited a nearly two-fold increase in large errors (0.04 to 0.076) relative to R5_2. However, in the low-relief Coastal Plain, deeper excavations resulted in a relative increase in errors of commission.

Comparing across different parameterizations of AGREE, increasing the extent of topographic reconditioning (R2_2, R5_2, and R10_2) led to greater error rates across all provinces except the Appalachian Mountains and Coastal Plain (Figure 2). Error rates achieved using a reconditioning width of two or five pixels were similar to those obtained using stream burning and normalized excavation, though a five pixel width resulted in greater error rates in both the Appalachian Plateau and the Piedmont. In addition, using a normalized excavation depth (NR5) gave a lower error rate than a uniform 2 m depth (R2_2, R5_2, and R10_2) across all provinces except the Appalachian Mountains and R2_2 in the Appalachian Plateau. NR5 also gave fewer errors than R5_10 in the Piedmont and the Coastal Plain.

Log-log scatter-plots of automatically delineated watershed areas illustrated the degree of correspondence between automated and manual results as well as the reductions in discrepancies achieved with the enhanced automated methods (Figure 3). Highlighted points below the 1:1 relationship lines resulted when an automated method omitted a sub-watershed (errors of omission). Highlighted points above the 1:1 line occur when an automated method erroneously included a neighboring watershed to give a subsequently larger watershed area (errors of commission). Many, but not all, of the discrepancies observed using the un-enhanced automated method (black points in Figure 3a) were corrected when enhanced algorithms were employed (Figure 3b through 3i). However, several parameterizations generated new discrepancies in watersheds that previously had little error (grey outliers in Figure 3b, 3e, and 3f). In this comparison, the tightest relationships between automated results and manual delineations were achieved using fixed 10 m (Figure 3c and 3g) or normalized (Figure 3h and 3i) excavation depths.

Pairwise comparison of automated to manual results revealed a highly significant difference between the distribution of watershed area from the un-enhanced automated method and the distribution of manually delineated area throughout the Chesapeake Bay drainage (Table 2). This overall result was corroborated by differences in the distribution of watershed area within three out of four physiographic provinces. In the Appalachian Plateau, the difference between manual and un-enhanced delineations was marginally significant ($p = 0.071$). In contrast to un-enhanced delineations, the enhanced algorithms did not exhibit the same trend and showed no significant difference ($p > 0.1$) with respect to manual delineations in all pairwise comparisons.

