Automatic Watershed Location and Characterization with GIS for an Analysis of Reservoir Sedimentation Patterns

by

David M. Mixon

B.S., University of Cincinnati, Cincinnati, OH 1998

A thesis submitted to the Faculty of the Graduate School of the

University of Colorado in partial fulfillment of the requirement for the

degree of

Master of Science

Department of Geological Sciences

2002

This thesis entitled:

Automatic Watershed Location and Characterization with GIS for an Analysis of Reservoir Sedimentation Patterns

written by David M. Mixon

has been approved for the Department of Geological Sciences

James Syvitski

Robert F. Stallard

Date ____________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Mixon, David Michael (M.S., Geological Sciences)

Automatic Watershed Location and Characterization with GIS for an Analysis of Reservoir Sedimentation Patterns

Thesis directed by Professor James Syvitski

A method for reservoir geolocation and watershed characterization using a geographic information system (GIS) is presented along with a series of statistical analyses and several regression models for predicting reservoir sedimentation in the United States. Several freely available geographic databases describing topography, rainfall, soils, and land-cover were manipulated and queried by a set of scripts written in the Arc Macro Language (AML). The information collected is compared to the reported sedimentation rates in the Reservoir Sedimentation Survey Information System (RESIS) and used to construct a regression model to predict reservoir unit sedimentation rates (volumetric sediment accumulation normalized to watershed area) for the entire United States as well as individual models for six physiographic divisions.

The methods presented in this thesis allow an unprecedented level of information about a large number of watersheds to be collected when compared to the methods of previous studies. A comparison of three sets of similar, proximal watersheds was performed to evaluate the potential effects of basin misidentification. The results indicate that in areas of limited geographic complexity the variation of hydrologically important parameters describing topography, climate, and land-cover between proximal watersheds is small. As geographic complexity increases, so does

the variation of these parameters. With this knowledge, it is possible to estimate the probability, and consequences, of basin misidentification if the geographic complexity of the region is quantified.

The regression models illustrate the dominance of the capacity-watershed ratio in determining reservoir unit sedimentation rates as well as the important role played by watershed land-cover, especially forests and agriculture. Basin elevation is also shown to incorporate the effects of a number of covarying parameters to explain a significant portion of the variation of unit sedimentation rates seen in several regions. It is shown that the dominant parameters affecting unit sedimentation rates vary from region to region and the accuracy of continent-scale models is limited because of this. The role of reservoir management is not evaluated, but is suspected to contribute significantly to the variation in volumetric accumulation rates mainly through its influence on sediment density. The literature surveyed and regression results stress the need to quantify land-use and land-cover as well as the changing nature of both when attempting to predict unit sedimentation rates for reservoirs with small watersheds.

Acknowledgments

I wish to thank the entire "Delta Force" for their support and patience throughout this project. I owe a debt of gratitude to my advisors, Bob Stallard, James Syvitski, and Nel Caine for their expertise and inspirations. I also want to thank David Kinner, without whom this project may have never been completed. The U.S. Geological Survey's Mississippi Basin Carbon Project provided funding for this research. Finally, I want to thank my wife for her patience, support, and for making me have fun when I needed a good laugh.

CONTENTS

List of Figures

List of Tables

Chapter 1 – Previous Reservoir Sedimentation Research

1.1 The Emerging Importance of Reservoirs in the Global Sediment System

Meade et al. (1990), estimated that "90% of the sediment presently being eroded off the land surface of the coterminous United States is being stored somewhere in the rivers systems between the upland and the sea." Modern data show large reductions in suspended sediment flux across many world rivers, despite known increases in erosion. It is quite reasonable to hypothesize that these reductions are, at least in part, the result of modern reservoir construction (Vörösmarty, et al, *in press*). Takeuchi et al. (1998) estimated the total number of dams larger than 15m was approaching 39,000, globally. No global estimates have been made of the number of smaller ponds, but a few million is probably not an unreasonable estimate (Verstraeten and Poesen, 2000). The storage capacity of the world's "large" reservoirs was estimated to be about 4000 km^2, which is approximately 1/3 of the total water content of the atmosphere (Fels and Keller, 1973). This number certainly has increased during the past three decades. Vörösmarty, et al, (*in press*), in an analysis of the hydrological and sedimentological impact of the world's large reservoirs reported that between 1950 and 1968 global sediment trapping increased from 5% to 15% and increased to 30% by 1985. Their analysis also suggests that globally, registered reservoirs may be trapping as much as 25-30% of all sediment flux, amounting to 4 to 5 Gt yr^{-1}. The impact due to smaller reservoirs has yet to be quantified.

Reservoirs provide significant benefits to society, including flow stabilization, recreation, municipal water supply, power generation, and temporary sediment storage in areas prone to landslides or debris flows. Reservoir sedimentation impedes the effectiveness of a reservoir in these capacities. The cost of removing sediments from a reservoir to restore lost capacity was reported to be approximately \$0.18 per cubic meter in 1974 (McHenry, 1974). A more modern estimate of the costs of dredging varies depending on access and distance to a disposal site but ranges from \$26 m^{-3} to \$3,500 m^{-3} (Kattelmann, 1997). If an average cost is assumed to be approximately \$500 m^{-3}, and the reported sedimentation rates for each reservoir in this study are assumed to represent current rates, then the cost of simply maintaining all of the reservoirs ($n = 534$) in this study at their current capacity would be over \$2 billion annually. However, most reservoirs are not maintained in this way in part because of the economic costs of dredging. Regardless, the estimated annual economic cost of reservoir sedimentation in the US was estimated to be somewhere around S820 million (Crowder, 1987). While it has been suggested that this material could be used to augment the dam itself, none of the reviewed research has presented this idea in practice. Typically, these sediments are removed to locations distal from the reservoir, with cost increasing as transportation distances increase.

Additionally, these sediments often are not simple admixtures of mineral and organic compounds from the watersheds, but also may hold relatively large stores of chemicals washed in from the watershed during the life of the reservoir. Pesticides and herbicides as well as industrial pollution have all been found in reservoir sediments (Dertine and Mendeck, 1978). Certain compounds, such as polychlorinated biphenyls (PCBs) and heavy metals are primarily associated with river sediments rather than river water (Meade et al, 1995). Because they tend to be associated with fine sediments, which are preferentially trapped in reservoirs, they may actually be concentrated in reservoir sediments.

In addition to storing pollutants and sediment, reservoirs have also been shown to be significant sinks for carbon and, along with rice paddies and floodplains, may even explain part or all of the "missing carbon sink" (Stallard, 1998). Several studies have attempted to estimate the magnitude of this sink for reservoirs including Mulholland and Elwood, 1982; Ritchie, 1989; Dean and Gorham, 1998; Stallard, 1998; and Smith, et al, 2001. The first three estimate the global flux of sediment carbon to reservoir systems to be 0.2, 0.3, and 0.16 Gt yr^{-1}. Stallard (1998, Table 11) estimated that the flux could range from 0.18 to 0.28 Gt yr^{-1}. Vörösmarty, et al (*in press*), estimate a much smaller sink of only 0.08 to 0.10 Gt yr^{-1} (assuming a 2% carbon content, by mass). However, the latter estimate is based only on registered impoundments; potentially millions of smaller reservoirs and farm ponds were not considered and are theorized to have an impact of similar magnitude to the larger reservoirs. The U.S. has one of the most developed reservoir systems in the world and Smith, et al (2001), have recently presented a sediment budget that suggests that U.S. reservoirs alone may be a carbon sink on the order of ~0.024 Gt yr^{-1}. Regardless, it seems clear that reservoirs play an important, though still uncertain, role in the global flux of carbon.

1.2 The compilation of the RESIS database

During the middle part of the last century, several programs were developed that ultimately lead to a boom in reservoir construction. The Flood Control Act of

1944 authorized flood control projects on eleven large watersheds in the US. The Pilot Watershed Program, begun in 1953, authorized similar work on 65 small watersheds. Public Law 566, enacted in 1954, "authorized the planning and construction of works of improvement on small watersheds in cooperation with local sponsors." This last program led to 100 plans for watershed development per year during the 1950's and 1960's. Combined, these programs have led to the construction of several thousand floodwater retarding and multi-purpose structures. There were 7,993 structures included in work plans submitted by June 1965 and 3,487 had been constructed. These programs were immensely popular and led to a significant change in the hydrologic and sedimentologic cycle for the United States. (Roehl, 1966)

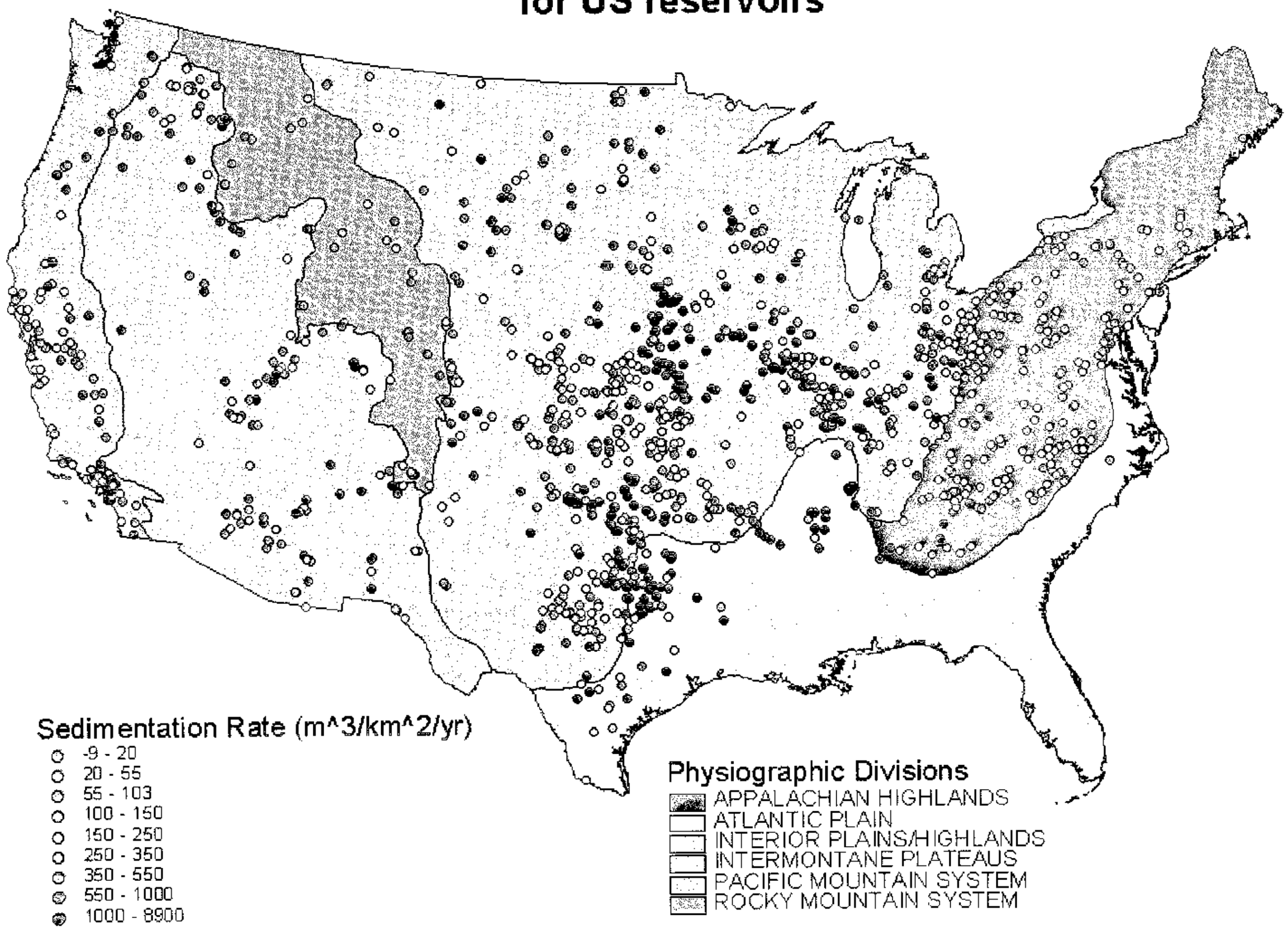
In response to the growing concern of reservoir siltation, the Subcommittee on Sedimentation of the Interagency Advisory Committee on Water Data began compiling sedimentation surveys for over 1,800 reservoirs across the United States. The result of this compilation is RESIS, the most comprehensive database of reservoir sedimentation in the world. The database contains nearly 6,000 surveys for 1,819 reservoirs across the coterminous United States (excluding Florida and Maine).

Periodically, the USDA summarizes the latest reservoir sedimentation data. The data in RESIS for the period ending 1965 was interpreted by Dendy *et al.* (1973). This summary included approximately 3,500 surveys. The data was summarized again for the period ending in 1975 by Renwick (1996). This summary included an additional 1,500 surveys. There is no published comprehensive analysis of the nearly 500 surveys entered into the database since 1975.

Renwick (1996) pointed out several notable biases likely present in RESIS. First, he noted that reservoirs with higher sedimentation rates are more likely to be surveyed than those with low rates. Second, many of the reservoirs in RESIS are located in agricultural sites, which tend to have smaller reservoirs than non-agricultural regions. Therefore, the average drainage area size sampled in RESIS is going to be disproportionately small compared to the true average drainage area of reservoirs in the US. Thus, the specific sediment yields (sediment yield per unit of drainage area) may be on the high end for the regions represented.

The database has been incorporated into a new database management package (Corel Paradox v9.0), given a new interface, converted to metric units, and the geographic coordinate information for most of the reservoirs has been updated. At the beginning of 1999, the database had latitude and longitude information for less than 800 of the 1,800 reservoirs in the database. Less than 300 had been "geolocated" on DEMs (enabling the delineation of the watershed). This made any large-scale use of RESIS with GIS systems impossible. Renwick (1996) used nearest town locations recorded in RESIS to approximate reservoir locations to within 20 to 50 km. With the ever-increasing number of sub-kilometer resolution databases of recent years, there is a growing need to find more accurate positions for these reservoirs. Recent work (outlined in section 2.2.4) has increased the number of reservoirs with reliable coordinates to 1,296. A map of those reservoirs with accurate geographic coordinates and symbols colored by their area-weighted (unit) sedimentation rates ($m^3\ km^{-2}\ yr^{-1}$) is shown in Figure 1.1. The topic of reservoir geolocation onto DEMs is discussed in greater detail in Chapter 2.

Area-weighted average sedimentation rates for US reservoirs



1.2.1 Survey Methods and Potential Error

Eakin (1939) outlines the steps for surveying a reservoir using *range* or *contour* mapping. These are the methods used for a majority of the surveys in the RESIS database. The principle advantage of the contour method is that it shows both the vertical and horizontal sediment distribution. However, more time is generally required for contour surveys and if the base map of the original valley is not of sufficient accuracy, significant errors can occur. The principle advantage of the range method is that it is quicker and it allows direct measurement of the sediment thickness. However, it is generally restricted to reservoirs where at least 50% of the sediment can be penetrated by a sampling spud (a large metal pole encircled by a stack of "cups" which is driven into the submerged sediment. When withdrawn, the depth of the sediment can be measured by examining the cups for the presence of sediment). A combination of both methods can be employed in many cases, where the lower part of the reservoir containing softer sediment penetrable by the spud is surveyed with ranges while the delta is surveyed by contouring. Previous maps can be checked by borings wherever possible.

In many of the surveys present in the RESIS database, areas were calculated to the nearest hundredth of an acre (if SCS guidelines were followed). Spuds typically measured sediment thickness to the nearest tenth of a foot. Rausch and Heinemann, (1984) report that ground and surface elevations are usually measured to an accuracy of ±3 cm and benchmarks are surveyed to within ±0.3 cm. For a reservoir averaging 2m of water and 1 m of sediment, this is an accuracy of ±1.5% for water volume and ±3% for sediment volume. Larger reservoirs will yield greater accuracy.

1.3 Previous Research Employing the RESIS Database

In a paper by Dendy et al. (1973), the data compiled in the RESIS database for 1,212 reservoirs through 1965 was summarized. For their compilation they divided the US into 79 sub-basins and summarized the results for each sub-basin. Lake Mead, one small reservoir in Utah that completely filled in one storm event and small debris basins and off-stream structures were excluded. The latest survey dates for their study ranged from 1918 to 1965. The authors point out that the quantity of sediment deposited in a reservoir is not equivalent to watershed sediment yield, as these surveys do not include sediments deposited above the spillway elevation and trapping efficiency is not considered, though watershed sediment yield is usually the predominant factor controlling unit sedimentation rates in reservoirs. While it is well established that watershed sediment yield tends to decrease with increasing basin area, their research reported a previously unnoticed "jump" in this trend between watersheds greater than 26 km^2 and those less than 26 mi^2. The reason for and validity of this "jump" has not yet been investigated.

It had become increasingly evident by the mid 1970's with the large RESIS database developing that the "relative importance of controlling factors varies from region to region and even within a region," and "...local parameters rather than climatic or geographic factors govern individual reservoir siltation rates" (Dendy et al., 1973). The large observed variation in unit sedimentation rates for the small reservoirs (<10mi^2) suggests wide local variations in characteristics such as vegetative cover, land-use, topography, etc. These factors are proposed to have a greater overall effect on unit sedimentation rates than do regional parameters such as climate.

Renwick (1996) took a more detailed look, examining reservoir sedimentation patterns taken from RESIS relative to regional topography, climate, and land-use. Specifically, the study re-examines the sediment yield data summarized by Dendy, et al. (1973) and attempts to determine broad environmental controls on specific sediment yield (SSY) and the generally negative SSY-drainage area relationship.

Renwick (1996) treats the phenomenon of reduction in specific sediment yield with basin area as a possible result of human modification to the landscape. For instance, he mentions, "the high rates of sediment accumulation in colluvial and alluvial deposits implied by sediment yield data probably cannot be sustained for long periods of time." This seems to suggest that the observed trend is not due to natural "buffering" by floodplains and naturally high erosion rates in the uplands, but rather to a pulse of sediment generated (or being generated) predominantly in the uplands, which is slowly working its way towards the lower basins.

Renwick (1996) uses the location of the nearest town as the reservoir location. He estimates that this gives an accuracy of about ±20 km in the eastern and central US and ±50 km in the western US. Land-use was taken from the 1982 National Resource Inventory MLRA-LRR (Major Land Resource Area - Land Resource Region) database. LRRs are roughly similar to physiographic regions. MLRAs and LRRs differ from strictly physiographic maps in that they incorporate land-use when defining the regions. Annual precipitation (P) and mean annual potential evapotranspiration (PE) were calculated using the Thornthwaite method. Topographic information was taken from 3-second DEMs. Local relief in the vicinity of the reservoir was determined by calculating the average value of local relief within a 5km radius of the town nearest the reservoir.

Of the 4 variables considered (drainage area, local relief, P-PE, and percent cropland in the MLRA) basin area had the strongest correlation with SSY. A weak, negative correlation between relief and SSY was attributed to the strong negative relationship between relief and percent cropland.

SSY was highest in the agricultural regions of the humid eastern and central states and in the Coast Ranges of California. Moderate SSY occurred in the western Great Plains, the semiarid western states, and in the Appalachians. Low SSY occurred in the forested areas of the northeast and northwest and in the northern Great Plains. In the western mountains, the wheat-growing regions of the Great Plains, and in the northeastern forests there was little or no downstream decrease in specific sediment yield. The percent agriculture in the MLRA has the strongest effect on both SSY and the SSY-Area relationship. The highest SSY and the greatest SSY-Area effect are present in the MLRA's with the highest percent agriculture.

Renwick (1996) groups the MLRAs into four major divisions:

1) **The forested mountains and uplands** have low sediment yields and little or no SSY-Area effect. The sediment ~~delivery~~ transport system may be more efficient in this region, or the uplands may not be providing proportionally more sediment.
2) **The semiarid uplands** have various SSY values and show a strong SSY-Area effect. Although these basins haven't been cultivated, they have

undergone significant increases in erosion since European occupation (according to Cooke & Reeves, 1976).

3) **The Great Plains** have generally low relief, although rangelands generally have higher relief than do agricultural lands. The erosion rates are lower than those found to the east in the Corn Belt and the region shows only a modest SSY-Area effect.
4) **The Corn Belt** and **the Piedmont** of the eastern slope of the Appalachians show the highest rates of sedimentation. The Corn Belt has 50-60% of its area in cropland while the Piedmont has only about 17% cropland (though the Piedmont was intensively cultivated during the 18th and 19th centuries). The acceleration of erosion due to man's influence is strongest in these areas. Relief is generally low and the SSY-Area effect is strong in these regions.

Renwick's analysis showed little correlation between local relief and specific sediment yield. The reason for this is not clear. It is possible that considering the entire dataset at once rather than grouping reservoirs by proximity allowed other parameters (such as basin area) to mask the influence of local relief. Alternatively, the method used to calculate local relief described above might not be adequate for characterizing true basin relief. While there is little doubt that basin relief does have an effect on sediment production, this effect is either not properly represented in his analysis or is overshadowed by other effects.

The conclusion states that spatial patterns of sediment yield are a consequence of natural features and human modification of the landscape. Human impact seems to be stronger (when analyzed this way) than the effect of natural features in controlling the spatial patterns of sediment yield.

The present research aims to improve upon the work already conducted by Renwick in several ways. Because Renwick (1996) was unable to geolocate any of his reservoirs, it was impossible to determine individual watershed properties. The watershed-level approach allows a more precise evaluation of the factors that affect sediment delivery to reservoirs. Additionally, land-use information is now available at a scale of 30-meters. The analysis by Renwick used the 1982 National Resource Inventory, which consisted of 181 major land resource areas averaging 43,000 km^2 in area. The present study also includes several variables describing watershed soil properties that were not considered by previous researchers.