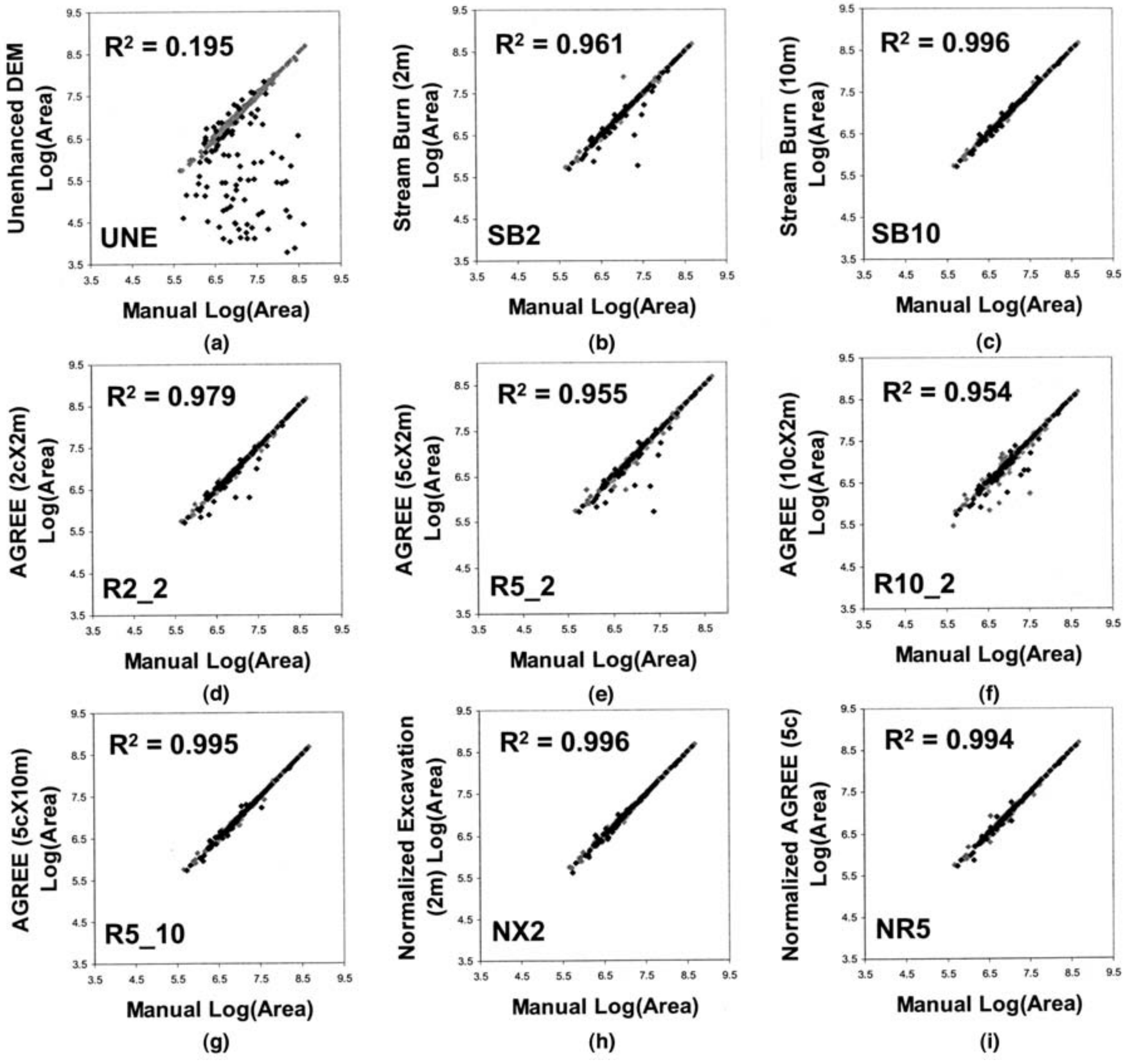

Figure 3. Scatter-plots of automated watershed area estimates versus manually delineated areas for un-enhanced DEM-based delineation (a), stream burning (b and c), surface reconditioning (d through g), normalized excavation (h), and normalized reconditioning (i). All axes are $\log_{10}$-transformed to emphasize smaller watersheds. Highlighted points (in black) indicate watersheds with >25 percent areal discrepancy relative to manual delineations using un-enhanced delineations (Figure 2).

Land Cover Areas

Differences in the distribution of watershed area were reflected in pairwise comparisons of watershed cropland area between manual and automated methods (Table 2). Across all provinces, un-enhanced delineations led to a significant difference in estimates of cropland area when compared to manual delineations. This pattern was consistent within the Appalachian Mountain and Piedmont provinces. In the Appalachian Plateau and Coastal Plain, differences in watershed area between manual and un-enhanced delineations did not result in significantly different estimates of cropland area. Watersheds delineated using enhanced algorithms showed no significant difference in cropland area with respect to manual delineations across all comparisons.

Table 2. Nonparametric Mann-whitney U Two-sample Tests for Identical Distributions in Watershed Area and Cropland Between Manual and Un-enhanced Automated Watershed Delineations

Extent of Analysis	Watershed Area	Cropland Area
Chesapeake Basin (420)	$p < 0.001$	$p = 0.049$
Appalachian Plateau (67)	$p = 0.071$	$p = 0.574$
Appalachian Mtn. (106)	$p = 0.012$	$p = 0.041$
Piedmont (127)	$p = 0.012$	$p = 0.049$
Coastal Plain (120)	$p = 0.032$	$p = 0.413$

Patterns in the estimation of cropland area were corroborated by plotting percent cropland-derived from automated methods against percent cropland derived from manual delineations (Figure 4). As with enhanced estimates of watershed area (Figure 3), percent cropland estimates from all automated methods showed a strong linear trend with manual results across all watersheds, but substantial discrepancies occasionally did occur within specific watersheds