1.4 Early Reservoir Research

Relating watershed parameters to reservoir sedimentation rates has been done many times in the past, but until recently watershed characterization has been a time-consuming task, involving detailed map and/or field analysis. The legacy of much of this work, along with new technologies in remote sensing and geographic information systems (GIS), has led to the growing availability of many digital data sources that can greatly simplify the task of watershed characterization. Outlined below are earlier findings and methods used in reservoir sedimentation research for comparison with the findings and methodologies used in this study.

Eakin and Brown (1939) outlined procedures for reservoir sedimentation surveys. They summarize some of the first organized surveys performed in the U.S. Their observations "emphasize the dependence of high rates of silting upon man-induced erosion and the general prevalence of these conditions over broad areas of the country." For each region (Southeast, Southern Great Plains, Southwest, and California), the predominant factors that influence erosion (leading to sedimentation) are reported below.

In the Southeast, Eakin and Brown (1939) note that rates of sedimentation were lower in mountainous areas with natural cover and higher in agricultural areas of the lower Piedmont. The Southern Great Plains showed higher rates within areas of soils covering sedimentary rocks particularly when augmented by agriculture and grazing land-use, though terracing, strip cropping, and similar control measures could effectively reduce sedimentation rates. They recommended the detention of sediment in broad tributary valleys and reduction in grazing as additional control measures.

In the Southwest, higher rates of sedimentation were commonly caused by sheet and gully erosion in areas subject to overgrazing. Eakin and Brown (1939) also note that due to the flashy nature of sediment delivery to reservoirs in this region, density currents will often enter the reservoir, displacing clean water over the spillway and filling the reservoir with sediment-laden water. This phenomenon points to the significance of engineering reservoirs to vent density currents (discussed later) as a measure for sedimentation control in this region.

In California, Eakin and Brown (1939) associate higher rates of sediment mainly with the occurrence of fires. Little more is said about the controls on reservoir sedimentation in this region, though they mention that sediment control practices have been effectively employed to reduce sedimentation rates.

In a broad sense, the Eakin and Brown (1939) report states that exorbitant rates of reservoir storage depletion are widespread and the problem of the protection of reservoirs from this problem "goes hand in hand with that of saving farm and range lands from impairment and destruction by uncontrolled erosion." They suggest that the most effective way to combat reservoir sedimentation is to eliminate erosion problems at their source through proven methods of erosion control.

1.5 Previous empirical models

Because of all of the factors affecting reservoir sedimentation, no physically based solutions have been created to predict silting rates and none are likely to be created soon. Thus, reservoir sedimentation has primarily been modeled using empirical relations (Singh and Durgunoğlu, 1989). A complete list of studies attempting to link reservoir sedimentation rates with watershed and reservoir characteristics is beyond the scope of this thesis, but in addition to the several studies outlined below, the following investigations can provide further insight: Woodburn (1955), Stall and Bertelli (1959), Ackerman and Corinth (1962), Farnham et al. (1966), and Singh and Durgunoğlu (1989).

Schumm (1956) investigated the relationship of several watershed parameters to the rates of sedimentation in thirty-five small stock ponds in Utah, New Mexico, and Arizona. The watershed relief ratio (the ratio of total basin relief to the length from the outlet to the most distant watershed point) was found to correlate highly with mean maximum slope, stream gradients, basin shape, and drainage density. Also, a strong positive correlation was found between the relief ratio and sedimentation rate. Schumm (1956) did not present a comprehensive model for predicting reservoir sedimentation rates and it was noted that regions with different climate and land-cover characteristics would need to be studied to better understand the processes controlling reservoir sedimentation. Previous papers have presented several regression models for predicting reservoir sedimentation (see Glymph (1954)). Gottschalk (1946) presented three equations using reservoir capacity, net drainage area, drainage density, reservoir age, and total precipitation to estimate volumetric accumulation rates. These equations were created from statistical analysis of surveys for 18 South Dakota reservoirs with similar watershed soils and with range-grass cover. Drainage areas ranged from 46 to 1635 acres (0.18 to 6.17 km^2). The third formula presented by Gottschalk (1946) (Equation 1.1) accounted for 89% of the data variability.

$$S = 0.0522C + 0.0027A + 0.2681T - 1.7974 \tag{1.1}$$

Where, S = Total sediment accumulation, in acre-feet
C = Capacity of the pond, in acre-feet
A = Net drainage area, in acres
T = Age, in years

This model is not likely to be very accurate outside of the area for which it was developed, as it does not account for differences in soil type, rainfall, and land-cover. However, it does illustrate that for studies of limited geographic range, the most variable parameters affecting sedimentation rates are often basin and reservoir geometry, and variables describing basin and reservoir size are able to explain a majority of the variation of sedimentation rates in these areas.

Anderson (1949) developed a similar equation (1.2), relating the natural log of the unit sedimentation rate (in acre-ft/mi^2) to peak annual discharge, area of the main channel of the watershed, and the cover density in the watershed. Equation 1.2 was developed from 23 cases and had an R^2 of 0.953.

$$\text{Log } e_D = 1.041 + 0.866 \log q + 0.370 \log A_{CH} - 1.236 \log C \qquad (1.2)$$

where,

e_D = Annual sediment accumulation, in acre-ft/sq mi
q = Maximum yearly peak discharge, cfs/sq mi
A_{CH} = Area of main channel of the watershed acre/sq mi
C = Forest cover density in the watershed, percent

This model was one of the first to incorporate a land-cover parameter. However, its widespread use is limited to those studies that are able to quantify both peak discharge and the area of the main channel.

Gottschalk and Brune (1950) presented equation (1.3) to predict total sediment accumulation in reservoirs if the watershed erosion rate is known or can be calculated.

$$\text{Log } S = 0.7664 \log 100W + 0.7867 \log T + 1.0545 \log E + 0.3701 \log C_r/W - 2.9127 \qquad (1.3)$$

where,

S = Total sediment accumulation in the reservoir, in tons
W = Net watershed area, in square miles
T = Age, in years
E = rate of gross erosion, in tons/sq mi/year
C_r/W = Capacity-watershed ratio of combined flood and conservation storage, in acre feet/sq mi of drainage area.

The equation was created to estimate the sedimentation rates in small reservoirs in the Missouri Basin Loess Hills of western Iowa. Surveys were made on 30 reservoirs ranging in age from 2.8 to 12.0 years. Drainage areas ranged from 0.098 to 107 square kilometers. The correlation coefficient for the above equation was 0.967. The use of this model along with an erosion prediction model such as WEPP or RUSLE to predict E has not been tested, but may be useful for similarly sized reservoirs in this region.

Around this time, several other authors proposed sediment yield relationships developed in a similar manner (Glymph, et al, 1951; Maner and Barnes, 1953; Kohler, 1954, cited in Glymph, 1954). These equations are not presented, though they are all similar in that they predict sediment yield based on measured or estimated erosion rates and watershed size for a limited geographic distribution of watersheds.

While these equations are likely to be unreliable if applied in regions different than those for which they were developed, they do illustrate that for any given continental region, a sedimentation equation based on a variety of watershed and reservoir characteristics can be produced if the watersheds all have similar land-cover. None of the reviewed empirical methods account for reservoir operational strategies or dam design. Glymph's (1954) review of these methods concluded that the shape of the drainage basin, channel density, rainfall distribution and topographic configuration all affect sediment yield and unless they are shown to be uniform should therefore be quantified for any robust treatment of sediment delivery to reservoirs. This demand for uniformity could well be extended to include watershed land-cover, land-use, and geology (i.e. soil type).

In a more detailed investigation of topographic effects, Flaxman (1966) used multiple regression and principle components multiple regression to examine the predictive capability of several variables for reservoir sedimentation rates in 30 reservoirs across the western US. Cultivated watersheds were excluded from this study. Sedimentation rates were obtained from the RESIS database. The parameters considered included climate, topographic factors, soil particle size distribution, plasticity index, soluble salts, density, and trap efficiency.

Flaxman (1966) concludes that runoff, or precipitation, is indicative of protective ground cover. Because of the positive influence of rainfall on ground cover, an increase in total precipitation leads to a decrease in sediment yield. So while land-cover was not explicitly considered in equation 1.5, rainfall acted in part as a proxy for land-cover.

The variable labeled X5 (1.5) below illustrates an attempt by the author to overcome a common problem in this type of investigation – that of representing the complex nature of watershed topography in one or two variables. While the topographic characterization of the watersheds in Flaxman (1966) was quite comprehensive, this characteristic makes the model dependent on a detailed DEM for each watershed. However, this study had the broadest geographic range of any presented thus far, and if a detailed watershed DEM can be obtained, it may be a useful model for Western reservoirs with uncultivated watersheds. Two equations are presented. The first is a simple multiple regression and the second is a "principle components multiple regression" with varimax rotation. The equations are shown below:

$$\text{Log } Y = 1.5945 + 0.6789\, X_2 + 0.0190\, X_3 - 0.00655\, X_7 + 0.1552\, X_{11} \qquad (1.4)$$

and

$$Y = -3.8182 - 0.5950 \log X_1 + 0.3517\, X_2 + 0.0258\, X_3 - 0.7165\, X_4 + 0.0066\, X_5 + 0.1405\, X_6 - 0.0066\, X_7 - 0.1219\, X_8 + 0.0423\, X_9 - 0.6127\, X_{10} + 0.0916\, X_{11} + 0.0235\, X_{12} + 0.0327\, X_{13} \qquad (1.5)$$

where,
Y = sediment yield in tons per square mile

$X1$ = average annual discharge in acre-ft/mi^2
$X2$ = precipitation intensity for the 5-year, 1-hour storm in inches
$X3$ = weighted average watershed slope
$X4$ = weighted average watershed slope / average annual runoff in acre-ft/ mi^2
$X5$ = a multi-dimensional topographic factor consisting of the product of weighted average slope, a value for drainage density, and a value for slope continuity
$X6 = X5$ / runoff
$X7$ = percent watershed soil particle size coarser than 1.0 mm (by weight)
$X8 = X7$ / runoff
$X9$ = plasticity index
$X10 = X9$ / runoff
$X11$ = percent of soluble salts in the watershed sample
$X12$ = density of reservoir deposits in lbs/ft^3
$X13$ = reservoir trap efficiency (%).

	Explained Variance	Stand. Error
Equation 1.4	$r^2 = 77$	0.4513 (log units)
Equation 1.5	$r^2 = 84$	0.5213

For the principal components multiple regression model, topographic factors and the bulk density of the deposits explained most of the variation (32% and 11%, respectively), while the proxy for vegetation density (topo./runoff) and trap efficiency explained 9% each. Soluble salts, precipitation intensity, percent finer than 1.0mm, log of runoff, and plasticity all contributed additional information to explain a further 17% of the variance.

A study investigating sediment deposition in 22 reservoirs in southwest Iowa and northern Missouri developed a regression model for predicting sediment retention in reservoirs (Farnham, et. al, 1966). The model predicted a combination of sediment delivery ratio and trap efficiency by using the ratio of the sediment deposited in the reservoir to the amount of calculated gross erosion (based on a modification of RUSLE and the Musgrave equation). They found that trap efficiency and delivery ratio were a function of five parameters: 1) ratio of mean direct tributary area to net drainage area, 2) relief-length ratio 3) ratio of non-incised channel length to total length of the drainage, 4) watershed shape factor (ratio of the watershed area to that of a circle with the same perimeter), and 5) slope of the highest order stream above the princip~~le~~al spillway. Farnham et al (1966) note that other parameters may be more important in different physiographic regions.

Paulet (1971) presented an analysis of fourteen reservoir-watershed combinations across southern Illinois (n = 8) and Indiana (n = 6). Grainsize distributions, organic matter percent, and bulk density were taken from the soils and sediments of these basins. Volumetric accumulation rates were also acquired for each reservoir. Many topographic and sedimentologic variables were measured. The most important parameters for determining reservoir sedimentation rates were logarithmic mean soil particle diameter (-), standard deviation of soil particle size (+), percent clay in soil (-), watershed area (+), stream length (-), order of the main stream (-), stream length ratio (ratio of average length of a given order stream to average length of the streams of the next lower order) (-), and capacity-watershed ratio (+). 73% of the variation of sedimentation rate was explained by five geomorphologic variables: area, stream length ratio, length of the main stream, order of the main stream, and C/W ratio (95% prob. significance).

A feature shared by each of these studies is a limited geographic range and a limited utility for large-scale reservoir sedimentation prediction. Modern earth science has seen a shift from a localized approach to a more geographically broad-based approach in an effort to deal with many of the large-scale problems such as global warming, the analysis of population pressure on natural systems, and water supply issues. Another problem with the aforementioned models is that most of them contain at least one variable, often more, which are difficult to tabulate for large numbers of watersheds (e.g. – grainsize distribution of incoming sediments, peak annual discharge of the main stream), especially for small watersheds that are typically ungauged.

The present study attempts to address this problem by utilizing new large-scale GIS databases for the US. The hypothesis is that easily obtainable databases could provide for a reasonably reliable model for reservoir sedimentation on large scales, when limited but meaningful information about each basin is known.

1.6 Land-use/Land-cover and Reservoir Sedimentation

Quantifying the effects of land-use and land-cover on erosion and hence reservoir sedimentation has proven to be difficult because local conditions seem to dominate. Much research has shown significant increases in sediment yield due to land-use change, particularly deforestation. These changes can have dramatically different effects on sediment yield depending on the type of change and the antecedent watershed conditions. For example land-use changes documented along the Yangtze in China show an increase in erosion by a factor of 3-4 while deforestation in Illinois resulted in a erosion rates 30-100 times previous measurements (Einsele and Hinderer, 1995). In some cases, the reservoir itself can become a driver of land-use change as the energy, water, and recreational opportunities that the reservoir provides result in increased local population and subsequent land-use change (Zhide and Yang, 1997).

Wolman (1989) presented data that shows a significant difference in sediment yield among basins with forested catchments (<50 metric tonnes km^{-2} yr^{-1}) and those that are predominantly agricultural (up to 350 tonnes km^{-2} yr^{-1}). Construction sites showed an even greater increase in sediment yields (1000's to 100,000's tonnes km^{-2} yr^{-1}) when compared with agriculture.

Land-use and land-cover effects seem to become less important as basin size increases and other parameters begin to dominate sediment production. However, the Basque River (or Reservoir) provides an example of the effects that land-use change and erosion conservation practices can have even in larger basins. In a 17-year period, this 1,666 mi^2 (4,315 km^2) basin showed a 38% reduction in sediment yield due to a comprehensive plan of erosion management (cf. 24-35% reduction for a 162 km^2 basin, 53% for a 36 km^2 basin, and 98% for a 5.85 km^2 basin) (Wolman, 1998).

Although Paulet (1971) did not consider land-cover or land-use change explicitly, he suggested that a major change in land-use would be expected to express itself in the geomorphic character of the watershed (e.g. bifurcation ratio, drainage density) with time. A sudden change in erosion rates due to land-use change would be detected in the reservoir sediments, but would not be reflected in the topography until the watershed had time to adjust. He concludes that the sedimentary records of the reservoirs in his study were long enough that the geomorphic properties of the watersheds had adjusted to any changes in land-use or land-cover. As a result, direct measurement of watershed land-use change was unnecessary. While this may hold true for studies with limited geographic range, this assumption is unlikely to be sufficient for continental-scale studies over diverse land-cover types, especially on a continent where land-cover change occurs rapidly and unpredictably.

Chapter 2 – Reservoir Geolocation and Watershed Characterization

2.1 The Databases

There are five large-scale datasets used in this study. They are described below, as they provide excellent resources for large-scale sedimentologic studies within the United States. The first and most integral dataset used is the Reservoir Sedimentation Survey Information System, or RESIS, described in detail in the previous chapter.

The second is the State Soil Geographic Database (STATSGO) for the conterminous United States (Wolock, 1997). Developed by the USDA, STATSGO summarizes local soil surveys on a national scale and provides many attributes, including organic matter content, K-factor (soil erodibility as defined in the Revised Universal Soil Loss Equation (RUSLE)), grain size information, available-water capacity, and permeability. A version of STATSGO that had been summarized on 1-km-grid resolution and is available at: http://water.usgs.gov/lookup/getspatial?muid is used to compile soil erodibility and organic matter content values. This version of STATSGO is convenient because it includes the information necessary for this study, but other versions of STATSGO are available through the USGS and USDA websites.

The third relevant database is the PRISM (**P**arameter-elevation **R**egressions on **I**ndependent **S**lopes **M**odel) rainfall database (Daly et al. 1994) that is maintained by the Oregon State Climate Center. The downloadable maps from the PRISM website at http://www.ocs.orst.edu/prism/_prism_products.html show mean annual rainfall for the United States. PRISM is a regression model that interpolates rainfall in mountainous regions where there are orographic controls over rainfall.

The fourth database is the USGS National Land-cover Dataset (NLCD). This dataset was generated through analysis of 1992 Landsat thematic mapper imagery and 1992 la consists of 21 classes of land-cover at a resolution of 30 m. Currently, these data are available online from the USGS at http://landcover.usgs.gov/natllandcover.html. Land-cover is tabulated from 1992 Landsat thematic mapper images and some supplemental data. Validation was performed using aerial photographs and limited field surveys. For this study, the 21 classes were simplified to 8 classes based on their estimated hydrologic and sedimentologic characteristics. These reclassifications are outlined in Appendix B. Although the NLCD distinguishes orchards and vineyards from other land-cover types, the near nonexistence of these types in the studied watersheds meant that there was no need to trouble with the decision of how to group these with other land-cover types and these areas were not included in the regression analyses.

The fifth dataset used is the National Inventory of Dams (NID). The NID contains general information, including reservoir capacity, dam height, construction date, and geographic coordinates for every dam in the United States that is over 6 feet (1.8 m) tall. This database was previously available online at the following address: http://crunch.tec.army.mil/nid/webpages/nid.cfm. However, national security concerns since September 11th, 2001 have led to its removal from public access indefinitely. This database is utilized in two ways. First, it is used during the preprocessing of the RESIS database to supplement the geographic coordinates of the reservoirs that were described in both databases. Second, the NID is used to determine the extent of reservoir "nesting" by recording the number of reservoirs present within the watershed area of each reservoir.

2.2 Preprocessing of the RESIS Database

2.2.1 Sedimentation Rates

The average sedimentation rate is calculated for each reservoir by summing the volume of sediments from each individual survey and dividing this value by the cumulative period of time between each of the surveys. This provides an average sedimentation rate in sediment volume per year ($kg^3\ yr^{-1}$) that is then normalized to total basin area ($kg^3\ km^{-2}\ yr^{-1}$) to calculate the unit sedimentation rate. Several of the reservoirs have one or more surveys that report negative values for the period (corresponding to an increase in storage). While most reservoir surveys show a decrease in storage volume with time, dredging, effective bottom-sediment resuspension and scouring by storms, or the raising of the spillway crest elevation may lead to an increase in storage. In this study, these increases were assumed to be the result of dredging or scouring. Because a desired product of this study is to develop a model that can evaluate the role of reservoirs as sediment and carbon storage devices, and dredging effectively releases stored sediment and carbon from the reservoirs, these negative values are included as part of the average unit sedimentation rate.

While most reservoirs are likely to demonstrate changing unit sedimentation rates from year to year, unit sedimentation rates in this study were assumed to be constant. Future analyses of the RESIS database may make use of temporally varying land-cover or rainfall information. For such studies, the time-varying unit sedimentation rates for each reservoir can and should be considered.

2.2.2 Density

The sediment quantities in the RESIS database are given as volumes. If one wishes to estimate basin sediment yield or nutrient or contaminant storage values, it is necessary to first convert these values to mass. Many surveys in RESIS present measured or estimated sediment dry weight values. However, there are many surveys that do not include dry weight values and other surveys only report estimates. Reservoir sediment density can be quite variable from reservoir to reservoir and even within a single reservoir. Butcher et al (1993), studying 28 small reservoirs in the Southern Pennine Region (UK) found variation in dry bulk density from 0.198 to 0.96 t m^{-3}. Because this thesis aims to aid future research of carbon storage in reservoirs, and carbon storage is usually computed as mass percent, a method for the prediction of sediment density as a function of known reservoir and watershed characteristics is investigated.

The primary factors affecting sediment bulk density (Brown, 1950, p. 785) are size frequency distribution of particles, time available for compaction, depth of burial, and the frequency and duration of subaerial exposure. Heinemann (1962) reported that percent clay, depth of sediment deposits, and sediment location relative to the dam accounted for 80% of the variation seen in density. Singh and Durgunoğlu

(1989) present a potentially useful equation for predicting reservoir sediment density, but the grain-size distribution of the incoming sediment must be known.

Flaxman (1966) also presented a "principle components multiple regression" equation that explained 62% of the variation of sediment density for 30 western reservoirs. This was dominated by the plasticity index (24%), followed by trap efficiency (13%), topo./runoff (9%), percent soluble salts (6%), precipitation intensity (5%), topo. factor (2%), percent finer than 1.0mm (2%), and log of runoff (1%). They theorize that a larger database, and one that would consider differences between submerged and subaerial sediments would help to explain a significant portion of the remaining uncertainty. Flaxman (1966) made no comment as to the possible reason for trap efficiency having such a strong effect on density, though this may be due to the influence of trap efficiency on sediment texture through the selective trapping of fine sediments.

While several factors ultimately affect the density of the sediments, the results of a study by Verstraeten and Poesen (2001a) also suggest that the factor most affecting sediment bulk density in reservoirs is subaerial exposure, which is quite difficult to estimate for large-scale studies. Attempts to estimate sediment mass for seven retention ponds in their study were reported to be in error by as much as 1 to 72%. This suggests that any study that attempts to convert sediment volumes to mass without a detailed investigation of the hydraulic history of the reservoir or direct measurement of sediment bulk density should be treated with suspicion.

Some authors have suggested that sediment deposited in the floodwater-retarding capacity (the portion of the reservoir volume that is reserved for absorbing incoming flood-waves and is typically unused for the much of the time) should be considered aerated sediment, subjected to alternate wetting and drying (Roehl, 1966). However, there is little practical way to determine which sediments were deposited in which pool for most of the reservoirs in RESIS. If the flood control proportion of the total storage can be estimated, it might be useful to consider this proportion to be aerated (thus higher density). Also complicating the matter, a large proportion of the coarse sediment arriving at the reservoir is deposited near the "entrance" to the reservoir and will therefore be more likely to undergo periodic subaerial exposure. However, the density of this coarse sediment is less affected by aeration than is the fine sediment (which is typically deposited nearer to the dam).

NRCS

The Soil Conservation Service has developed a simple system for estimating the proportion of sediments that would be aerated based on the topography of the watershed, the grain size of the incoming sediment, and method of transport (suspended v. bedload) for the incoming sediments (Roehl, 1966). However, without accurate grainsize information for the reservoirs in the present study, and without knowledge of the accuracy of this system, it is concluded that the possibility of introducing additional errors out-weighed the potential benefits provided by sediment density estimates calculated by this system.

Because no method suitable for the scale of investigation of the present study has yet been developed to predict reservoir sediment density, total masses and mass accumulation rates reported in Appendix D are based only on recorded densities from the RESIS database. Where no density value was recorded, the mean density for the reservoirs studied is used to provide an estimated sediment mass. Because this method does not provide unique densities for each reservoir, its use is likely to limit accurate prediction of mass accumulation through multiple regression analysis. Accordingly, the model results presented in the present study are only for volumetric accumulation rates.

2.2.3 Trap Efficiency

Reservoirs do not hold all of the sediment that is carried into their waters. While the overwhelming factor controlling reservoir sedimentation is the yield of sediment from the upstream basin, using the sedimentation rates of reservoirs to make interpretations about upstream erosion rates or basin sediment yield without correcting for the "trap efficiency" of the reservoirs can lead to serious errors. Trap efficiency is defined as the percentage of sediment entering the reservoir that is not discharged over or through the spillway. If one can accurately predict the trap efficiency of a reservoir, then the mass of sediment in the reservoir can be used with relative confidence to estimate the total basin yield.

The estimation of trap efficiency is complex and has been found to be a difficult parameter to quantify. Empirical methods for estimating trap efficiency have been proposed by many authors (see Brune, 1953; Churchill, 1948; Brown, 1943; Gill, 1979; Heinemann, 1984; and Verstraeten and Poesen, 2000). Typically, these models relate trap efficiency to the capacity-watershed ratio (C/W) or capacity-inflow ratio (C/I), where reservoirs with larger capacities relative to their watershed area or annual inflow will typically have higher trap efficiencies. Other models also incorporate grain-size characteristics of the incoming sediment into the prediction, increasing its accuracy, but reducing their widespread applicability. Each method has some advantages and disadvantages, but all share in common an inability to accurately predict long-term trap efficiencies for reservoirs in diverse geographic settings. Operational strategies of reservoir managers are also difficult to account for when dealing with large numbers of reservoirs, and the trap efficiency of reservoirs that are not normally ponded throughout the year prove especially difficult to predict.

Trap efficiencies are also not stable over time. For individual events, the trap efficiency of a given reservoir can vary greatly depending on the water level in the reservoir, the timing of the sediment arrival to the reservoir, the temperature and chemistry of the reservoir waters, and the management system used, if any, for the reservoir. Over longer timespans (e.g. greater than 10 years) trap efficiency is more stable, though it has been shown to decline with time as sediment accumulates in the reservoir, effectively reducing the C/W ratio. But even the long-term trap efficiency can be influenced by extreme events. Mclean, et al (1991) found that intense storm events can reduce long-term sediment retention by a factor of 3-4 due to bottom current resuspension. For this study, trap efficiency was assumed to remain constant, though more detailed future analyses of the same data could employ a time varying trap efficiency for each reservoir based on the recorded reductions in reservoir capacity with each survey.

no change in TE w/ time

These analyses should consider the pattern of reservoir sedimentation outlined by Lajczak (1996). The author details two phases of sedimentation that often occur in reservoirs. During the first phase, nearly all grain-sizes are trapped in the reservoir. As the reservoir shallows, the mean particle size trapped by the reservoir decreases, and assuming a constant incoming grain-size distribution, the total trap efficiency also decreases. The first phase of silting ends when the mean depth of the reservoir reaches a critical value (dependent on hydrological conditions – but for the Polish reservoirs studied, the depth was about 5 m in medium to large mountainous reservoirs, 3-4 m in medium to large lowland reservoirs, and about 2 m in small, shallow reservoirs along stream courses). At this stage, trap efficiency is lower and some years may actually show negative infilling rates. The second stage ends when a new river channel is formed along the completely filled up reservoir basin. The duration of both phases depends on the geological, morphological, and climatic conditions in the catchment areas. No quantitative description of these relationships has been developed, but could be useful for studies considering the effects of the passage of time and reservoir "aging".

When studying sediment yields over decadal timescales for large, normally ponded reservoirs, the empirical method proposed by Brune (1953) is the most widely used and probably produces satisfactory results. However, for small ponds no reliable empirical relationship has been developed and physically based models such as the STEP (Sediment Trap Efficiency for small Ponds) model (Verstraeten and Poesen, 2001b) or FLOODSIM (Bechteler and Nujic, 1998) require too much detailed information about reservoir geometry and sediment characteristics to be applied to large numbers of reservoirs.

As mentioned above, Brune's equation is more commonly used in the literature, and has shown to be more reliable for normally ponded sediments. However, because inflow values were recorded for only a small number of the reservoirs in the present study, Brown's (1943) equation is employed to estimate trap efficiency for the reservoirs. This equation (2.1) needs only the capacity watershed ratio (C/W) to estimate trap efficiency (TE).

$$\mathrm{TE} = 100\left(1 - \frac{1}{1 + \mathrm{K(C/W)}}\right) \tag{2.1}$$

Where C is the capacity of the reservoir, in acre-feet, and W is the area of the watershed in miles. The curve can be modified using the constant, K, which accounts to some extent for the variations due to different reservoir practices (such as sediment flushing) as well as differing inflows between reservoirs with similar C/W ratios. Brown found K to range between 0.046 and 1.0 with a median of 0.1.

For the present study, because reservoir practices were not known, it is assumed that the K for each reservoir would vary as a function of rainfall. To produce a similar distribution of K values as found by Brown (1943), the maximum, minimum, and median values for mean annual rainfall for all of the studied watersheds were plotted against the maximum, minimum, and median values of K reported by Brown. A power function was fit through the three points made by these pairs and the power function was used to predict K values as a function of rainfall. Using this method, for two reservoirs with similar C/W ratios the reservoir receiving a greater mean annual rainfall total is assigned a lower K value, and thus a lower trap efficiency. The distribution of estimated trap efficiencies is presented in Figure 2.1.

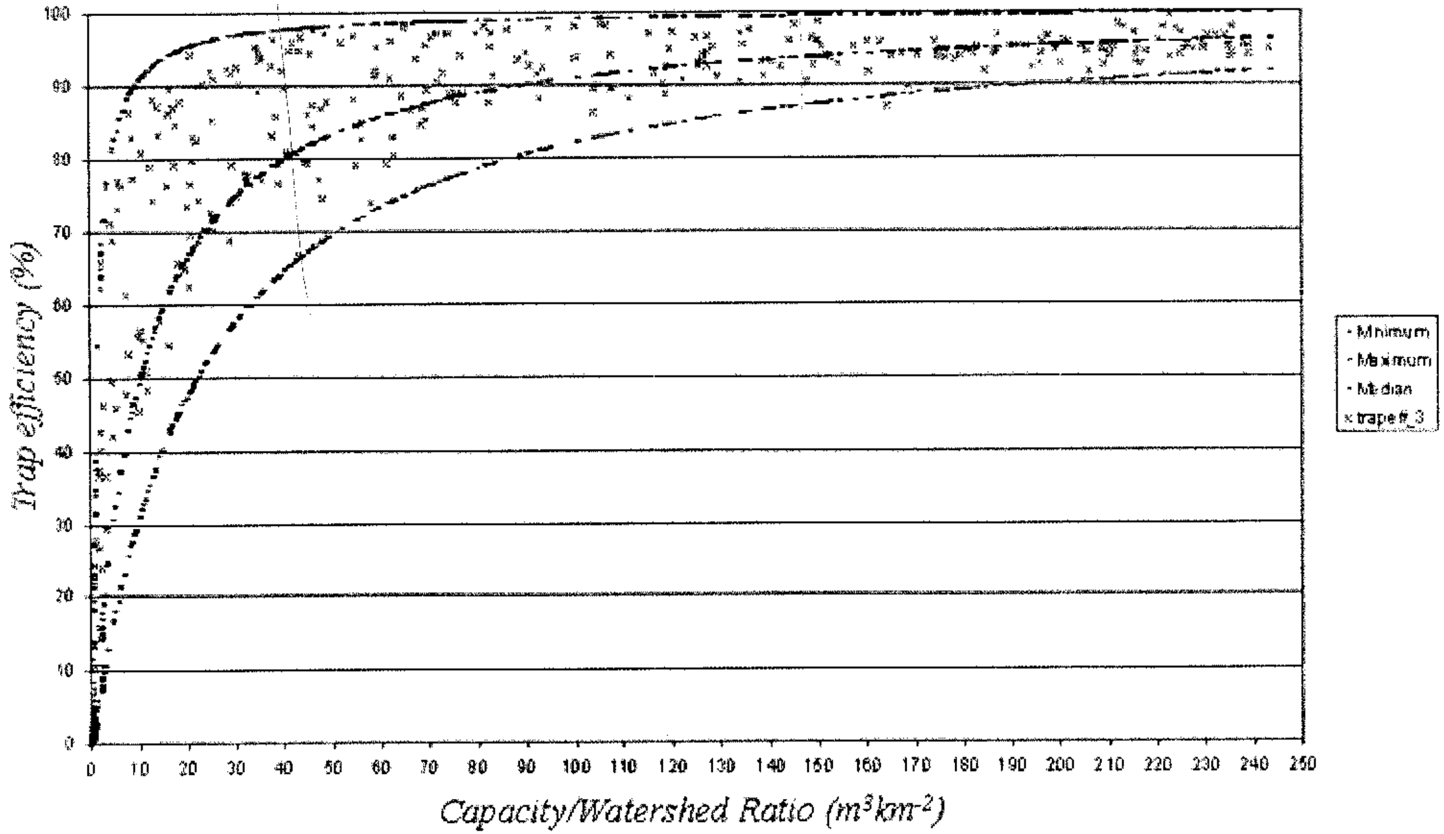


***Figure 2.1** – Estimated trap efficiencies of the 537 reservoirs for which there were original capacity values as well as the maximum, minimum, and median trap efficiencies possible using Brown's equation.*

For the present work, although trap efficiency for the reservoirs is computed, the trap efficiency-corrected sedimentation rates (i.e. basin yield) were subject to such a variety of possible errors from density and trap efficiency correction that multiple regressions were unable to adequately predict basin sediment yields. Nevertheless, the preceding discussion is presented as a cautionary note to those who wish to use reservoir sedimentation rates to make interpretations about erosion patterns. good pt.

2.2.4 Starting Coordinates

Fewer than 900 reservoirs in the RESIS database had geographic coordinates that could be used without alteration for input into the geolocation program. A small number (334) had previously been located on topographic maps and given new coordinates that would match their location to their watershed "outlets" on standard USGS DEMs. For the remainder, several techniques were applied to create "starting coordinates" that could be used in the geolocation program detailed below. The starting coordinates are required to create the "area DEM" representing the topography in the region near the reservoir. In this study the area DEMs were constructed from 9 7.5-minute DEMs arranged in a 3 by 3 matrix. Because the average width of a 22.5-minute DEM in the US is ~32 km, the starting coordinates must be accurate enough to ensure that an appropriate area DEM is generated to contain the full extent of a 250 km^2 watershed.

The first method involved matching reservoirs in RESIS to those in the NID based on reservoir name, as well as recorded city, county, and state in which the reservoir is located. It was assumed that the coordinates in the NID are more reliable than those in the RESIS database and visual inspection of coordinates for a few of these confirmed this. Accordingly, 735 reservoirs were given improved coordinates from the NID (note - in some cases, the NID coordinates superceded the coordinates already recorded in RESIS).

If latitude and longitude were not provided in either the NID or RESIS databases, the second method uses the recorded Public Land Survey (PLS) coordinates (if they existed) to determine approximate latitude and longitude. This is performed using a program called TRS2LL developed by Marty Wefald, that returns the latitude and longitude for the center of a PLS section for the much of the US. The program is available at http://www.geocities.com/jeremiahobrien/trs2ll.html. Because latitude and longitude are for the center of the section in which the reservoir was located, these coordinates may be in error by as much as ~0.7 mile. There were 50 reservoirs that were given new coordinates with this method.

The third method of "starting coordinate" creation was used when all of the above methods were not usable. This method used recorded "nearest post office" value in RESIS. Like the method used by Renwick (1996), the USGS "places" database is queried for the coordinates of the town center for the nearest post office. These coordinates are assigned as the reservoir's "starting coordinates." The previous methods are not always possible and the nearest post office is available for nearly every RESIS reservoir. This method represents the worst-case scenario for starting coordinate determination. The estimated positional error of this method is variable, depending upon which part of the country the reservoir is in; the error is higher in the western states because the relative density of post offices is lower than in the eastern

US. This method of starting point determination was required for 427 of the reservoirs in the RESIS database.

2.3 Geolocation

2.3.1 The Arc Macro Language Method

The original plan for implementing this GIS was to locate the reservoirs by hand and digitize the coordinates at the top center of the dam. It soon became apparent that this was intractable. Many of the reservoirs in the RESIS database do not show up on standard 7.5-minute topographic quadrangle maps, because 1) the most up to date map available has not been revised since the construction of the reservoir; 2) the reservoir is too small, or is dry for too much of the year to show up on the aerial photos used to construct the maps; or 3) the reservoir had been drained or completely filled with sediment at the time of the compilation of the topographic map available. Thus, an alternative method was needed for the geolocation of these reservoirs if they were to be characterized in a GIS.

Arc/Info was used to find the best coordinates given the limited starting information. The hypothesis is that if geolocation by hand is impractical or impossible, then collecting data from a "similar" watershed in close proximity to the dam will provide in most cases a reasonable description of the watershed above the dam. A "similar" watershed is defined as a watershed that lies at the same elevation and has the same total drainage area as the watershed described in the RESIS database. For the purposes of a broad sedimentary environment characterization on a continental scale, the present research aims to answer the question: does it matter whether data were collected from the watershed above the dam or whether data were collected from a nearby equivalent? The following sections describe this method and present the results of an examination of the variability of several GIS-derived parameters between nearby watersheds of similar size and at similar elevations.

The program was developed in the Arc Macro Language (AML). The RESIS database contains a general description of nearly every reservoir, including values for watershed area, maximum and minimum elevation in the watershed, and elevation of the spillway and dam. If watershed boundaries were derived for a large number of these reservoirs, a GIS can be used to collect a vast amount of data about the watershed for each reservoir using freely downloadable grid and polygonal coverages describing rainfall patterns, soil characteristics, land-use and land-cover, and of course, topography. A broad variety of geospatial data beyond the data used in the present study is available but was not considered in this investigation.

To delineate the watersheds, it was necessary to have a representation of the local topography. In this case, a Digital Elevation Model (DEM), which represents elevation as a continuous grid of elevation values, was chosen. These are the most widely available forms of topographic data for the US and are easily used for topographic characterization. The USGS has digitized most of the 7.5-minute topographic maps for the continental U.S. Although a large-scale seamless DEM for the continental US (The National Elevation Dataset (NED)) was made available during the course of this thesis, the funds for its purchase were not available. The necessary 7.5-minute, 30-meter resolution grids were acquired from the USGS WebGLIS site developed at the EROS Data Center and later from http://www.gisdatadepot.com/ when the responsibility of DEM distribution was transferred from the USGS to a private company. The DEMs were stored and most GIS calculations were performed on a SUN mainframe computer. ESRI's software packages Arc/Info 8.0.2 and ArcView 3.2 for UNIX were installed on a 400MHz SUN Enterprise 6500 running SunOS version 5.8.

The AML script was originally designed to run on only small watersheds (those with a watershed area of approximately 250 km^2 or less). Most natural watersheds of this size are likely to be contained on a 3 by 3 mosaic of 7.5-minute, 30-meter resolution DEMs (at least in the continental US). To construct these 22.5-minute "area DEMS", a polygon coverage showing the name and location of all USGS 7.5-minute topographic maps was acquired. This coverage contains the USGS Quad-ID code, which is based on a grid numbering scheme that can be reproduced with a fairly simple algorithm. This permits the creation of an AML script that returns the name of the 8 quads surrounding a given 7.5-minute quad. A modified version of this script was used to mosaic the DEMs around the best available coordinates for the dam (the compilation of the starting coordinates is not a trivial matter and is discussed in detail in section 2.2.4 above labeled "Starting Coordinates").

Once the DEMs are properly mosaiced, the next step is to "fill" them, eliminating small errors in the DEM called "sinks", which consist of a cell or small group of cells lower than all surrounding cells. Generally, these sinks do not represent the true topography (and thus hydraulic flow) of the area, though karst terrain and glacial till plains are examples of areas where true sinks occur (few reservoirs were located in such places, so no effort was made to insure the preservation of true sink holes or kettles).

Once the area DEM is prepared, a "spillpoint" is determined for each possible reservoir location. The information provided in RESIS for each reservoir made this possible. Knowing the approximate location of the reservoir, the minimum and maximum elevation in the watershed, the elevation of the dam spillway, and the watershed area allows one to relatively accurately locate the spillpoint on a topographic map or DEM by hand. Using GIS commands to determine the elevation and contributing area of each grid-cell allows the procedure to be automated.

One problem with identifying the true edge of the dam is that on some DEMs the reservoir appears as a topographic feature (i.e. a large flat area at the elevation of the spillway), while other DEMs instead show the original topography underlying the reservoir surface. Thus, there is no consistent automatic way to identify which point along the stream course is the true spillway. So, the program looks to see if the lowest elevation in the watershed is recorded (it is not for all reservoirs in RESIS) and if it is not it uses the spillway elevation. In some cases this means that the watershed is delineated slightly above or slightly below the true spillway. But, the mean watershed values for parameters like slope, soil properties, and land-cover that were collected for statistical analysis are not greatly affected by these deviations. Using this elevation, the DEM is queried for elevations falling within 10% of this value. This produces a binary grid with "rings" of potential reservoir sites around high and low features (as illustrated in Figure 2.2).

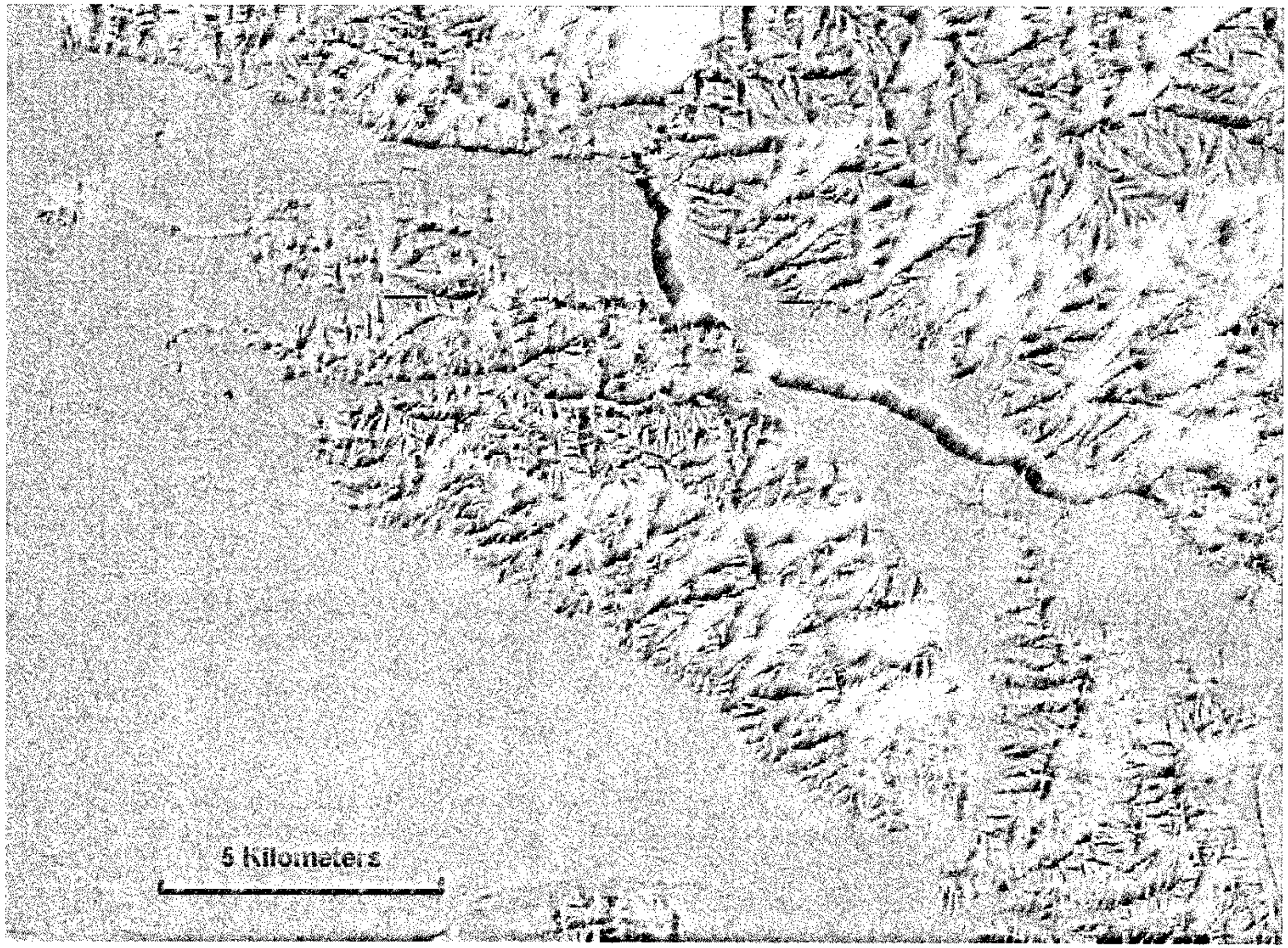


Figure 2.2 *– The first step in the geolocation process using the area near Pasadena, California as an example. Area shown in blue represents the selection of a 10% range of elevations around the minimum recorded watershed elevation.*

With the majority of the DEM area eliminated as potential dam locations, the next task is to further limit the search using total watershed area. Thus, the *Flowdirection* and *Flowaccumulation* commands were run on the area DEM around each reservoir to calculate the watershed area for every point on the DEM. Then, using the total watershed area recorded in RESIS, this grid was queried for accumulation areas within 10% of this value, producing "strings" along stream courses showing potential gridcells for the dam location (as illustrated in Figure 2.3).