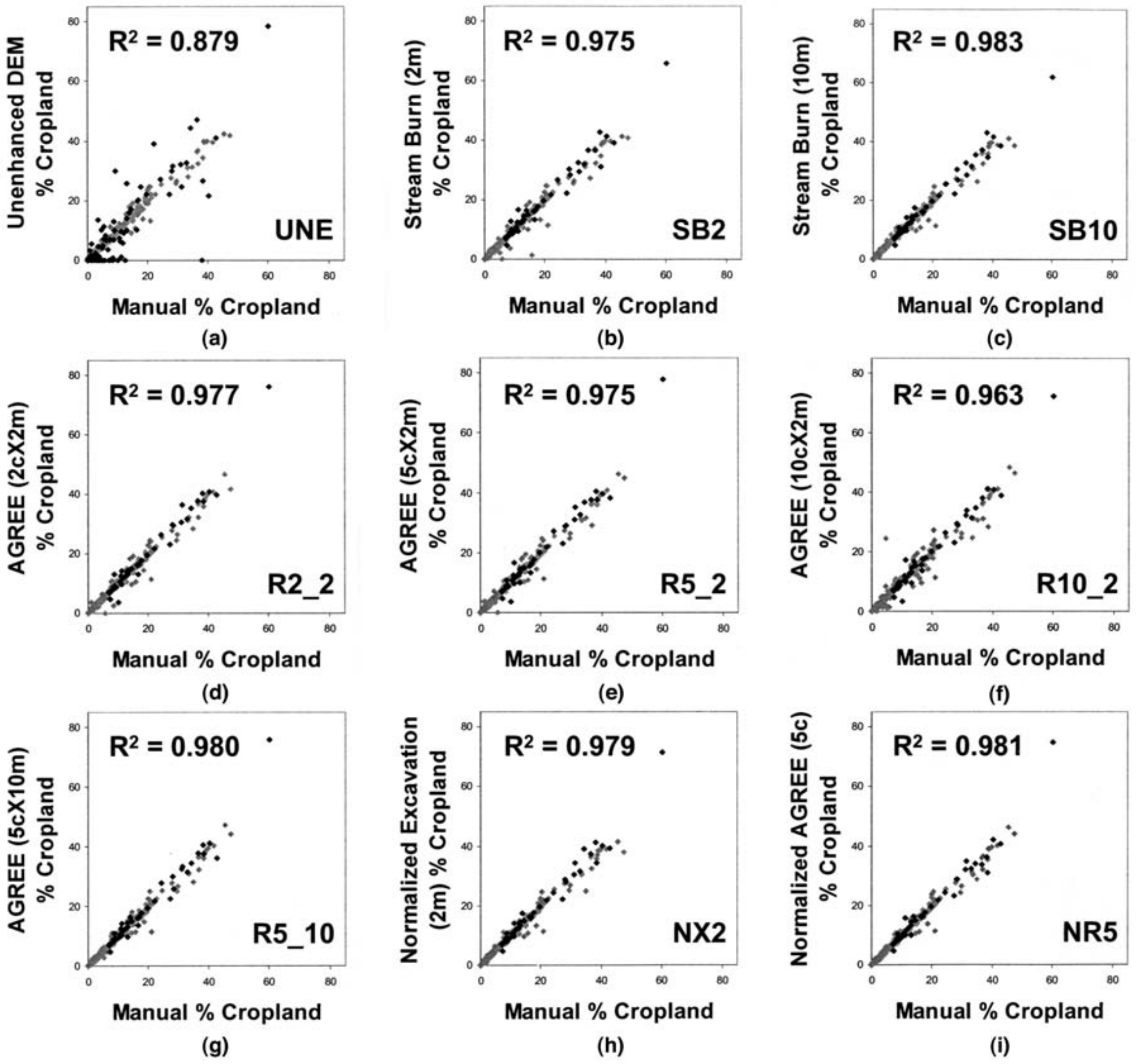

Figure 4. Scatter-plots of automated watershed area estimates versus manually delineated areas for un-enhanced DEM-based watershed delineation (a), stream burning (b and c), surface reconditioning (d through g), normalized excavation (h), and normalized reconditioning (i) methods relative to manual delineation. Highlighted points (in black) indicate watersheds with >25 percent areal discrepancy relative to manual delineations using un-enhanced delineations (Figure 2).

TABLE 3. COMPARISON OF NITRATE-N CONCENTRATION (UG/L) REGRESSED AGAINST PERCENT CROPLAND AS DETERMINED BY MANUAL, UN-ENHANCED, AND SELECTED ENHANCED AUTOMATED WATERSHED DELINEATION METHODS

Physiographic Region (N)	Delineation Method	Regression Results		
		R^2	slope	90%CI
Appalachian Plateau (67)	MAN	47.3	153	±33
	UNE	36.5	137	±37
	SB2	50.9	158	±32
	NX2	50.7	158	±32
	R5_2	51.5	157	±32
	NR5	51.1	157	±31
Appalachian Mountain (106)	MAN	75.2	193	±18
	UNE	75.0	187	±17
	SB2	78.2	203	±17
	NX2	76.7	202	±18
	R5_2	77.3	203	±18
	NR5	77.3	204	±18
Piedmont (127)	MAN	68.1	325	±33
	UNE	65.8	293	±31
	SB2	66.5	325	±33
	NX2	66.5	332	±35
	R5_2	65.4	325	±35
	NR5	65.7	333	±35
Coastal Plain (120)	MAN	62.0	84	±10
	UNE	57.1	79	±10
	SB2	61.7	85	±10
	NX2	61.7	84	±10
	R5_2	63.0	81	±10
	NR5	64.1	83	±9

(Figures 4a, 4b, and 4f). Many of the errors in percent cropland resulting from un-enhanced watershed delineation coincided with large areal discrepancies (highlighted points in Figure 4a) and were subsequently reduced using enhanced techniques.