Finally, these two binary grids were multiplied to produce a final binary grid. The points on this grid represent "candidate points" with an elevation and contributing area within 10% of the true values for the reservoir. Because of the 10% window of elevation and watershed values selected, many of the candidate "points" were actually composed of strings of gridcells. This is illustrated in Figure 2.4. The final "spillpoint", defining the best guess for the location of the dam, was chosen as the candidate point closest to original starting coordinates. In some cases, this meant that a point with an elevation and/or contributing area having an error of 10% was chosen despite the fact that a more suitable point (i.e. a point with a lower deviation from the recorded value) may have existed just a little further from the starting point. Modifications were made to a subsequent version to alleviate this problem. Although this modified version has been applied to a small number of reservoirs for test purposes, it was not applied for the collection of the data in this thesis and will not be presented here. Nevertheless, the tolerance level for both elevation and contributing area (currently 10%) can be changed in the current version of the algorithm, based in part on estimated terrain complexity, which is discussed more in a later section.

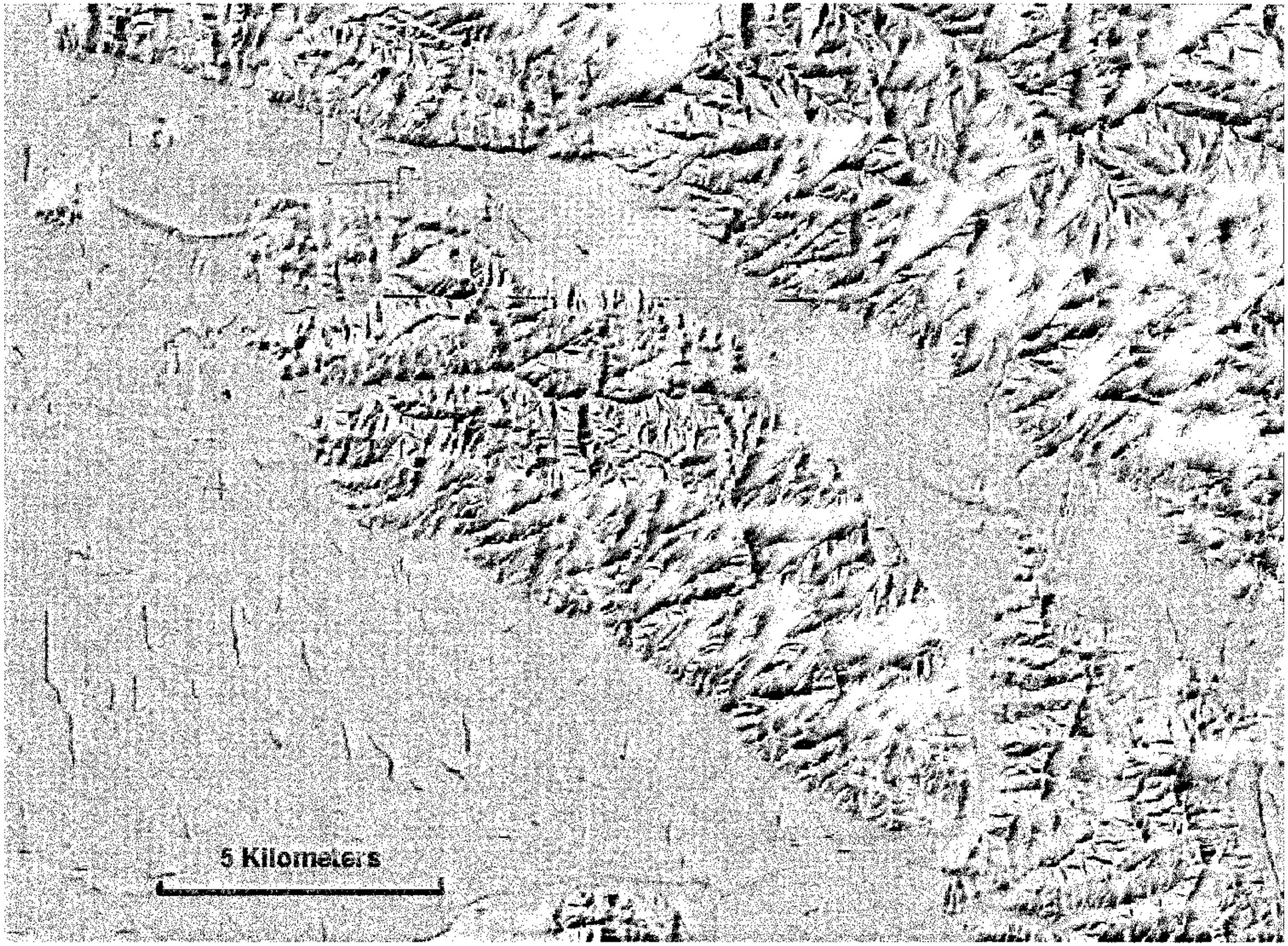


Figure 2.3 *– The second step in the geolocation process. Selection of a 10% range of total drainage areas around the recorded total drainage area..*

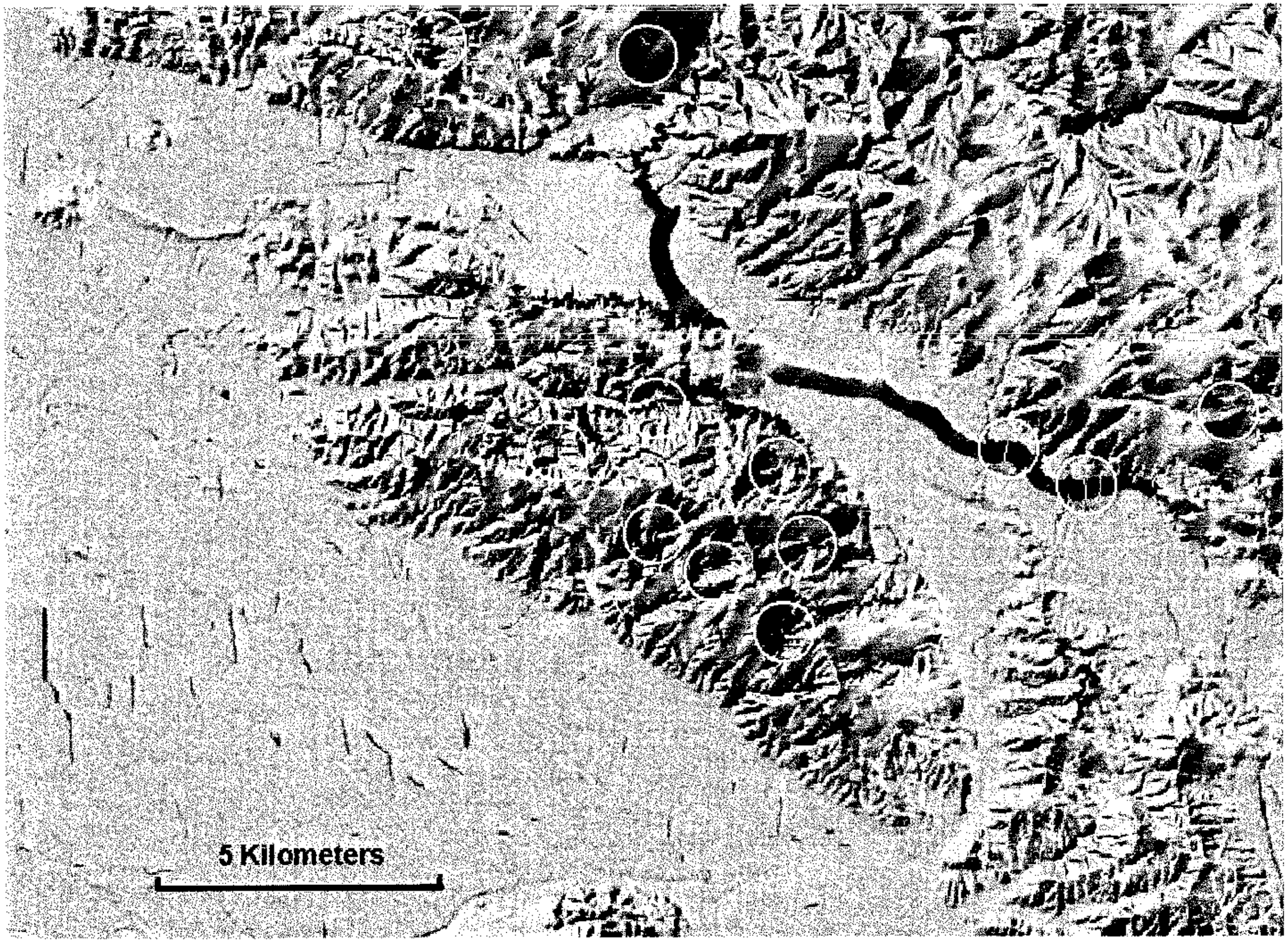


Figure 2.4 *– The third step in the geolocation process. "Candidate points" (circled in yellow) are defined as areas with the correct elevation and total drainage areas. These represent potential reservoir locations. The "candidate point" closest to the originally recorded geographic coordinates is chosen for basin delineation and characterization.*

Once the spill point is defined, the *Watershed* command is used to delineate the watershed with the DEM. This creates a binary grid that can be converted to a polygon coverage. This watershed polygon is used to clip the DEM of the local area so that parameters can be collected describing the watershed topography (see Appendix A for a description of each parameter collected, and Appendix D for the recorded values for each reservoir). For most watersheds, it was observed that the 30-meter resolution of the DEM was sufficient for characterization of the general topographic character of the watershed. However, for watersheds with areas less than ~0.5 km^2 (n = 52 for this study), many of the derived topographic variables may not be reliably represented at this resolution.

The watershed polygon was also used to clip several additional grids (described above) characterizing soil, rainfall, and land-cover. These values were added to a new database file keyed to the original datasheet numbers from the RESIS database (each reservoir is assigned a unique 4- or 5-digit datasheet number (DSNUM), where the first 1 or 2 digits identifies the sub-basin in which the reservoir lies. These sub-basins are shown in Appendix F, Figure F.1).

2.3.2 Problems Encountered

In some cases, the DEMs were provided in inconsistent vertical units (i.e. sometimes in meters, sometimes in feet). It was necessary then to convert the 9 grids to be mosaiced into one consistent set of units. As the RESIS database was compiled in English units, and the majority of the DEMs used feet for elevation units, all metric elevation grids were converted to English units. Watershed values such as relief were later converted back to meters.

DEMs do not patch together seamlessly. Figure 2.5 shows an example of a slope grid generated from a watershed DEM containing one type of seam error. In this case, the DEM to the south shows elevations along the valley bottom near the seam to range from 1161 to 1165 feet above mean sea level (msl). The DEM to the

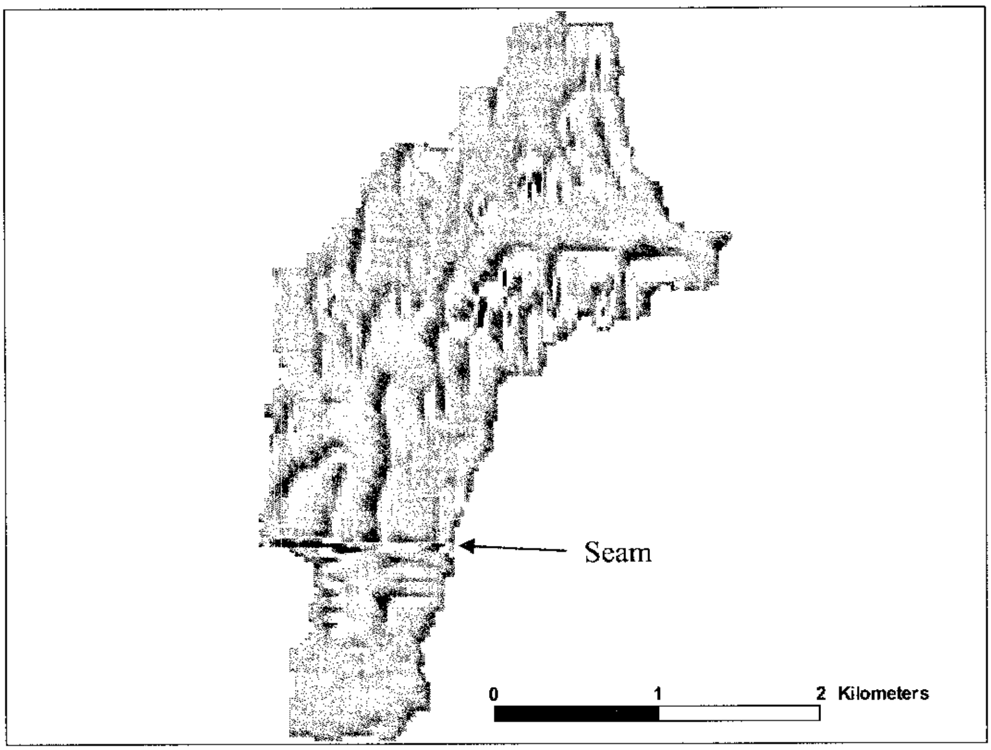


Figure 2.5 *– An example of a slope grid containing an extreme seam error (indicated by linear area of unusually steep slopes in southern part of watershed). Parameters such as hypsometric interval and mean slope are probably affected by this type of error. It was present in 9 out of 537 delineated watersheds.*

north shows elevations in the valley bottom near the seam to range from 1135 to 1141 feet above msl, creating an artificial "cliff" in the watershed. This type of seam error occurred with varying severity in 44 cases. However, in no instance was it found that the watershed boundary was significantly affected by the presence of this seam - Figure 2.6 (next page) illustrates this. The topographic map for this watershed is overlain by two boundaries. The geolocation program derived the red boundary while the black boundary was traced on screen over a scanned topographic map in ArcView. The two blue lines indicate the locations of two level 3 seams (seams were categorized based on relative severity from 1 (barely detectable) to 3 (significant elevation difference across the seam). There is no visible effect on the boundary of the watershed in the vicinity of these seams.

In many cases, another type of seam error occurred where adjacent DEMs did not meet exactly, leaving a gradually widening gap along the seam between them. This is fixed with an edge-matching algorithm that calculates a moving average using a 4x4 window for any blank cells between the adjacent DEMs. This leads to some minor topographic aberrations, but they are assumed to have a negligible effect on the mean watershed values for the topographic parameters examined in this thesis.

Another phenomenon encountered while compiling the watershed DEMs (I call it a phenomenon and not a problem because I do not believe it has any significant effect, even in its most extreme examples, on the analysis) was a "granularity" or a mesh of abnormally high and/or low values imposed upon the natural topography. This effect is due to the sparse sampling grid that was used for the construction of some DEMs, and is particularly noticeable in areas such as Ohio that have gentle, repeating topography (though it has been seen in much more topographically complex areas like Boulder, CO). While in its extreme cases it may lead to a slight increase in

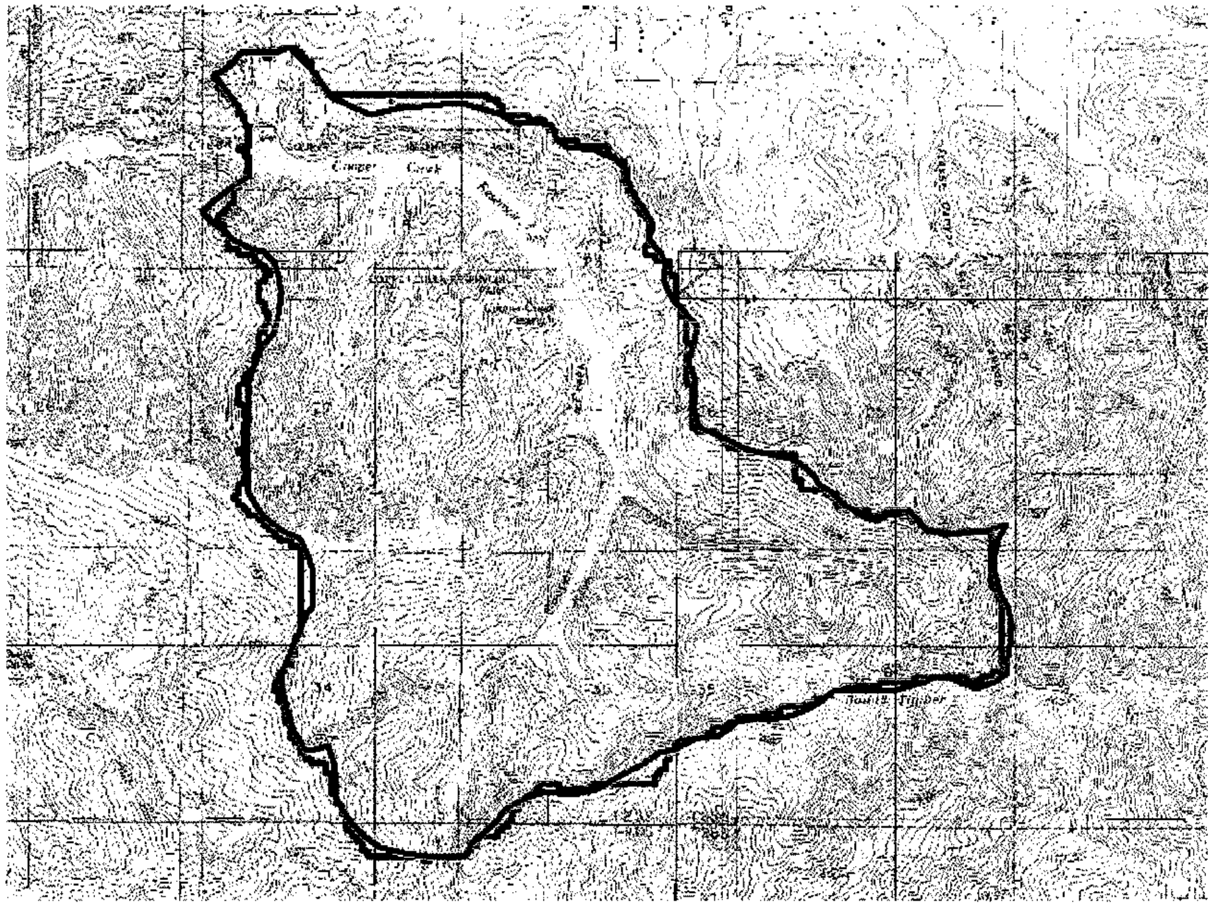


***Figure 2.6** – Seam effects on boundary delineation. The red line was created by automated watershed delineation from a DEM with two major seam errors (shown by blue lines). The black line is a basin boundary traced onscreen "by hand" using the digitized topographic maps comprising the base for this map. The difference between the two delineated basins is minimal.*

the mean slope of the affected watershed, it is generally assumed to be an insignificant effect.

There were 1,490 reservoirs with appropriate information for geolocation, though only 573 had sufficient information for statistical analysis (580 watersheds were ultimately delineated and characterized). 384 reservoirs could not be located because no candidate point satisfied the requirements of the geolocation program. In many cases, this is due to incorrect information (such as minimum watershed elevation) recorded in the RESIS database. The remaining watersheds were not delineated because one or more DEMs for the area surrounding the starting coordinates recorded for the reservoir had not been acquired at the time of program execution. While every effort was made to collect a complete set of DEMs, DEMs for some areas were still not available at the WebGLIS ftp site at the time the data was acquired, and the download interface at www.gisdatadepot.com prevents quick acquisition of large numbers of DEMs.

2.4 Basin Characterization

The section of the program that characterized the watersheds is a fairly straightforward data-mining operation. The watershed boundaries, once generated, are used to clip the grids described above and collect a variety of data for each reservoir-watershed system. A complete list of the variables collected, their source, units, and a brief description are included in Appendix A. Some further explanation of variables that either are not included in the regressions or are new to this type of research are discussed below.

The number of reservoirs (as reported in the NID) falling within the bounds of the watershed of each study reservoir was tabulated (*numres*). Although there is much evidence to show that upstream reservoirs can have a strong effect on sediment delivery to downstream reservoirs (Lajczak, 1996), this effect was not strong enough to cause the "*numres*" variable to appear in any regression models presented below. Nevertheless, any model that does not explicitly deal with upstream reservoirs, when upstream reservoirs are known to exist, should be treated with caution, as predicted sedimentation rates may be erroneously high.

The length of main flowpath through the watershed was calculated using the *Flowlength* command in Arc/Info. This length variable was used to calculate the relief ratio (relief / max. basin length) and was also highly correlated with total drainage area (r = 0.973). Because of this, *length* was assumed not to provide any additional information and is not included in the final regression model experiment. (Previous experiments in which it was included as an input variable showed that it was never selected by the step-wise regression process if total drainage area and relief-ratio were also included).

The number of candidate points is tabulated for each reservoir. This number was essentially a function of two variables: basin area and terrain complexity. The topic of terrain complexity and its effect on the basin delineation and characterization procedure are discussed in Chapter 3.

2.4.2 Quantifying Depositional Zones

I developed two variables describing topography that are not seen elsewhere in the literature. Several studies have suggested that hillslope and floodplain sediment storage may significantly affect basin yields, often accounting for more than 50% of the measured erosion (see Spomer, et. al. (1985), Spomer and Mahurin (1987), and Trimble (1999)). The following parameters represent two methods developed to quantify the extent of potential depositional areas within the basin using a single number derived from the basin topography.

The first parameter is defined as the percentage of the watershed classified as a potential "depositional zone" (*depzon*). A depositional zone is defined as a grid cell having both a slope of less than 2 degrees and a negative, or concave up, profile curvature. The second parameter used to estimate in-basin sediment retention is a "flatness" coefficient (*flat*). This value is calculated as the percentage of the watershed with slopes of less than half of the median watershed slope. It is theorized that a basin with a few regions of high slopes, but for which the majority of the area has low slopes stands a greater chance of storing mobilized sediment before it has a chance to reach the reservoir than a similarly sized watershed with an even distribution of medium slopes.

While the hypsometric integral (*HI*) gives some indication of the position of low slopes relative to the reservoir, an additional experiment was performed to more explicitly analyze the distribution of slopes in each watershed and compare that distribution to unit sedimentation rate. The mean and maximum slope values at four positions along the course of the watershed are tabulated using the "distance grid" (a grid, generated by the *distancegrid* command, that calculates the total flow distance from the basin outlet to every point on the grid) to divide the watershed into four parts. The results of this analysis are discussed later in this chapter. This sort of method could be used to quantify the position of low and high slopes, but because this requires a detailed watershed DEM, this value would be much more difficult to quantify for large numbers of watersheds than would a simple mean slope or the flatness coefficient, which could be calculated and applied at the county level for areas of limited topographic complexity. This would make it easier to apply the model to a large database such as the National Inventory of Dams.

2.4.3 Variables from the RUSLE Model

This analysis includes two parameters from RUSLE (the Revised Universal Soil Loss Equation) in the regressions with sediment data. RUSLE was initially designed by the USDA to estimate soil losses from agricultural fields (Renard, et al, 1997; USDA Handbook). By carefully examining soil losses with time from over 10,000 plot-scale measurements, the USDA developed the USLE (Universal Soil Loss Equation). RUSLE was then extended from agricultural settings to grasslands and woodlands, and there are now parameter values for all of the land-cover in the United States (Renard, et al, 1997).

The key to the model is the following simple multiplicative formula for calculating average erosion losses from a field:

$$A = R \cdot K \cdot L \cdot S \cdot C \cdot P \tag{2.2}$$

where A is the soil loss per unit area (tons/ac/yr), R is the rainfall-runoff factor, K is the soil erodibility factor, L is the slope length factor, S is the slope factor, C is the cover and management factor and P is the practice support factor. The L and S factors are usually combined to form a single factor that expresses the topographic potential of erosion. Although the factors are estimated empirically, some parts of the model like the LS factor can be estimated using physical first principles (Moore and Wilson, 1992).

Despite its plot-scale origin, RUSLE has some potential for estimating large-scale erosion over landscapes. The K- and *R-factor*s have been mapped throughout the United States. In the present study, RUSLE is not applied as a complete model because it is difficult to obtain accurate values of C and P for agricultural watersheds without having some knowledge about the style of agriculture in a given area. Instead, RUSLE parameters are used as descriptors of erosion conditions in the regression models.

R-factor

The R-factor examines the combined affects of rainfall volume and intensity. The amount of rainfall impacts erosion transport capacity, and the kinetic energy from high-intensity rainfall helps detach soil particles and make them available for erosion. The intensity of rainfall also determines whether there is infiltration-excess overland flow.

R values can be calculated by multiplying the maximum storm 30-minute intensity I_{30} by the storm energy E (in ft · tonf · acre^{-1}) for each storm and then summing the storms. E can be calculated by the following equation:

$$E=916+331\,log_{10}I$$

where I is the intensity in inches per hour. The maximum value of I is 3 in/hr because above this intensity, drop size and thus rain drop impact does not increase. R-factor values are also modified for snowmelt runoff.

The calculation of an R-factor for a national-scale watershed study would be tedious because it would involve analysis of local rain gauge records. Fortunately, values of R have been mapped for the United States and can be found in the RUSLE handbook. This map was scanned and digitized to produce a 1-km grid of estimated R-factor values for the continental US. The national scale-map is fairly well defined for the Eastern United States, but it is more difficult to predict R in the western United States because of the orographic effect of mountain ranges. The map of R which is used in this analysis varies from <10 in the desert southwest to >700 in Louisiana.

K-factor

In RUSLE, the K-factor is defined as the amount of erosion from a unit plot under continuous fallow under a representative range of storm conditions. A unit plot is a plot that is 72.6 ft (22.13 m) long with 9 percent slope, and continuous fallow is land that has been tilled and kept free of vegetation for more than 2 years. Because the plot is kept free of vegetation and measured under standard topographic and rainfall characteristics, the K-factor is only measuring the soils inherent erodibility, or lack of cohesion, without vegetation.

The K-factor has been estimated for soils throughout the United States. K values can also be calculated locally through the nomograph method by taking the amount of sand, silt, organic matter, soil structure and permeability and using the charts in (Renard, et al, 1997). The dependence of the K-factor on soil properties shows that erosion is dependent on soil texture and organic matter content. Statistical analysis of the data in the present work has also shown that K-factor is positively correlated with the occurrence of agricultural land ($R = 0.526$). However, it is not clear whether years of agriculture have lead to change in the nature of the soil (i.e. cohesiveness and/or organic matter) or whether the characteristics of highly erodible soils simply make them ideal for agriculture.

No attempt was made to convert R-factor and K-factor values to metric units. This was to circumvent the necessity of the complex conversion for future researchers dealing with similar values or comparing these watersheds to others. The multiple regression equations require all other values to be converted to appropriate metric units, as outlined in the variable list in Appendix A.

Chapter 3 - Terrain Complexity and the "Similar Watershed" Idea

While many of the watersheds have had their positions verified on topographic maps it was impossible to verify the positions of many of the smaller reservoirs. This chapter outlines an effort to quantify the potential errors encountered when one delineates a watershed that lies close to, but is likely not, the watershed for a particular reservoir.

Figure 3.1 (next page) shows the number of candidate points (as determined by the geolocation script) plotted versus the recorded drainage area in RESIS. It shows that a strong negative logarithmic relationship exists between these values. Given a constant nine-quad (22.5-minute) area, the maximum number of watersheds of a given size is constrained. As the size of the watershed in question increases, the number of potential watershed points in the nine-quad grid is reduced.

The scatter away from the trend in Figure 3.1 is not random but is geographically controlled. Almost all of the points that fall distant from this trend are located in the Pacific Mountains. In fact, the reservoirs in mountainous areas fall consistently further from the trend line than do reservoirs in flat regions. An examination of several DEMs for the areas surrounding reservoirs close to and far from the line shows that the number of candidate points per unit drainage area seems to be controlled predominantly by "terrain complexity" (though it is also controlled to a lesser extent by the relationship of the minimum watershed elevation to the mean elevation in the chosen DEM area). In other words, an area with steep mountains bounded by broad flat plains or an area with rolling hills cut by a large floodplain will

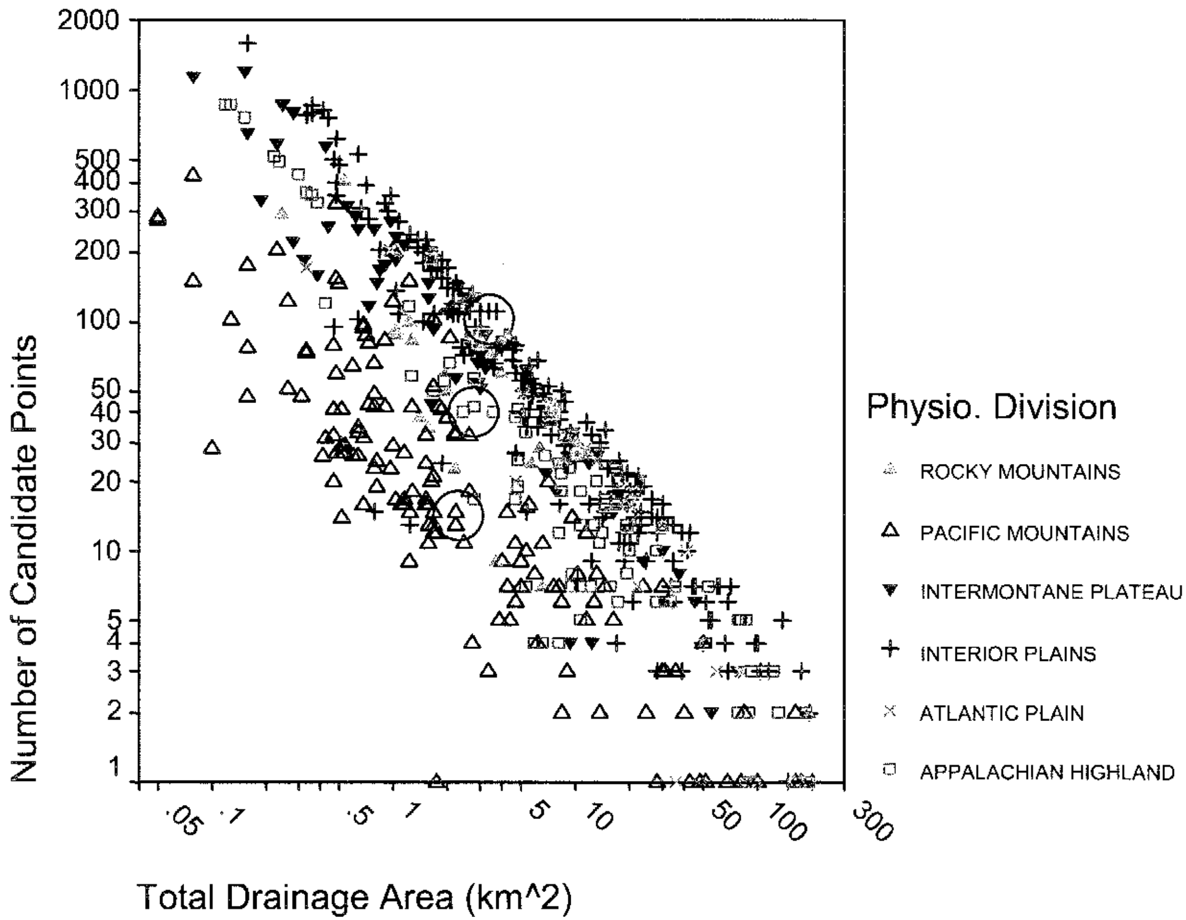


Figure 3.1 *– An illustration of the power law relationship between the number of candidate points and the total drainage area. This figure also shows how geographic/topographic complexity (here represented by physiographic division) affects this relationship. The circles mark the location of the three reservoirs whose surrounding areas are analyzed in Chapter 3.*

typically have a smaller number of watersheds that satisfy both the elevation and watershed area constraints imposed by the geolocation program. As is shown by the data presented below, this behavior has two competing consequences. Although more complex terrain areas have fewer candidate points, reducing the chances of picking the wrong watershed, the consequences to model development of choosing the wrong watershed in topographically complex areas can be significantly greater than in more homogenous topographies.

To quantify this effect, three areas were chosen to represent terrains considered complex, moderately complex, and simple. The level of complexity is assumed to be represented by the distance each point falls from the observed power law relationship shown in Figure 3.1. This distance has been quantified by fitting a line through the upper limits of the data scatter with approximately the same slope defined by this upper bound. The "complexity" number (C) was then calculated as the difference between the value predicted by this line (C_p) and the true number of candidate points (C_a) divided by true number of candidate points (Equation 3.1).

$$C = (C_a - C_p) / C_a \tag{3.1}$$

where C_p is predicted with basin area in km^2 (A) using Equation 3.2:

$$C_p = 25670A^{-0.9017} \tag{3.2}$$

For each area, the ten (10) candidate watersheds (as defined by the "candidates" coverage produced by the geolocation program) falling closest to the best available coordinates for the actual reservoir are chosen. Once delineated, each watershed is characterized by a set of relevant variables chosen to illustrate the potential for misrepresentation of the watershed.

The three area DEMs chosen are centered on the Pasadena, California quadrangle, the Lancaster, Kansas quadrangle, and the Toccoa, Georgia quadrangle. The locations of these three areas on the complexity plot (Figure 3.1) are indicated. The three areas represent complexity values of 0.37, 3.4, and 15.8 for Kansas, Georgia, and California, respectively. The locations and boundaries of the 30 watersheds considered are shown in Figures 3.2, 3.3, and 3.4.

The average distance of the nearest 10 candidate watersheds to the starting coordinates increases with terrain complexity. Near Pasadena, the watersheds are spread out along the foothills of the ranges and the distance from the starting point to the furthest watershed exceeds 18 km. The watersheds near Lancaster, however, are clustered tightly around the starting point and the distance between the starting coordinates and the most distant of the 10 nearest spillpoints is approximately 4.6 km.

Some summary statistics for each region considered in its entirety as well as the statistics describing the 10 watersheds from each area are presented in Table 3.1. The variability of nearly every topographic and climatic parameter characterized for the watersheds near Pasadena, California is higher than that seen for the watersheds near Lancaster, KS. The mean values and variation for the Toccoa, Georgia area watersheds generally fall between those for the other two.