Nutrient Predictions

Despite the moderating effect on delineation errors of using percent cropland rather than cropland area, watershed delineation method affected the results of regressing nitrate-N concentrations against percent cropland (Table 3). In most provinces, using percent cropland estimated from automated watershed delineations did not have a strong impact on model fit (R^2). However, un-enhanced delineations gave the lowest R^2 values in three of four provinces and, in the Appalachian Plateau, un-enhanced delineations resulted in a substantial reduction in fit compared to manual delineations (>10 percent drop in R^2). The slope of nitrate-N versus percent cropland relationship was generally similar to manual results when using enhanced delineation algorithms. However, the slope for un-enhanced delineations was the lowest in all four provinces and, in the Piedmont, was significantly different from the manual-delineation slope at the 90 percent confidence level.

Discussion

Watershed Differences

In comparing watershed areas, our primary purpose was to evaluate explicitly the performance of enhanced and un-enhanced delineations relative to manually delineated watersheds. In general, the un-enhanced automated delineation method originally described by Jenson and Domingue (1988) produced watersheds similar to those derived from manual interpretation of contour maps. Yet in certain topographic settings, the un-enhanced automated methods were inadequate and gave frequent large errors when compared to manually delineated boundaries (Figures 2 and 3a). Un-enhanced Coastal Plain delineations in particular had many errors due to the low-relief of drainage divides and the extent of ditching.

The distribution of large discrepancies (>25 percent; black points in Figure 3a) illustrates two distinct, but common errors in the un-enhanced delineations. Highlighted points below the 1:1 line are sites where the un-enhanced method identified and delineated a smaller sub-basin within (or, in a few rare cases, outside) the focal watershed. In automated delineations such as ours that utilize pre-existing geographic data, this could result from persistent and erroneous sinks in the DEM (despite a sink-filling step) leading to subdivision of flow accumulation within a watershed and incorrect repositioning of the outlet point within the manual boundary. Highlighted points above the 1:1 line were usually smaller errors, and often represented sites with outlets near a tributary confluence where the un-enhanced method erroneously included the adjacent drainage in the resulting delineation. Streams often join in relatively flat areas, so subjective manual interpretation of contours can produce boundaries that include DEM cells from neighboring basins. Because we adjusted outlet points within the manual boundaries to ensure that outlet locations occurred at locally maximal flow accumulations, these small erroneous areal inclusions in the manual boundaries near outlets could lead to large errors during any of the automated delineations. However, incorporating information from vector stream maps by using enhanced algorithms usually improved the accuracy of automated methods and often eliminated both additions of neighboring drainages as well as exclusions of upslope contributing areas (Figures 2 and 3).

Enhanced delineation methods substantially reduced the number of large areal discrepancies with the manual watersheds, but the improvement was not consistent across methods. Among the methods utilizing a 2 m excavation, reconditioning using AGREE led to greater errors as the extent of reconditioning increased (Figure 2). At this relatively shallow excavation depth, both stream burning and AGREE failed to correct several errors in the un-enhanced delineation and generated new delineation errors (Figures 3b, 3d, 3e, and 3f). In contrast, normalized excavation appeared relatively immune to such errors (Figures 2, 3h and 3i). Across all methods, increasing the excavation depth from 2 m to 10 m also appeared to remedy these problems (Figures 2, 3c, and 3g).

The parameterization of enhanced algorithms can introduce additional errors into watershed delineations. For example, although greater excavation depths appeared to reduce the frequency of large errors in some provinces, this was not necessarily the case in the Coastal Plain (Figure 2). Instead, digging a deeper channel in the DEM led to a decrease in errors of omission accompanied by an apparent increase in errors of commission (areal increases relative to manual delineations). These additional errors occurred when the channel excavation depth approached or exceeded the topographic relief of the focal watershed. While this is not a concern in some regions, excessively deep excavations are known to result in boundary distortions in other low-relief landscapes (Saunders, 2000).