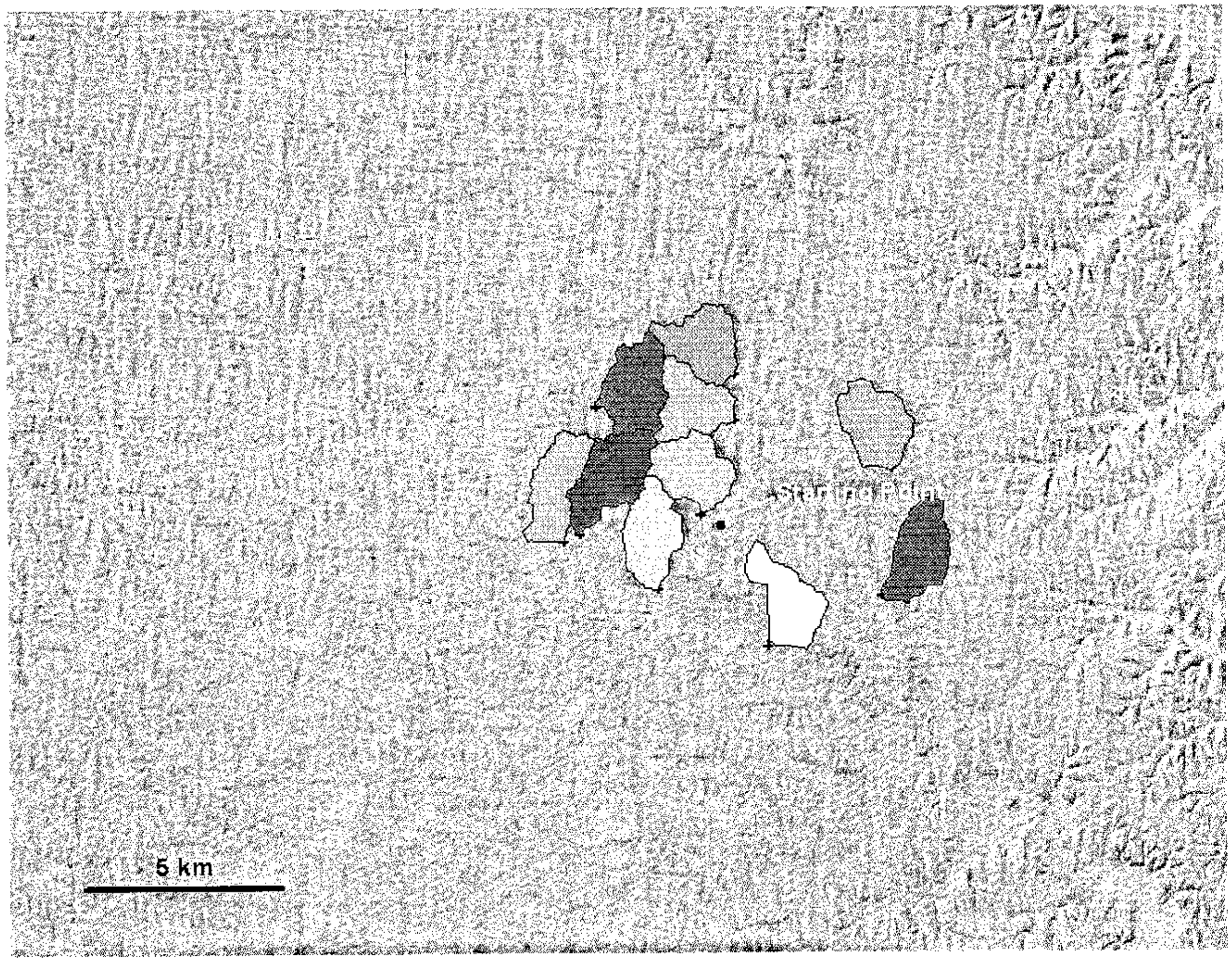


Figure 3.2 – *Candidate watersheds delineated near Lancaster, Kansas.*

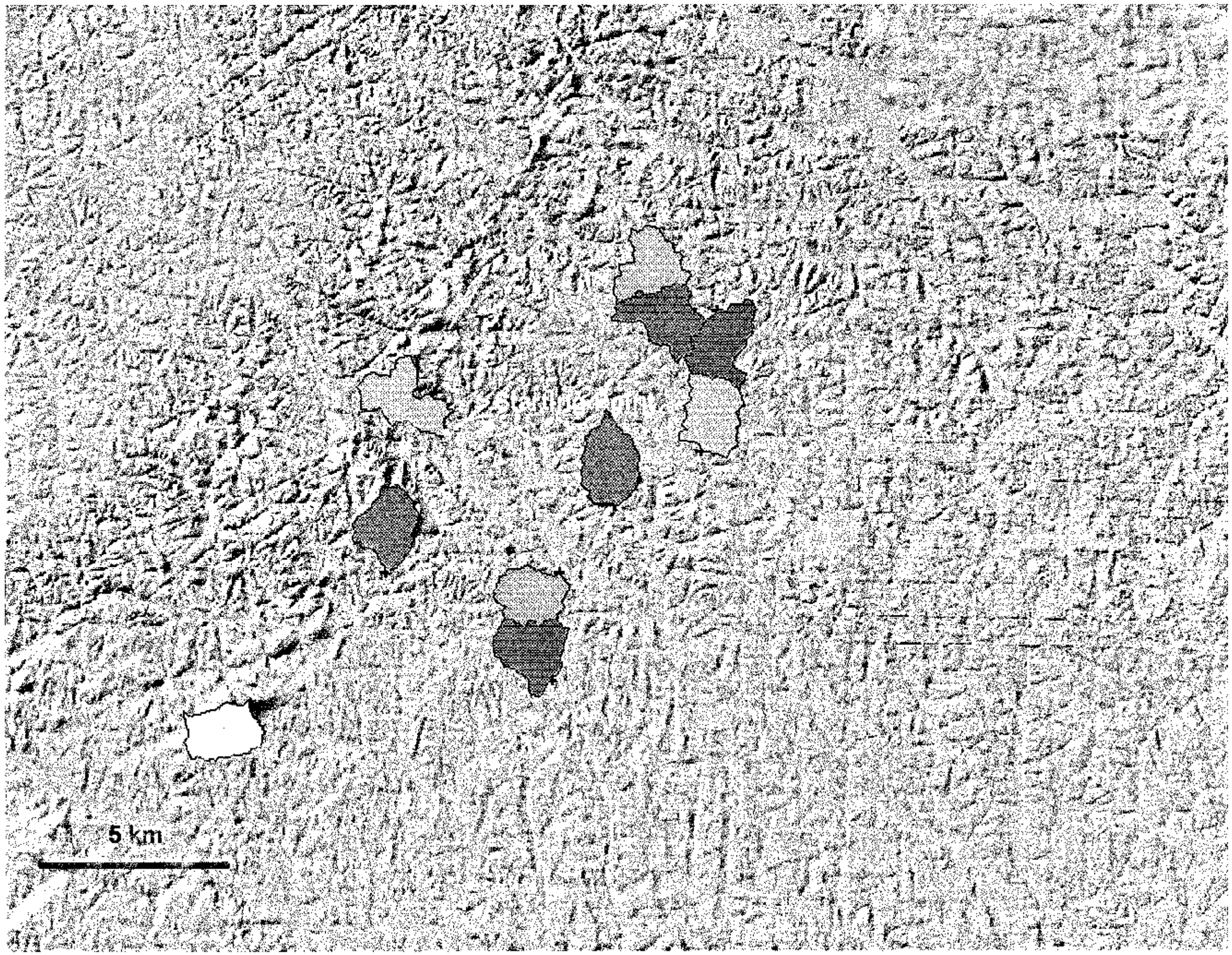


Figure 3.3 *– Candidate watersheds delineated near Toccoa, Georgia.*

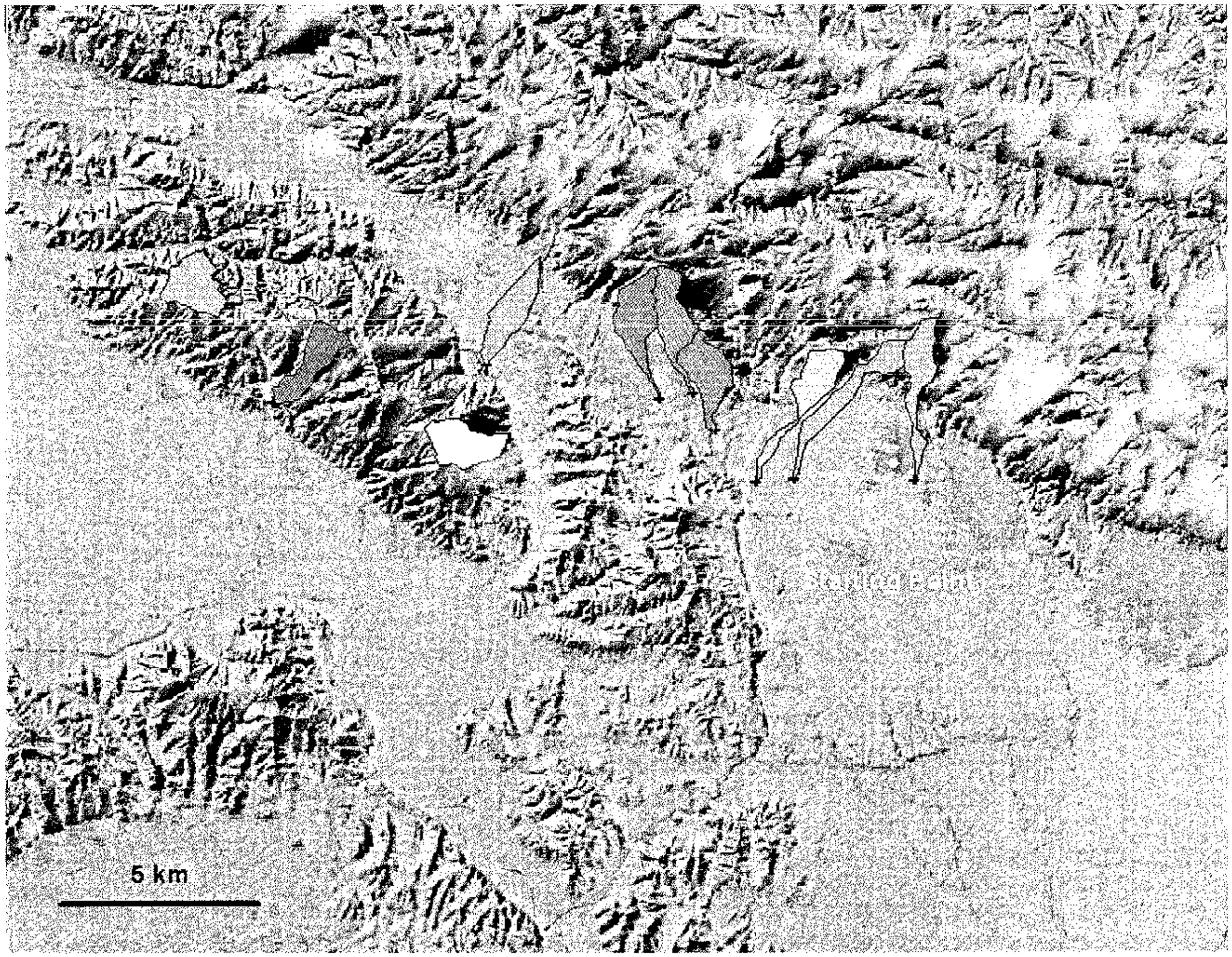


***Figure 3.4** – Candidate watersheds delineated near Pasadena, California.*

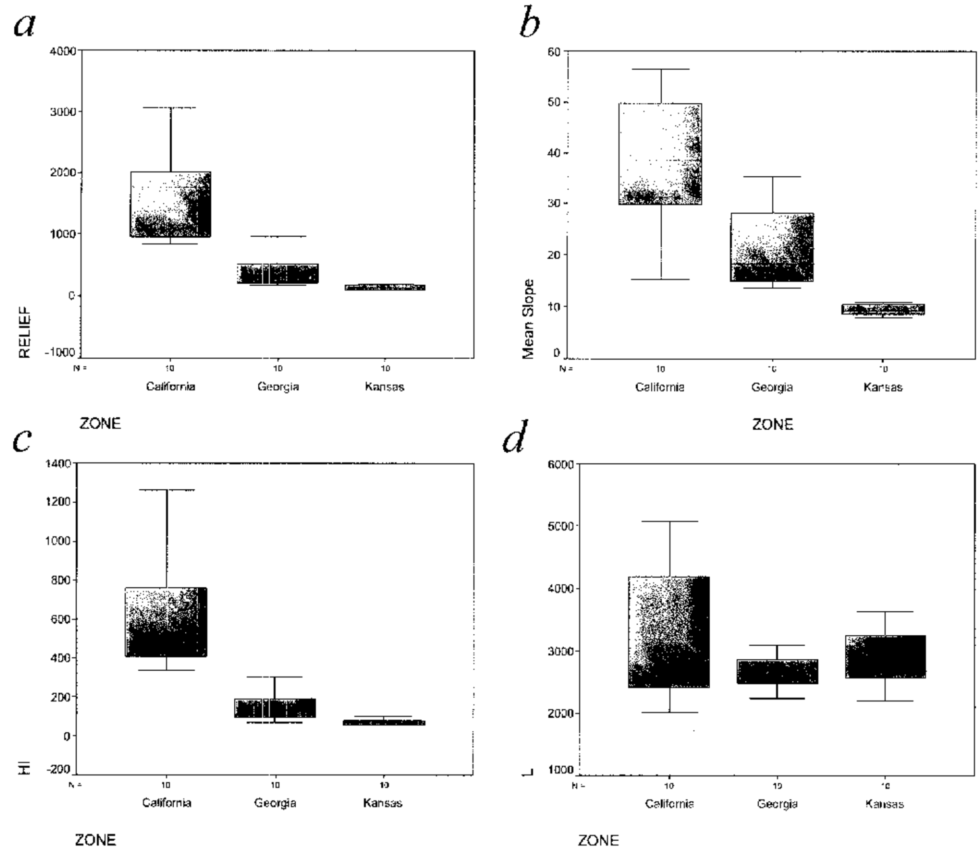
As is discussed in more detail later, topographic variables such as the flatness coefficient, hypsometric integral, and mean slope all contribute to the amount of eroded sediment that makes it to the reservoir. Table 3.1 illustrates that in every case the California watersheds show greater variability than the Kansas watersheds. For most variables, the Georgia watersheds show levels of variability that are more than the Kansas watersheds and less than the California watersheds.

Decreasing geographic complexity →

	California			Georgia			Kansas		
Land Use	*mean*	*st. dev.*	*coeff. variation*	*mean*	*st. dev.*	*coeff. variation*	*mean*	*st. dev.*	*coeff. variation*
Agriculture	0.01%	0.03%	3.16	3.42%	3.04%	0.89	51.68%	10.00%	0.19
Barren	0.26%	0.21%	0.79	1.29%	2.38%	1.85	0.00%	0.00%	-
Forest	13.02%	4.30%	0.33	72.22%	17.80%	0.25	4.08%	2.26%	0.55
Grasses	9.18%	5.66%	0.62	6.94%	5.74%	0.83	42.17%	10.39%	0.25
Shrubland	53.19%	22.19%	0.42	0.00%	0.00%	-	0.02%	0.06%	3.16
Urban	24.36%	22.18%	0.91	15.56%	16.03%	1.03	0.84%	1.37%	1.64
Water	0.00%	0.00%	-	0.56%	0.76%	1.36	1.22%	0.42%	0.34
Soil/Climate	*Sum*	0.55	6.23	*Sum*	0.46	6.20	*Sum*	0.24	6.13
Mean K-Factor	2.69E-01	1.21E-02	0.04	2.88E-01	1.41E-03	0.00	3.66E-01	5.69E-03	0.02
Mean Rain	5.73E+04	5.98E+03	0.10	1.48E+05	4.46E+03	0.03	9.30E+04	4.18E+02	0.00
Mean R-Factor	9.65E+01	9.98E+00	0.10	2.72E+02	9.27E-01	0.00	1.92E+02	2.97E-01	0.00
Topography									
HI	658	285	0.43	141	79	0.56	71	17	0.24
L	3387	1059	0.31	2662	259	0.10	2903	461	0.16
O	5.4	0.52	0.10	6	0	0.07	6	0	0.07
RELIEF	1660	742	0.45	339	257	0.76	133	30	0.22
SLOPE_MEAN	39.0	13.6	0.35	21	8	0.36	9	1	0.12
SLOPE_STDV	16.6	5.7	0.34	10	3	0.29	5	1	0.26

Table 3.1 *– Yellow boxes highlight the standard deviation of the dominant land use classes for each region. Red boxes highlight the highest coefficient of variation for each soil, climate, and topographic variable and green highlight the lowest coefficient.*

Figures 3.5a-d show boxplots for four of the topographic parameters considered. All four plots illustrate the more dramatic topography seen in the California watersheds compared to the Kansas watersheds. Although the standard deviation for the California watersheds is large, the values for relief, mean slope, and hypsometric integral for all California watersheds is distinguishable from the 10 Kansas watersheds. The plots also show that most of the Georgia watersheds have means between those of the other two regions for all of the parameters shown, though the mean total stream length (L) for the watersheds of each region are similar despite differences in complexity.



Figures 3.5a-d *– The variation of several topographic variables for each zone.*

Examining the land-cover values, the coefficient of variation (standard deviation divided by mean) for the dominant land classes in each area is a function of that area's complexity. Shrubland is the dominant land-cover in the mountainous watersheds of the Coastal Ranges of Southern California and shows a standard deviation of ~22%. The dominant land-cover of Georgia, forest, shows a standard deviation of ~18%. Dominant land-cover in the Kansas watersheds is divided between agriculture and grasses (pasture). The standard deviation for each is about

10%. The sum of the standard deviations and coefficients of variation for all land-cover classes of each region is also shown. This presents a measure of overall variation of land-cover in each region. These values also illustrate the observed pattern of greater variation in the California watersheds and lesser variation in the Kansas watersheds.

This complexity factor is merely an approximate number representing the variation from the normal number of expected candidate points based on the given watershed area. Caution should be used when interpreting these results. For instance, it is quite possible for a watershed of a given size to fall into a homogenous (i.e. simple) terrain, but to have a low number of candidate points (leading to a high complexity factor) because the reservoir falls into an unusual topographic feature (often a floodplain) that does not otherwise characterize the topography of the region. This is shown by the shaded relief map of candidate points for reservoir number 31021 (DEM – Blue Rapids Northeast, KS) shown in Figure 3.6 below. Thus, it might be more appropriate to say that the complexity factor is a combination of both the relative geographic complexity of an area as well as the degree to which a given watershed is representative of the local topography.

There is a slight negative trend between the calculated complexity and the total drainage area of the watershed. This appears to be due to the large number of watersheds (n = 97) in Los Angeles County, California with watershed areas (mean = 958 ha) below the mean for the entire database (mean = 1673 ha). The mean estimated complexity ratio for these watersheds is 16 compared to the database mean of 5.7.

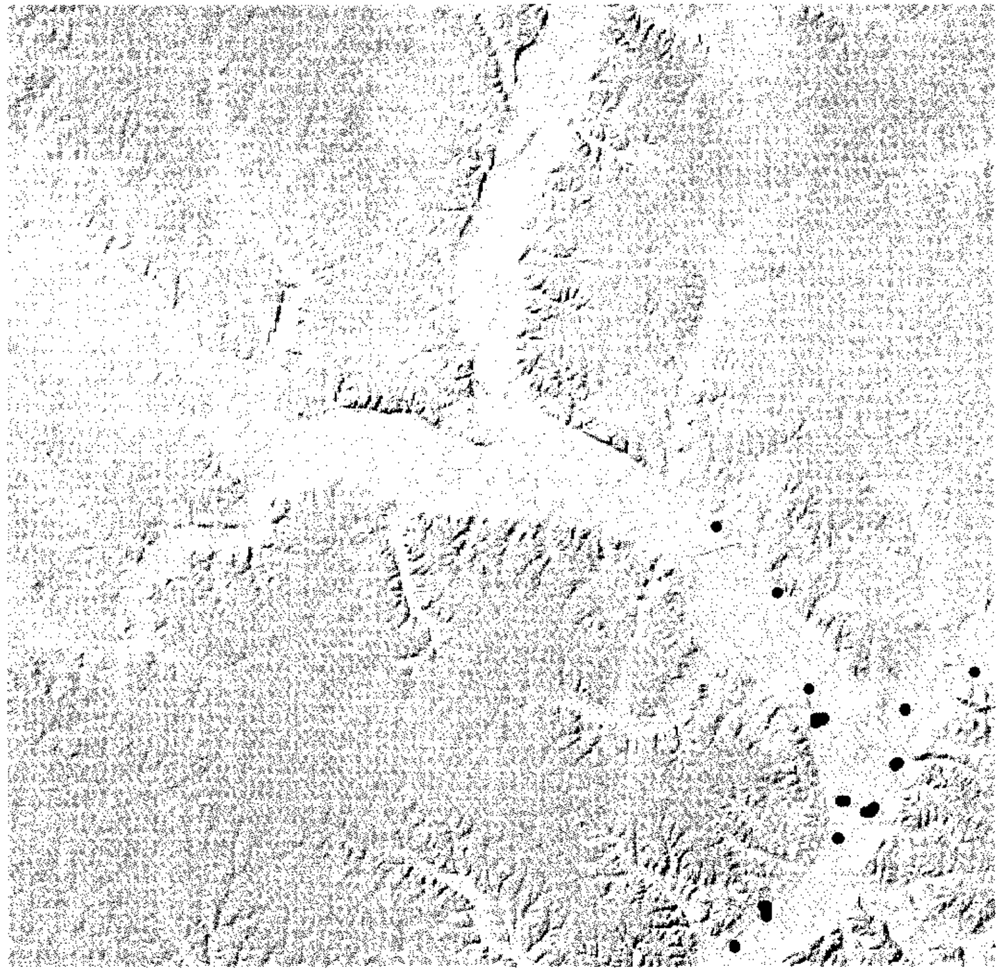


***Figure 3.6** – An example of how the complexity value can be influenced by reservoir location. Blue dots represent "candidate points."*

In summary, it seems that variation of topographic parameters between nearby watersheds in topographically complex areas can be expected to vary by up to an order of magnitude more than the same parameters in a topographically simple area. Land-cover variables show a similar trend, with more than twice the variation in the topographically complex region. The extent to which this variation might affect regression models developed for each area type (complex, moderate, simple) is difficult to quantify, but is discussed more in the section describing the models.

Chapter 4 – Statistics and Model Results

All statistical analyses are performed considering all reservoirs grouped together and all reservoirs grouped by physiographic division. The physiographic divisions used are based on the "Physiographic Divisions of the United States" map created by Fenneman and Johnson (1946) and were acquired from the US Geological Survey Water Resources NSDI Node: http://water.usgs.gov/lookup/getspatial?physio. This dataset divides the continental US into 8 physiographic divisions, 24 provinces, and 75 sections. Although some of the physiographic sections contain sufficient numbers of reservoirs for regression analysis, the majority did not. Similarly, many of the physiographic provinces are not well represented enough for regression analysis. Thus, the physiographic divisions are used to group reservoirs into areas of similar physiography. Because small numbers of reservoirs are located in the Laurentian Uplands and Interior Highlands divisions, these divisions were not used and the reservoirs in them were grouped with the Appalachian Highland reservoirs. The remaining six divisions are used to group the reservoirs in the regression analyses presented below.

Many sets of statistical analyses were performed upon the data prior to settling upon a final set of models. Only the most pertinent results are presented in this thesis. Regressions, correlations, and factor analyses were all performed in the SPSS for Windows software package, version 7.5.1. All parameters from the RESIS database that were required for the statistical analysis were summarized and added to a single dBase IV table with the information collected by the geolocation scripts.

Three subsets of data were created. The total number of reservoirs in each subset and the number in each physiographic division are shown in Table 4.1. The first subset included all reservoirs for which complete watershed data were available. Because it could not be done for all reservoirs, no attempt was made to separate those reservoirs that may have incorrectly delineated watersheds from those that were checked for accuracy. This subset contained 535 reservoirs. A map showing the location of the subset 1 reservoirs is presented in Figure 4.1 (next page).

	Total	Appalachian Highlands	Atlantic Plain	Interior Plains	Intermontane Plateaus	Pacific Mountain System	Rocky Mountain System
Set 1	535	89	17	203	57	146	23
Set 2	415	64	9	151	43	127	21
Set 3	281	57	9	132	26	47	10

Table 4.1 *– Total number of reservoirs in each subset and the number for each physiographic division.*

The second subset is the same as the first, but only those reservoirs with at least 10 years of elapsed time between the original and final survey are included. This was done because the small watersheds considered in this study are likely to be quite sensitive to the effects of extreme events. Thus longer records are more likely to integrate the short-term effects of extreme events with the long-term trend of slower sediment accumulation giving a more reliable average unit sedimentation rate. This reduced the total number of reservoirs to 415.

For the third subset, a simple averaging technique was tested to both reduce the number of reservoirs in those areas with high numbers of reservoirs and to reduce the effects of accidental basin misidentification. Bahr and Syvitski (*unpublished*) point out that small streams only partially obey Horton's Laws as

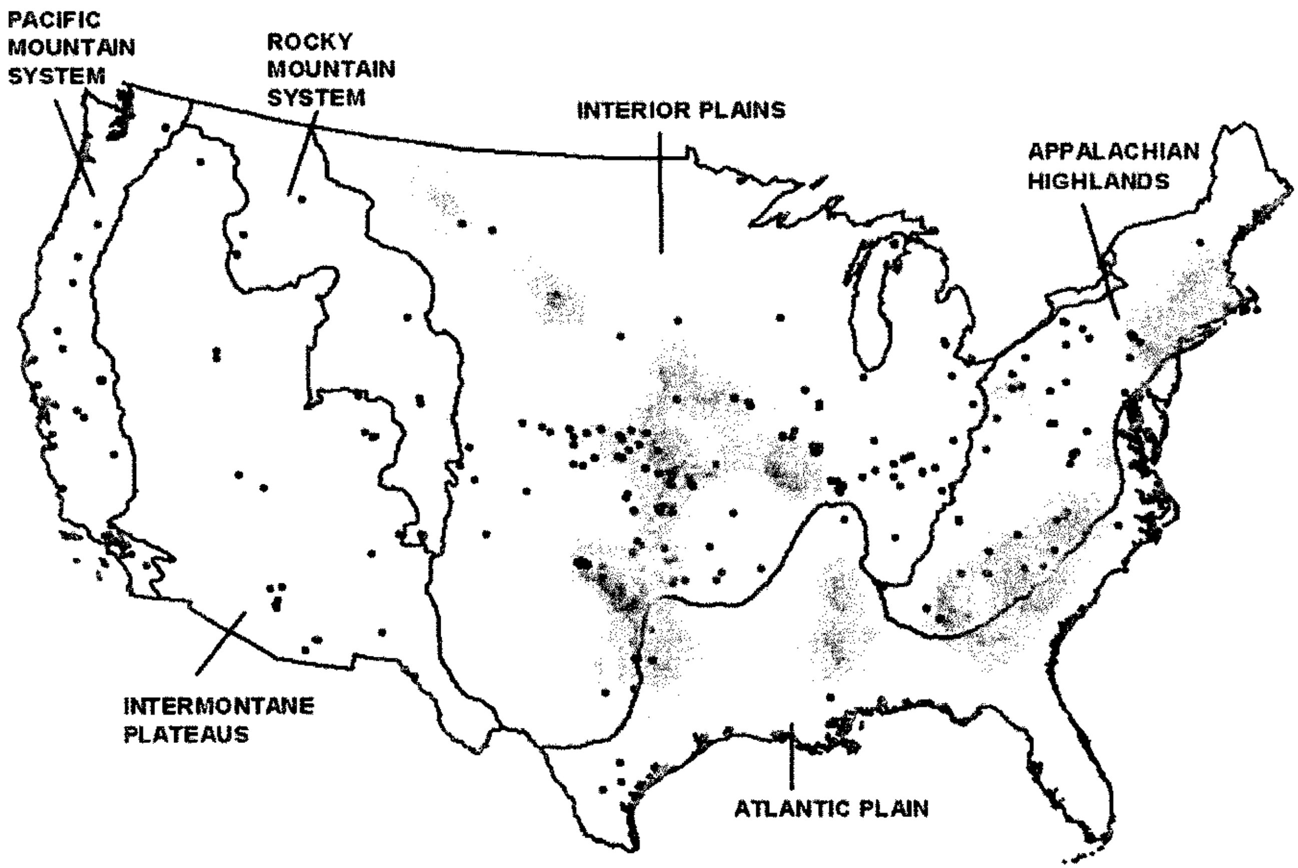


Figure 4.1 *– Location of the 537 reservoirs used in this analysis and the outlines of the 6 physiographic divisions into which they fall. In red is the relative density of NID reservoirs for comparison (darker red indicates a higher density of reservoirs).*

concerns discharge, sediment load, basin area, basin relief, etc. For rivers (defined as streams with discharge larger than 30 m^3 s^{-1}), Horton's Laws are much more stable and less subject to fluctuation due to local conditions. They find that if the average characteristics of many smaller streams are used, the Horton relationships are more likely to be satisfied by these smaller streams.

Most of the watersheds in this study support only small to medium sized streams with watersheds of less than 190 km^2. To test the possibility that this averaging technique would improve the performance of the regression model for this study, the reservoir unit sedimentation rates and watershed characteristics of small watersheds of similar size that were near each other (within ~ 2 km, more in certain cases) or were determined to be tributaries of the same medium-order stream are averaged. When the proximal watersheds are averaged, there are a total of 281 watersheds.

For a description of all variables considered in regressions and factor analysis, see Appendices A and B. Several variables, as noted in Appendix A, were not normally distributed. All non-normal variables were positively skewed, and were log transformed to produce a more normal distribution prior to regression or factor analysis. For the purposes of this thesis, the phrase "unit sedimentation rate" will refer to the volumetric accumulation rate per unit watershed area (in m^3 km^{-2} yr^{-1}) uncorrected for trap efficiency.

4.1 Discussion of Land-cover characterization

Land-cover, as derived from the National Land-cover Dataset (NLCD) was quantified in several different ways for each reservoir to determine if one method of quantification would prove more useful in determining unit sedimentation rates than another. The first method was simply to tabulate land-cover type distribution as the percentages of the watershed that fell into each of the eight land-cover types (see Appendix B). The second method involved tabulating the land-cover type distribution in the same way for only the lower 25% of the watershed. This was done to examine the extent to which the land-cover type in close proximity to the reservoir may affect sediment yield. Finally, land-cover percentage for the high-slope areas (>10 degrees) was tabulated as both the land-cover percentage of the total high-slope area and as the percentage of the watershed in which both high slopes and the land-cover type are indicated. This method of tabulation demonstrated that forests, for instance, were the most dominant land-cover type in the high-slope regions of many watersheds.

4.1.1 Land-cover relationships for the entire US

For both the averaged and unaveraged complete datasets, forest cover shows a negative correlation with unit sedimentation rate, regardless of the method of characterization. However, both datasets show a modest strengthening of this effect if forests are located near the reservoir.

Agriculture, barren land, and grassland do not exhibit strong correlations with unit sedimentation rates except for a weak, but significant, negative correlation for grasses on high slopes.

For sets 1 and 2 (the unaveraged datasets) shrubland is consistently positively correlated with unit sedimentation rates, and even more so if the percentage of the watershed having both shrubland and high slopes is considered. However, when averaging was employed, a process that greatly reduced the number of watersheds from Los Angeles County (the locations with the highest percentages of shrubland, high slopes, and sediment output), the significance of the correlation between shrubland and unit sedimentation rate is all but eliminated except for the value describing the watershed fraction that is shrublands and high slopes.

All data sets, including those averaged, show a progressively increasing weak, but significant, correlation between urban-lands and unit sedimentation rates as total urban-land percentage is considered alone, on high slopes, and finally near to the reservoir. However, the improvement of the correlation coefficient between the lowest and highest (r = 0.112 and 0.258, respectively for Set 2) is considered modest. The results still suggest that urban-land has the strongest influence on unit sedimentation rates when located near the reservoir.

For all sets, there is a slight positive correlation between unit sedimentation rates and the percentage of land classified as water and the percentage of water near the reservoir. This is actually opposite of what was expected. Wetlands, natural lakes, and reservoirs are classified as water in the NLCD and all were assumed to serve as sinks, reducing overall sediment delivery to the reservoir. Of course, percent water near the reservoir for many watersheds included the area of the reservoir itself. As is discussed later, as reservoir capacity (and thus area) goes up relative to basin area, the trap efficiency (and thus unit sedimentation rate) also goes up. Therefore, it is likely that the value describing percent water near the watershed is also serving to some extent as a proxy for the capacity-watershed ratio. Visual inspection of a scatter plot between these two parameters confirmed their positive relationship.

4.1.2 Land-cover relationships by physiographic division

The effect of land-cover on reservoir unit sedimentation rates becomes more apparent when examining these relationships with reservoirs grouped by physiographic division. However, while the correlations between total watershed land-cover and unit sedimentation rate are found to generally increase after grouping, the difference in correlation coefficients between the various methods of land-cover representation for a given land-cover type is less pronounced, suggesting that a general tabulation of land-cover percentage for the entire watershed is probably sufficient for unit sedimentation rate prediction.

Agriculture and nearby agriculture show significant (at the 0.05 level) positive correlations with sediment output for the Rocky Mountain division, but the difference between the two is negligible. The same is true for barren land and nearby barren land for the Rockies.

Forestland shows a strong negative correlation with unit sedimentation rate for the Appalachian Highlands, Pacific Mountains, and the Rocky Mountains. Only for the Appalachian Highlands did consideration of high-slope land-cover improve the correlation. Both the Interior Plains and Intermontane Plateaus show a significant positive correlation between nearby forests and unit sedimentation rates.

The influence of grass and pastureland was inconsistent from region to region. Grass and pasture land and nearby grass and pasture land show a strong positive correlation with unit sedimentation rates in the Atlantic plain, and all representations of grass and pasture correlated positively with unit sedimentation rates in the Rocky Mountains. Conversely, grass and pastureland in the Pacific Mountains and Interior Plains show a negative correlation with unit sedimentation rate (significant in the Interior Plains only for total percent and nearby percent). These inconsistencies could be the result of several factors, including sampling bias. Perhaps the most likely source of this inconsistency is the varied use of land classified as grasses in this study. While some of this land is likely to be hay fields, natural prairie, lawns, golf courses, and sports fields (probably not big sediment producers) other areas are likely post-fire successional grasslands or cattle rangelands (areas known for inconsistent but often extremely high sediment yields). It is beyond the scope of this study to further examine the relationship between grassland and unit sedimentation rates, but it perhaps bears looking into in future analyses.

Shrubland and nearby shrubland in the Intermontane Plateaus show a negative trend with unit sedimentation rate, but the Pacific Mountains show a very strong positive correlation between shrublands (of similar levels despite the method of tabulation) and unit sedimentation rate. It is likely that the strong correlation of this land-cover type with unit sedimentation rates in the Pacific Mountains is due more to the occurrence of repeated or major fires in these areas and their occurrence on extremely high slopes than to the nature of the cover itself. This finding is not inconsistent with the findings of other authors (Eakin, 1939).

When divided into physiographic regions, the effect of the urban-land and water land-cover types is weak and inconsistent. While positive correlations between total urban-land percentage are seen for the Appalachian Highlands and Intermontane Plateaus and for nearby urban-land percentages for Pacific and Rocky Mountains, watershed averaging weakens all correlations below the level of significance except for the positive correlation of total urban-land in the Intermontane Plateaus.

The positive correlation between unit sedimentation rate and total water and nearby water percentages seems to be mainly due to the Interior Plains reservoirs, which are the only reservoirs showing a significant correlation. Again, this is most likely due to the percentage of water acting as a proxy for the capacity watershed ratio, and thus trap efficiency.

In summary, it seems that the effects of land-cover on reservoir sedimentation, while not the most influential factors, are certainly detectable even with this somewhat simple approach. A more comprehensive estimate of land-cover and land-use history would provide even more insight into the variability of unit sedimentation rates in otherwise similar watersheds. Also, while the classification of high-slope land-cover seems to give higher correlations in some cases, for the majority of cases the quantification of the land-cover of the entire watershed or the land-cover near the reservoir provide similar and more reliable results. Because these values are also the most easily obtained for large datasets, it is concluded that the overall watershed land-cover values are the best for model development.

4.2 Factor Analysis

Although factor analyses of a variety of parameter and reservoir groupings were employed to better understand parameter covariance, the results of most of the factor analyses are not presented as they do not provide any significant information not already covered by the multiple regression analyses discussed below. However, the factor analysis technique was used to aid in the consolidation of the land-cover variables during an analysis of a smaller portion of the database. Because the groupings derived from this analysis are fairly obvious (deciduous, coniferous, and mixed forest types are all fairly covariant; light and heavy residential areas and commercial areas are also covariant), the analysis will not be outlined here. Appendix B shows the original NLCD land-cover classes and the reassigned classes in bold letters heading each group.

Factor analysis results are not relied upon exclusively for land-cover reassignments. For instance, urban grasses, which were loaded highly on the same factor as high and low intensity residential land, were nevertheless grouped with other grasses because of the much different influence they are likely to have on runoff, and thus erosion, compared to other residential land-cover types (which are assumed to provide little sediment, but high runoff due to the presence of pavement and man-made drainage ditches).

Factor analysis was used to evaluate the suitability of the physiographic divisions used to divide the dataset. Figure 4.2 (next page) shows the standard deviations of the factor loadings of each reservoir on factor I of an R-mode analysis

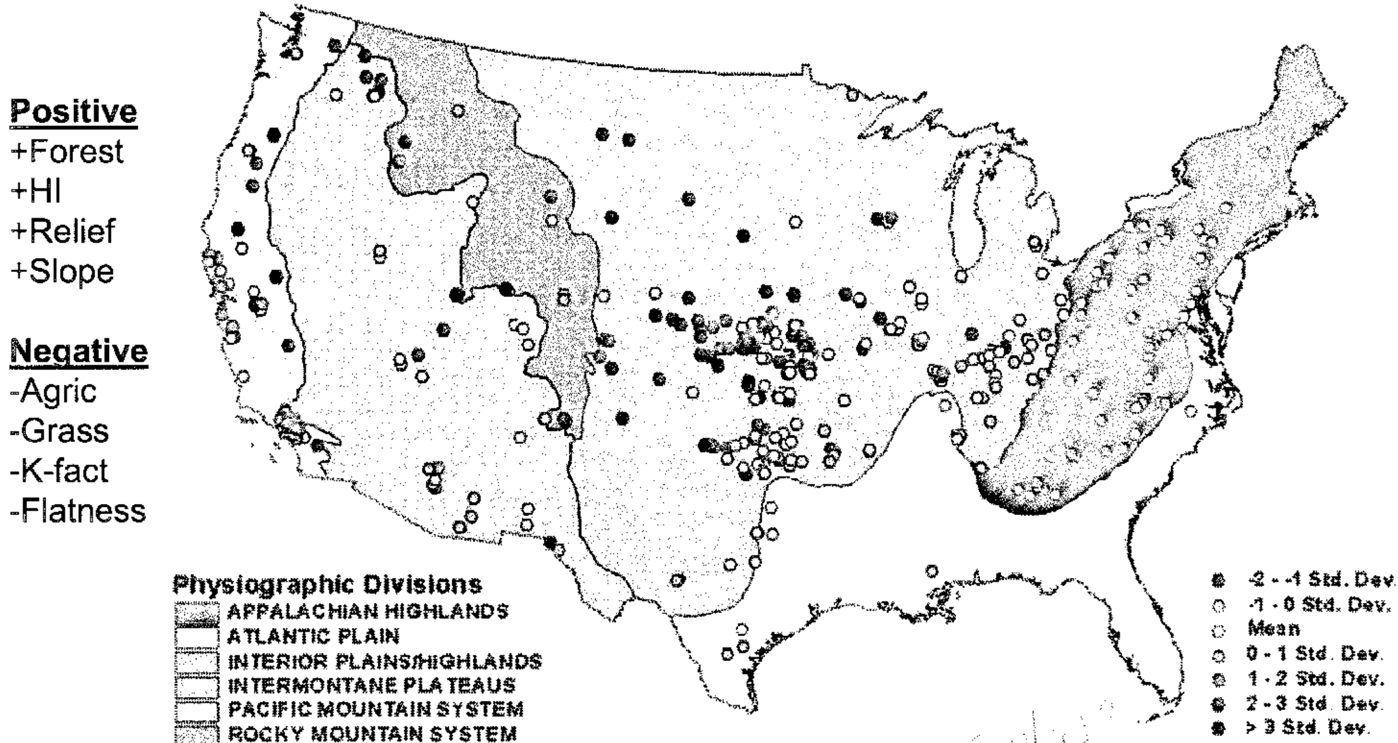


***Figure 4.2** – A plot of the standard deviations of the Factor I scores for an R-mode factor analysis of the entire dataset. Parameters that have the strongest positive and negative loadings for this factor are indicated at left.*

done on the entire dataset. The rotated component matrix for this analysis is shown in Table 4.2.

Rotated Component Matrix

	Component					
	1	2	3	4	5	6
AGRIC	-.609	-.204	-.270	-9.245E-02	.261	-.172
BARREN	5.379E-02	6.088E-02	.123	5.707E-02	-.673	-.328
BDL	8.297E-02	-.148	-.476	.191	-7.314E-03	.531
DEPZON	-.846	-3.045E-02	3.615E-02	.190	5.130E-02	-1.700E-02
ELVMAX	.355	.820	.217	1.820E-04	-5.956E-02	-.104
ELVMIN	.178	.777	3.079E-02	-.210	-.136	-.115
FOREST	.431	-.194	.758	1.320E-02	.142	-8.062E-02
GRASS	-.108	-.157	-.752	4.399E-02	.214	-.174
KFACT	-.406	-.451	-.542	-5.890E-02	-4.279E-02	4.426E-02
L_C/W	-3.280E-04	-.557	-.483	.344	-.220	.221
L_OMH	-1.838E-02	-7.852E-02	8.843E-02	8.457E-02	.708	-.158
L_R	.616	.503	.437	.348	2.024E-02	-.103
L_RR	.617	.521	.405	-.316	-1.838E-02	-8.156E-02
L_SL	.794	.356	.408	7.407E-03	5.461E-03	-.112
L_AREA	-3.473E-02	-.140	1.255E-02	.949	8.316E-02	-9.498E-03
LN_HI	.625	.479	.434	.316	9.139E-02	-.146
LNATB	-.868	-6.420E-02	2.423E-03	1.904E-02	3.601E-02	.143
RAIN	.110	-.661	.550	2.382E-02	.309	-1.904E-02
RFACT	-.166	-.874	-5.236E-02	9.138E-02	4.394E-02	8.370E-02
SHRUB	.113	.625	-1.773E-02	1.702E-02	-.539	.175
FLAT	-.529	-.254	-.270	.514	-.154	4.470E-02
URBAN	-.148	-9.753E-02	.122	-5.729E-02	-1.723E-03	.769

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 14 iterations.

***Table 4.2** – Rotated component matrix from R-mode factor analysis of all study reservoirs in the U.S. The scores for factor I are plotted in Figure 4.2. Note that Factor I describes covarying aspects of both topography and land cover.*

Factor I has high positive scores for variables such as percent forest, HI, relief, and slope and high negative scores for percent agriculture, percent grass, K-factor, and flatness coefficient. In essence, this map shows that most reservoirs in a given region have a group of variables that tend to vary similarly and these covariances are different from region to region. The Intermontane Plateaus and Rocky Mountain System show the widest internal variation in Factor I scores. The Pacific Mountain

System reservoirs typically show strong positive scores on Factor I. The Appalachian Highlands have moderate scores and the Interior Plains typically show lower scores.

Variation of nearly every variable considered across the entire US is large and the multiple regression equations derived from the entire database are expected to be different than those developed for each region, where variations for certain variables are high, while other variables are held nearly constant. This should then influence the multiple regressions accordingly, providing a model that explains how the differences within a given region influence unit sedimentation rates.

4.3 Multiple Regression Analysis

4.3.1 All Reservoirs Considered Together

The same set of regressions was performed for the entire dataset as was performed for each physiographic division. The results of the analysis of the entire dataset are presented first, followed by discussions of the unique results for each physiographic region. Because of the unbalanced effects that extreme events are likely to have on the unit sedimentation rates for reservoirs with short records in Model Set 1, which includes all delineated watersheds regardless of the length of the sedimentary history for the reservoir, this set is not discussed unless its behavior seems significant.

Step-wise multiple regression was used to create the models. This regression method adds variables progressively based on their cumulative ability to explain the variation of the dependent variable. The variable that explains the most variance at each step is added until the set significance of the F value is exceeded (Norusis, 1999). The initial step-wise regression models were constructed using the standard significance of the F value of 0.05 and 0.10 for inclusion and exclusion of variables. Experimentation showed that the model R^2 values and standard error of the estimates were both sufficiently improved by increasing these values to 0.10 and 0.20 to justify the possible negative effects. Increasing the F significance for entry and removal introduces a greater probability of including a variable that does not truly influence sediment output, but because all of the variables considered in the model have been shown to influence sediment yields in previous research, the chances of this type of error were thought to be small. The adjusted R^2 and standard error of all of the regressions are improved by these changes.

The regression results that show the addition of each variable and the variable's effect on model performance are included in Appendix E (Tables E.1 – E.6). Table 4.3 (next page) shows the coefficients of selected models.

The best single model created to predict reservoir unit sedimentation rates for the entire US has an R^2 of 0.668. The standard error of the estimate is 0.921. Because the error is given in log units of reservoir sedimentation, an error of less than 1.0 indicates that the model can, on average, predict unit sedimentation rates within an order of magnitude. It should be noted that this means, for a single reservoir, a predicted unit sedimentation rate of 9.0 $m^3\ km^{-2}\ yr^{-1}$ with a standard error of 1.0 log units represents a range of possible unit sedimentation rates from 0.9 to 90 $m^3\ km^{-2}\ yr^{-1}$, but the error on the average of many reservoirs is much smaller.

avg rate = 9 m³/km²/yr
could be as low @ 0.9 or as high
as 90 →

The parameters accounting for most of the variation in this model (Model Set 2a) are the logarithmic capacity-watershed ratio (+), percent forest cover (-), the flatness coefficient (-), R-factor (+), and soil bulk density (+). The full model also includes logarithmic mean slope (+), K-factor (+), logarithmic relief (-), maximum elevation (+), percent depositional zone (+) (which replaces the flatness coefficient in the final model), and percent agriculture (+). Most of the variables used for model creation have been found to correlate with unit sedimentation rates or basin yield in previous studies. However, the influence of certain variables is not necessarily consistent with previous experience. The influence of several individual variables is discussed in more detail in later sections.

All of the models for the entire dataset (Set 1a, 2a, and 3a) exhibited some similarities. The model that considers all reservoir-watershed pairs regardless of the length of record (Set 1a) is similar to the other models in the parameters it includes, but the R^2 (0.572) and standard error (1.02) are the worst of the three models. The final step-model also requires the greatest number of input variables (n = 13) of any of the three models.

The averaging technique employed for Set 3a did not seem to have a positive effect on the model developed for the entire US. The standard error is about the same, increasing from 0.921 for Set 2 to 0.927 for Set 3a and the R^2 is worsened, from 0.668 for Set 2a to 0.565 for Set 3a. In addition, the degrees of freedom for the final model for Set 3a were only 251 (12 variables are included in the model for 264 reservoirs) compared to the 315 for Set 2a (10 variables are used to predict unit

	R^2	Std error
M1	0.572	1.02
M2	0.668	0.921
M3	0.565	0.927

Model Coefficients

			Land Cover					Soils			Topography									Climate	
	constant	L CAP WAT	AGRIC	BARREN	FOREST	SHRUB	URBAN	K-FACT	SOIL DENS.	L OM	DEP ZONE	FLAT	HI	L RELIEF	L REL-RATIO	L SLOPE	L AREA	ELEV MAX	ELEV MIN	MEAN RAIN	R-FACT
Model 1a	0.4	0.32	1.1		-1.0		1.3	2.9	2.4	-0.2		-0.5		-0.5		0.8		8.7E-04	-5.0E-04		4.1E-03
Model 2a	0.6	0.36	0.7		-1.2			2.1	2.1		3.9			-0.6		1.1		5.5E-04			4.9E-03
Model 3a	-0.9	0.27	0.9		-1.0			1.6	3.0					-0.7	-0.4	1.3		9.5E-04	-6.2E-04	-8.70E-06	6.5E-03
Model 2																					
APP	0.8	0.27				67.5						-1.3	1.86		-0.6				-4.1E-03	1.43E-05	
ATP	3.1	0.37																-2.6E-02	4.1E-02		
INT	2.0	0.51	0.8		-0.9									-0.4	-0.3	0.9					4.3E-03
PLT	4.1	0.54				-1.4	21.6			0.3							-0.3				2.5E-02
PAC	0.2	0.29				2.3		-15.4	3.6		14.0										2.9E-02
RKY	-1.8	0.27		12.2													-0.2				4.4E-01
Model 3																					
APP	0.8	0.27				67.5						-1.3	1.86		-0.6				-4.1E-03	1.43E-05	
ATP	3.1	0.37																-2.6E-02	4.1E-02		
INT	0.0	0.42	1.3		-0.7				2.3					-0.3		0.7					5.7E-03
PLT	0.2	0.64					43.1			0.5	-11.8				-0.8		-0.5				6.5E-02
PAC	-2.8					1.8		-13.6	6.0		20.4						0.2		7.1E-03		3.2E-02
RKY	-9.8	-0.09						-7.5	0.9		42.0										2.8E-01

Table 4.3 *- The regression coefficients for selected models. The abbreviations in the first column are for the physiographic divisions: APP - Appalachian Highlands, ATP - Atlantic Plains, INT - Interior Plains, PLT - Interior Plateaus, PAC - Pacific Mountain System, RKY - Rocky Mountain System*

sedimentation rates for 326 reservoirs). So this model requires more variables to create a model based on fewer cases and still has a lower R^2 and higher standard error. Based on these numbers, the model developed for Set 2a is more likely to be reliable.