We tested a range of various parameters in the AGREE algorithm and found a larger reconditioning extent lead to increased delineation errors. These errors occurred when the reconditioning extent crossed topographic ridgelines, when stream lines intersected with valley walls, or when stream profiles did not show a downstream descent. In other words, surface reconditioning can erroneously pull neighboring drainages into the delineation by altering ridges. When mapped streams lie on a local topographic high, surface reconditioning can create a downstream "dam" by raising the elevation of the valley bottom. In our evaluation, this happened when the reconditioning extent was sufficiently broad to encompass narrow watershed outlets and stream valleys or when the excavation depth (smooth drop) was not as deep as upstream lows. Increasing the reconditioning extent increased the potential for such systematic errors, but only when we employed shallow excavations. Smaller reconditioning extents certainly reduce such errors, but they also make the AGREE algorithm functionally indistinguishable from simple stream burning. Our results may not apply universally, but they do represent a range of physiographic conditions typical of the eastern U.S. that may help researchers and managers in other regions develop expectations about the likely performance of different delineation algorithms.

In a study of the watersheds of two Texas bayous, Saunders (2000) compared delineations using surface reconditioning and various excavation methods with topographic and hydrologic information from USGS 7.5-minute quadrangles. In his analysis, channels were excavated by 1,000 m (*fillburn*), normalized to an exponential drop within reaches (*expocurv*), or smoothed to ensure continuous descent between upstream and downstream ends of a reach and excavated by 2 m (*tribburn*). Extreme excavation values, such as the 1 km used in the *fillburn* method, were far more likely to result in watershed boundary distortions than the more moderate, but computationally intensive, *tribburn* method. Saunders (1999) found that his *tribburn* excavation method resulted in fewer boundary errors, but more stream-mapping errors than reconditioning and concluded that reconditioning was preferable. Like earlier analyses of DEM-based flow accumulations (e.g., Jenson and Domingue, 1988), Saunders did not distinguish between performance in the separate tasks of delineating drainage divides and mapping stream flow paths. In contrast, when topographic data are augmented by the integration of hydrographic information, we view the two goals of delineation and correcting internal flow lines as potentially distinct processing steps, each requiring its own evaluation.

Data Legacies

Many of the discrepancies among delineations efforts resulted from the different data sets used in our analyses and the way we located sampling sites. Our approach purposefully utilized publicly available data because our goal was to evaluate delineation methods under realistic constraints. In our case, initial watershed boundaries were manually digitized from 7.5-minute paper topographic maps following reconnaissance, site selection, and field sampling. For the many sites near stream confluences, slight errors in mapping outlets from paper maps could result in large differences in the automated delineation of upslope drainage areas. Both un-enhanced and, to a lesser degree, reconditioning methods seemed particularly sensitive to such errors. When watersheds are automatically delineated from outlet locations selected directly from digital stream maps, this type of error is less of a concern because outlet alignment with either the original DEM or a modified DEM is guaranteed (e.g., King *et al*., 2005). Automated analyses also require that both digital elevation and digital stream data be described at similar resolutions to avoid further error during the conversion of vector hydrography to raster maps of single-cell strings (Saunders, 2000). Although the USGS produced the DEMs and vector stream maps from the same 7.5-minute source maps, there are certainly discrepancies among the digitizing efforts and within the DEM, especially across map edges and regional jurisdictional boundaries. This will not surprise anyone familiar with the generation of digital data sets across such broad landscapes, but it can influence the accuracy of automated procedures, particularly when the procedures assume perfect relationships among data sets.

Our experience with data legacies does not necessarily represent a shortcoming of our approach. Many historic or GPS-derived data sets will have alignment problems or similar challenges. Therefore, our experience is likely to be quite representative though our strategy of utilizing manual delineations as an initial guide for automated delineation may be less typical. Furthermore, using multiple data sets combined through enhanced, objective, automated delineation techniques can help identify gross errors in manually-delineated boundaries. Thus, our methods provide an example approach likely to be useful in quality assurance and quality control efforts.

Implications for Land-cover and Water Quality Analyses

Significant differences between manual and un-enhanced automated watershed delineations observed across and within physiographic provinces led to a similar pattern of differences in estimates of cropland area. Differences within estimates of Piedmont and Appalachian Mountain watershed areas led to significantly different estimates of cropland area (Table 2). Such differences are of interest to researchers and managers concerned with predicting nutrient loading to downstream water bodies and non-point source pollution control (Carpenter *et al*., 1998). However, the fact that different delineation methods result in large errors of omission or commission did not guarantee that cropland estimates were affected. Coastal Plain watersheds exhibited significant differences in watershed area yet the rank order of cropland area estimates failed to reflect these boundary differences (Table 2). This pattern may result from strong sub-regional patterns of agricultural practice, where inner Coastal Plain watersheds have substantially less row crop agriculture than outer Coastal Plain watersheds (King *et al*., 2005). Thus, differences in watershed delineation caused by regional patterns of topographic relief and hydrology can interact with regional patterns of land-cover to create unpredictable, yet significant errors in the estimation of row crop agriculture. This finding suggests that any areal analysis of watershed land-cover based on un-enhanced delineations is vulnerable to similar errors. In this context, it is worth noting that in every comparison, enhanced algorithms appeared to correct both the watershed delineation errors as well as the estimation of cropland area.

Discrepancies among percent cropland estimates from manual and un-enhanced watershed delineations were more muted than differences in watershed or cropland area (Figure 4). Most of the percent cropland differences observed were the result of large areal discrepancies in watershed area (highlighted points in Figure 4). Areal differences among methods had to be both compositionally distinct and of sufficient size relative to watershed area to yield a large change in percent cropland. This would be more likely when a specific discrepancy added or removed cropland cells from small or moderately-sized watersheds, or when the discrepancy reduced watershed size in a predominantly agricultural landscape. We would expect such errors to be more common in physiographic contexts such as the Coastal Plain or Appalachian Mountains, where strong regional land-use patterns result in row crops agriculture in uplands or along valley bottoms, respectively. Relative to un-enhanced delineations, enhanced methods reduced substantially the number and magnitude of discrepancies in percent cropland. Nevertheless, estimates of land-cover percentages appear more robust to delineation error than do estimates of land-cover area. However, this also means that error attributable to delineation method is difficult to distinguish from variation potentially due to land-cover arrangement or within-watershed processes (e.g., King *et al*., 2005).

Enhancement of automated methods had a strong impact on the observed relationship between stream nitrate-N concentration and percent cropland relative to un-enhanced delineations. Although the observed differences among regression slopes were of marginal statistical significance, they could have a large influence on predictions based on statistical models. For example, using our regression models (Table 3), two hypothetical Appalachian Mountain watersheds with 25 percent and 75 percent cropland would have predicted nitrate-N concentrations 4.7 mg/l and 14 mg/l using the un-enhanced regression. However, if watersheds were delineated using a 2 m stream burn, the increase in predicted nitrate-N (5.1 mg/l and 15.2 mg/l) would mean predicted differences of 0.4 mg/l and 1.2 mg/l, respectively for downstream water bodies. Because such regression results are very sensitive to sample size, it is likely that other studies with much smaller or much larger numbers of samples will show a variety of responses to delineation errors. Nevertheless, our example illustrates that the choice of watershed delineation method might well have an important effect on the implied relationships between human activity in the landscape and any stream response correlated with land-cover proportions. All enhanced methods consistently resulted in regression slopes that agreed more closely with manual predictions than the un-enhanced method (Table 3), suggesting that some form of enhanced automated delineation procedures can improve the overall accuracy of aquatic assessment.

Directions for Applied Automated Delineation

Automated watershed delineation methods may be useful for evaluating manually delineated boundaries, but un-enhanced automated algorithms can not be reliably employed in aquatic assessment in all regions. Integrating mapped stream channels is one refinement that substantially improved the accuracy of resulting watershed boundaries in all four physiographic provinces. Thus, it appears that incorporating information about known drainage patterns improves the reliability of automated delineation regardless of topographic relief. Nevertheless, our results show strong differences among physiographic provinces in the performance of enhanced methods involving extensive DEM modification relative to manual delineations, indicating that the accuracy of enhanced automated delineations should continue to be evaluated on a regional basis.

Generally speaking, enhanced approaches involving some form of channel excavation appeared to be highly effective at reproducing the results of manual watershed delineation, as well as estimates of cropland area and relationships of land-cover to nutrient concentration within the study area. Utilizing the additional and sometimes extensive topographic modifications of surface reconditioning did not improve delineations appreciably compared to simpler enhancements, and sometimes resulted in additional errors. On the other hand, all enhancements resulted in a marked improvement over un-enhanced delineations. Our results illustrate that the consequences of using an un-enhanced delineation approach in watershed-scale studies will not necessarily be apparent from land-cover proportions or statistical model fits, but may yet alter the observed relationships between land-cover and stream response. Our case study involving land-cover analysis and predicting nutrient concentrations demonstrated that, relative to un-enhanced delineations, enhanced methods not only increase the accuracy of watershed boundaries, they also provide landscape characterizations and analyses based on landscape metrics that are more consistent with manually-delineated results.

When DEM surfaces are enhanced through integration of mapped hydrology, watershed delineation and internal flow pathways can be influenced by the enhancement approach and choice of parameters. Our results demonstrate that, when appropriately parameterized, any of the enhanced methods can be highly effective for watershed delineation within the Chesapeake Basin. The question then becomes how does one select the appropriate parameters so as to optimize delineation performance while minimizing the distortion and modification of the DEM surface? We evaluated using local normalization to contribute to delineation accuracy while minimizing unnecessary topographic modification. At the very least, this approach lowered stream pixels to downstream elevation values or the elevation of proximate valley bottoms, but it also reduced the magnitude of the elevation change in relatively flat landscapes when deep excavations had undesirable effects on watershed boundaries. By using local elevation minima in normalized excavation and normalized reconditioning approaches, we achieved improved correspondence with manual watershed boundaries without distorting the landscape with very large excavation values that could lead to new errors. Thus, our findings suggest that whether the goal is simple delineation or delineation combined with internally-corrected flow lines, incorporating local topographic information through normalization increases the accuracy of both stream burning and AGREE delineations while eliminating the otherwise arbitrary choice of a uniform excavation depth.

As broad-scale analyses in ecology and management become more common and more sophisticated, standardized delineation methods that integrate a variety of data sources may be desirable. For example, explicit analyses of land-use arrangement and flow lengths along surface flow paths (e.g., King *et al.*, 2005) do not function properly unless watershed boundaries align with the DEM. An approach to automated watershed analyses including (a) some form of normalized channel excavation for watershed delineation followed by (b) some form of surface reconditioning for improved internal flow lines may offer the best solution for future aquatic assessments.

Acknowledgments

We thank Ryan King, Cassandra Carmichael, Kathleen Boomer, and three anonymous reviewers for providing helpful comments on earlier drafts of this manuscript. David Correll played a vital role in starting and sustaining the water sampling described in this paper. Data collection was supported by grants from the National Science Foundation Ecosystem Studies Program (Grant Numbers DEB-93-19768 and BSR-9085219), the National Oceanic and Atmospheric Administration Coastal Oceans Program (NA66RG0129), the government of Charles County, Maryland, and the Smithsonian Institution Environmental Sciences Program. The current research was funded by a grant from the U.S. Environmental Protection Agency's Science to Achieve Results (STAR) Estuarine and Great Lakes (EaGLes) program to the Atlantic Slope Consortium (USEPA Agreement #R-82868401). Although the research described in this article has been funded by the United States Environmental Protection Agency, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

References

Allan, J.D., and L.B. Johnson, 1997. Watershed-scale analysis of aquatic ecosystems, *Freshwater Biology*, 37:107–111.

Carpenter, S., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley, and V.H. Smith, 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen, *Ecological Applications*, 8: 559–568.

Castillo, M.M., J.D. Allan, and S. Brunzell, 2000. Nutrient concentrations and discharges in a midwestern agricultural watershed, *Journal of Environmental Quality*, 29:1142–1151.

Dunne, T., and L.B. Leopold, 1978. *Water in Environmental Planning*, W.B. Freeman and Co., New York, 818 p.

Gergel, S.E., M.G. Turner, J.R. Miller, J.M. Melack, and E.H. Stanley, 2002. Landscape indicators of human impacts to riverine systems, *Aquatic Sciences*, 64:118–128.

Griffith, J.A., 2002. Geographic techniques and recent applications of remote sensing to landscape-water quality studies, *Water, Air, and Soil Pollution*, 138:181–197.

Griffith, J., E. Martinko, J. Whistler, and K. Price, 2002. An analysis of interrelationships among landscapes, NDVI, and stream condition in the U.S. Central Plains, *Ecological Applications*, 12(6):1702–1718.

Hellweger, F.L., 1997. AGREE – DEM surface reconditioning system, URL: *http://www.ce.utexas.edu/prof/maidment/gishydro/ferdi/research/agree/agree.html*, University of Texas, Austin, (last date accessed: 09 November 2005).

Herlihy, A.T., J.L. Stoddard, and C.B. Johnson, 1998. The relationship between stream chemistry and watershed land cover data in the mid-Atlantic region, US, *Water Air and Soil Pollution*, 105:377–386.

Hunsaker, C.T., and D.A. Levine, 1995. Hierarchical approaches to the study of water quality in rivers, *Bioscience*, 45:193–203.

Jenson, S.K., and J.O. Domingue, 1988. Extracting topographic structure from digital elevation data for geographic information system analysis, *Photogrammetric Engineering & Remote Sensing*, 54(11):1593–1600.

Johnson, L.B., C. Richards, G.E. Host, and J.W. Arthur, 1997. Landscape influences on water chemistry in Midwestern stream ecosystems, *Freshwater Biology*, 37:193–208.

Jones, K.B., A.C. Neale, M.S. Nash, R.D. Van Remortel, J.D. Wickham, K.H. Riiters, and R.V. O'Neill, 2001. Predicting nutrient discharges and sediment loadings to streams from landscape metrics: a multiple watershed study from the United States Mid-Atlantic Region, *Landscape Ecology*, 16:301–312.

Jordan, T.E., D.L. Correll, and D.E. Weller, 1997. Effects of agriculture on discharges of nutrients from Coastal Plain watersheds of Chesapeake Bay, *Journal of Environmental Quality*, 26:836–848.

King, R.S., M.E. Baker, D.F. Whigham, D.E. Weller, T.E. Jordan, P.F. Kazyak, and M.K. Hurd, 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams, *Ecological Applications*, 15(1):137–153.

Langland, M.J., P.L. Lietman, and S. Hoffman, 1995. *Synthesis of Nutrient and Sediment Data for Watersheds within the Chesapeake Drainage Basin*, U.S. Geological Survey Water Resources Investigations Report 95–4233, Lemoyne, Pennsylvania.

Liu, Z-J., D.E. Weller, D.L. Correll, and T.E. Jordan, 2000. Effects of land cover and geology on stream chemistry in watersheds of Chesapeake Bay, *Journal of the American Water Resources Association*, 36(6):1349–1365.

O'Callaghan, J.F., and D.M. Mark, 1984. The extraction of drainage networks from digital elevation data, *Computer Vision, Graphics, and Image Processing*, 28:323–344.

O'Neill, R.V., C.T. Hunsaker, K.B. Jones, K.H. Riiters, J.D. Wickham, P.M Schwartz, I.A. Goodman, B.L. Jackson, and W.S. Baillargeon, 1997. Monitoring environmental quality at the landscape scale, *Bioscience*, 47:513–519.

Omernik, J.M., A.R. Abernathy, and L.M. Male, 1981. Stream nutrient levels and proximity of agricultural and forest land to streams: Some relationships, *Journal of Soil and Water Conservation*, 36:227–231.

Reed, T., and S.R. Carpenter, 2002. Comparisons of P-yield, riparian buffer strips, and land cover in six agricultural watersheds, *Ecosystems*, 5:568–577.

Richards, C., R.J. Haro, L.B. Johnson, and G.E. Host, 1997. Watershed and reach scale properties as indicators of macroinvertebrate species traits, *Freshwater Biology*, 37:219–230.

Roth, N.E., J.D. Allan, and D.L. Erickson, 1996. Landscape influences on stream biotic integrity assessed at multiple spatial scales, *Landscape Ecology*, 11:141–156.

Saunders, W.K., 2000. Preparation of DEMs for use in environmental modeling analysis, *Hydrologic and Hydraulic Modeling Support with Geographic Information Systems* (D. Maidment and D. Djokic, editors), Environmental Systems Research Institute, Inc., Redlands, California.

Sokal, R.R., and F.J. Rohlf, 1981. *Biometry: The Principles and Practice of Statistics in Biological Research*, Second edition, W. H. Freeman and Co., New York, 859 p.

Strayer, D.L., R.E. Beighley, L.C. Thompson, S. Brooks, C. Nilsson, G. Pinay, and R.J. Naiman, 2003. Effects of land cover on stream ecosystems: Roles of empirical models and scaling issues, *Ecosystems*, 6:407–423.

Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models, *Water Resources Research*, 33(2):309–319.

United States Environmental Protection Agency (USEPA), 2000. Multi-Resolution Land Characteristics Consortium (MRLC) database, URL: *http://www.epa.gov/mrlcpage*, (last date accessed: 09 November 2005).

United States Environmental Protection Agency (USEPA), 2001. *Protecting and Restoring America's Watersheds: Status, Trends, and Initiatives in Watershed Management*, USEPA, Office of Water, EPA-840-R-00-001, 56 p.

Vogelmann, J.E., T. Sohl, and S.M. Howard, 1998a. Regional characterization of land cover using multiple sources of data, *Photogrammetric Engineering & Remote Sensing*, 64:45–67.

Vogelmann, J.E., T. Sohl, S.M. Howard, and D.M. Shaw, 1998b. Regional land cover characterization using Landsat Thematic Mapper data and ancillary data sources, *Environmental Monitoring and Assessment*, 51:415–428.

Weigel, B.M., 2003. Development of stream macroinvertebrate models that predict watershed and local stressors in Wisconsin, *Journal of the North American Benthological Society*, 22:123–142.

Weller, D.E., T.E. Jordan, D.L. Correll, and Z.-J. Liu, 2003. Effects of land use change on nutrient discharges from the Patuxent River watershed, *Estuaries*, 26:244–266.

(Received 14 September 2004; accepted 25 January 2005; revised 02 March 2005)