Looking only at the variables included, Model Set 2a and 3a are quite similar. The only major differences in the structure of the models after averaging is the inclusion of log of relief-ratio, mean rain, and minimum elevation in place of percent depositional zones in the unaveraged model.

While the predictive capabilities of the models presented above for the entire US may be questionable for small numbers of reservoirs, they do illustrate the influence of land-cover variables and suggest the need for their inclusion in similar analyses. Land-cover variables behaved in a manner consistent with previous investigations (e.g. forests tend to reduce, whereas agriculture tends to increase unit sedimentation rates). The models also illustrate the relative influence of climate, soil type, and topography. The results suggest that the presence of forests in a watershed has more influence on reservoir unit sedimentation rates than does soil type, rainfall intensity, or relief. It should be stressed, however, that few of the variables considered in this investigation can be considered to be truly independent, and the presence of forests may say as much about the topography, soil type, and rainfall of a watershed as it does about the stabilizing effects of the forest surface and subsurface environment. As will be discussed below, the relative influence of each variable is found to change from region to region within the US, suggesting the importance of local conditions over more general, continental patterns.

Use this in Summary

True to all physical process.

4.3.2 Appalachian Highlands

For the Appalachian Highlands division, the most reliable models are those developed with data sets 2 and 3, which were identical (there were no proximally averaged watersheds in this region). The R^2 for this model is 0.641 and the standard error is 0.728 with 48 degrees of freedom (6 variables are used to predict unit sedimentation rates for 56 reservoirs). Elevation and the capacity-watershed ratio describe most of the observed variation in unit sedimentation rates. The variables included in this model and their coefficients (see Table 4.3 and Appendix E, Figure E.10) are largely expected, though it is surprising that percent shrubland is included rather than a more dominant land-cover for this area.

Because shrubland is so rare in this area, this relationship is built from a small number of watersheds and is probably suspect. Because the majority of watersheds in the Appalachian Highlands do not contain any shrubland, it would be best to include a different variable. The land-cover variables most highly correlated with unit sedimentation rate, forest and urban-land, also happen to be correlated with other variables already included in the regression, which explains their ultimate absence from the model. Forest is positively correlated with mean rainfall ($R = 0.367$) and logarithmic relief-ratio ($R = 0.654$) and negatively correlated with the flatness coefficient ($R = -0.560$). Urban-land is negatively correlated with logarithmic relief-ratio ($R = -0.441$) and positively with the flatness coefficient ($R = 0.346$). Thus, in the Appalachian Highlands, forests tend to dominate the steep, mountainous watersheds while urban-land is found most often in flatter areas. Land-cover effects consistently show these interactions with topographic and climate parameters.

Consequently, isolating the effects of land-cover from those of topography and climate and adequately characterizing those effects for model construction has been difficult using the step-wise multiple regression techniques employed in this study.

4.3.3 Atlantic Plain

Models 2 and 3 for this division are the same because there were no proximally averaged watershed-reservoir systems. The model developed for the Atlantic Plain physiographic province is unlikely to be broadly applicable. Only 9 of the delineated watersheds had reservoir sedimentation histories of greater than 10 years. Consideration of those with shorter records (total n = 17) produces a model that is dominated by an elevation effect. While the effects of elevation in the Atlantic Plain division certainly appears to be consistent for the reservoirs considered and the R^2 and standard error suggest that the model performed admirably, the geographic distribution of the Atlantic Plains reservoirs is the least matched of any of the physiographic divisions to the distribution shown in the National Inventory of Dams (see Figure 4.2). If more reservoirs from this division were characterized and included in future regressions, elevation would most likely lose its predominance as the parameter explaining most of the variation. The elevation effect is discussed in more detail in Section 4.4.5.

4.3.4 Interior Plains

The models for this division are based on the largest number of reservoirs. The calculated R^2 (0.619) is the lowest of any for Model Set 2. This model also contains the greatest number of parameters (n = 7) of any model for a physiographic division. Depending on whether one examines the averaged (Set 3) or unaveraged model (Set 2), the dominant variables influencing unit sedimentation rates are the capacity-watershed ratio, agriculture, and mean basin slope or rainfall intensity (R-factor). It is interesting to note that log of basin relief and relief-ratio are negative coefficients in this model. Although counterintuitive, this may be the result of the interaction between land-cover and topography. Relief is positively correlated with total drainage area, and though drainage area is not included in the model, relief rather than drainage area may reflect the influence of reduced sediment output from larger drainage basins.

In this division, forests have a negative influence on unit sedimentation rates. The actual correlation coefficient between forest and unit sedimentation rate is positive, but this is due to the positive correlation between forest and both the logarithmic capacity-watershed ratio and mean slope and a negative correlation with maximum elevation (a result of the positive correlation between longitude and forests and the negative relationship between elevation and longitude). Forestland has a negative coefficient in the model once unit sedimentation rates are corrected for capacity-watershed, maximum elevation, and mean slope. Agriculture has a positive coefficient in the model, implying that agricultural watersheds tend to have greater unit sedimentation rates, but it is negatively correlated with relief and relief-ratio.

For this region, averaging proximal watersheds did not seem to improve the model. After averaging, the R^2 dropped from 0.619 to 0.579 and the standard error increased from 0.628 to 0.666. These changes occur despite a reduction in the number of cases from 151 to 132. The model coefficients are quite similar, although the negative coefficient for relief-ratio is replaced by a positive term using soil bulk density.

It is interesting that the model developed for this physiographic division has such a low standard error compared with the other regions, but this is partially a result of this division having the lowest coefficient of variation of unit sedimentation rates (0.168). This may also be due to the large number of reservoirs for this region. While it has been typically found with this dataset that models become less reliable as reservoirs are added (this is discussed more later) it may be that once a critical number is reached, the reliability of the model can be improved with greater numbers of cases provided that the region being grouped is fairly uniform in properties. It is suspected that this number has not generally been reached for most physiographic divisions or for the entire US in the present study. The fact that averaging, which reduced the number of cases considered, actually increased the standard error also hints at this possibility.

4.3.5 Intermontane Plateaus

This division shows the highest standard error of all regions for Model Set 2 (SE = 0.931), though the R^2 (0.671) is better than that for the Appalachian Highlands or the Interior Plains. The use of averaging for this division nearly halves the number of cases considered from 42 to 26 and increases the number of required variables from 6 to 7, though the model R^2 (0.847) and standard error (0.745) are both substantially improved. The averaged model suggests that depositional zones actually have a positive influence on unit sedimentation rates, but since depositional zones are negatively correlated with unit sedimentation rates, this is most likely the result of other variable interactions with depositional zones. However, examination of the correlation matrix did not reveal that any one interaction explained this behavior.

Both regression models (averaged and unaveraged) illustrate the positive influence that capacity-watershed ratio, soil organic matter, R-factor, and urban-land have on the unit sedimentation rates in this division. Both models also include total drainage, which shows a negative influence on unit sedimentation rates.

4.3.6 Pacific Mountain System

The unaveraged model for this division (Model Set 2) uses 50 reservoirs. Although this is a similar number to that used for the Appalachian Highlands division, this model requires one fewer parameter ($n = 7$) and has a much higher R^2 (0.890) and a slightly lower standard error (0.745).

While Model Set 2 for the entire country shows the expected positive influence of K-factor on unit sedimentation rates, both the averaged and unaveraged model for this division show that the influence is negative, despite the fact the correlation between unit sedimentation rate and K-factor is positive. But, because K-factor is positively correlated with the other three variables that are included in this model before its entry, and because it is negatively correlated with slope and relief (which are positively correlated with unit sedimentation rates), K-factor may be acting as a proxy for these two topographic variables. Certainly, this relationship is not unexpected and the negative relationship between slope and soil erodibility is seen in other regions as well. This again points to the difficulty in using multiple regression analysis when many of the variables are not truly independent. However, it would be difficult to adequately describe a watershed-reservoir system for model construction without considering the individual influence of these covarying parameters. This suggests that model development methods other than step-wise multiple regression may prove more useful if they can adequately account for the varying influence of soil type, land-cover, and topography despite the possible covariance among these parameters.

After averaging was employed (Model Set 3) this division is the only model for a single physiographic division or for the entire US that does not include the capacity-watershed ratio as an input variable. However, it is probable that this is due to the presumably coincidental but nonetheless strong correlation between soil bulk density and capacity-watershed ratio. Also, capacity-watershed ratio is positively correlated with total drainage area, which is included in the model. The averaging technique does not seem to produce a better model for this region, as the R^2 and standard error are both worse than for the unaveraged model and the number of cases considered (and thus the degrees of freedom) is lower. For this division, the model developed for Model Set 2 is probably the most reliable.

4.3.7 Rocky Mountain System

Aside from the Atlantic Plains division, this division has the fewest number of cases used for model development. Nonetheless, for Model Set 2 the R^2 (0.989) and standard error (0.290) for this region suggest that the model may still be reliable for most reservoirs in this region (if the reservoirs used for model development are representative of the population in the division). Using only 4 variables (R-factor, capacity-watershed ratio, total drainage area, and percent barren land), most of the variation of unit sedimentation rates is explained.

After averaging (Model Set 3), the R^2 (1.00) and standard error (0.005) for this division were the best for any model presented in this work, though the model was developed on only 9 reservoirs. Strangely, this model was the only example that shows a negative relationship between capacity-watershed ratio and the unit sedimentation rate. Both unit sedimentation rate and capacity-watershed ratio are strongly correlated with minimum watershed elevation, suggesting that minimum elevation, included as the first variable in the averaged model, may have been surrogated for capacity-watershed ratio.

Although a larger dataset would be useful to make any definite statements about reservoir sedimentation in this area, there seems to be a strong effect due to elevation for the watersheds studied. This could be tied to several things, including the influence of orographic effects on rainfall totals and intensity as well as (and perhaps as a result) land-cover.

4.3.8 Final Comments on Multiple Regression Analysis

While the relationships of several previously investigated parameters (e.g. R-factor, total drainage area, percent forest cover) show relationships to unit sedimentation rates that are consistent with previous research, the behavior of many parameters is clouded by significant, but geographically varying, interactions. This suggests a need for a model development technique that is not as sensitive to these interactions as is step-wise multiple regression. Nevertheless, this technique does demonstrate the need to somehow characterize land-cover when attempting to predict unit sedimentation rates, as every model presented (excepting those for the 7 Atlantic Plain reservoirs) includes a land-cover variable. It also illustrates the overwhelming effect of capacity-watershed ratio (a proxy for trap efficiency) on unit sedimentation rates.

To visually compare the predictive capabilities of the model developed for the entire U.S. to those developed for each physiographic region, the recorded logarithmic unit sedimentation rates are plotted against the unstandardized predicted value (from Model Set 2) for each reservoir in Figures 4.3 and 4.4. The 95% confidence intervals are also shown. The deviations of the predicted values are much smaller for the individual models developed for the physiographic regions than for the single model for the entire U.S., especially for those reservoirs with lower unit sedimentation rates. This illustrates the influence of *spatial autocorrelation* on the multiple regression technique employed here. Spatial autocorrelation is a tenet of Tobler's First Law of Geography (Tobler 1979), which states: "Everything is related to everything else, but near things are more related than distant things." For example,

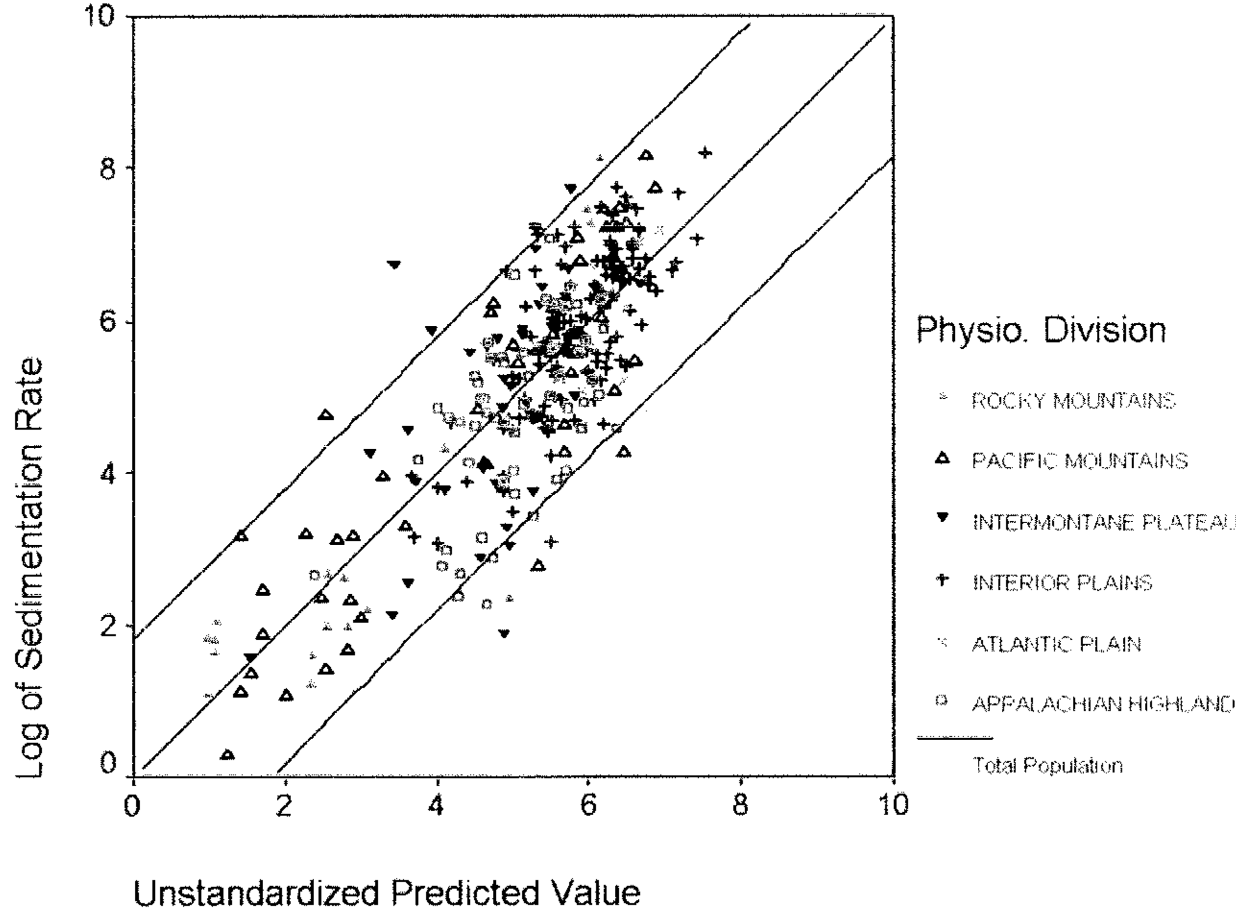


Figure 4.3 *– Actual vs. Predicted log sedimentation rate as predicted by Model Set 2a and the 95% confidence interval of the prediction.*

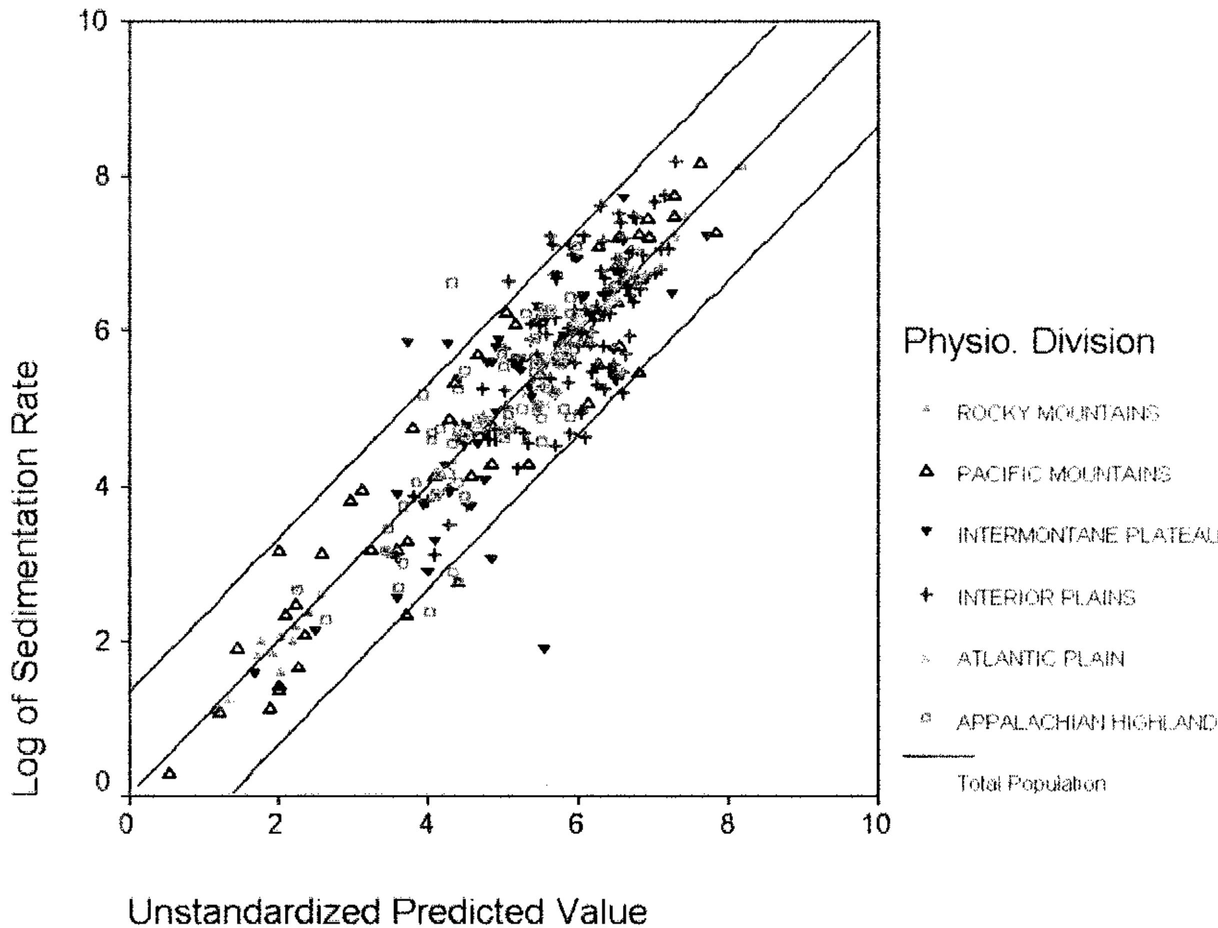


***Figure 4.4** – Actual vs. Predicted log sedimentation rate as predicted by the physiographic division models from Model Set 2 and the 95% confidence interval of the prediction.*

the standard deviation of rainfall rates compiled for most individual counties in the U.S. is smaller than the standard deviation of rainfall considered at the state level, and significantly smaller than that of the entire continental U.S. As the geographic range of the watersheds considered for model development is narrowed, the predictive capability of the model increases because the variation of the parameters that affect the unit sedimentation rates, but are not included in the model, is reduced.

4.4 Discussion of Individual Parameters

4.4.1 Slopes

Watershed slope undoubtedly has an effect on the amount of sediment delivered to a reservoir. Previous reservoir sedimentation studies have typically characterized slopes using the mean slope for the watershed, though other methods have been employed (see Flaxman, 1966). A qualitative comparison of two similar watersheds by Verstraeten and Poesen (2001c) suggested the importance of the distance between steep slopes and the catchment outlet on unit sedimentation rates.

A variety of methods for characterizing watershed slopes were employed to evaluate the relative efficacy of each characterization for predicting reservoir sedimentation. In addition to the rather common values of mean, minimum, and maximum slope, the mean and maximum slope were recorded for each "flowlength range" of the watershed. The watershed was divided into four flowlength ranges based on the *flowlength* command. This command uses the flowdirection grid (described earlier) to trace the length of the flow path from each gridcell to either the divide or the basin outlet. In this case, the distance from the outlet was used to separate each watershed into four zones, equally spaced from outlet to the maximum distance to the divide.

In almost all cases, the slopes contained in the 50-75% range of the watershed were most similar to the mean slope of the watershed. While the correlation between the various slope measurements and sediment accumulation was different for each physiographic division, in every case, the mean slope was one of the most highly correlated with unit sedimentation rate and any improvement in the correlation between sediment accumulation and a slope variable besides mean slope was not significant enough to merit the extra complexity involved in its calculation.

4.4.2 Depositional Zone

While it was found that the percent of depositional zones does correlate negatively (as expected) with reservoir unit sedimentation rates in certain physiographic regions (Atlantic Plain, Interior Plains, Intermontane Plateaus), the relationship is significant at the 0.05 level only for the Intermontane Plateaus region. Unexpectedly, percent depositional zone actually shows a significant positive correlation with unit sedimentation rate in the Appalachian Highlands. However, this may be due to the weak correlation between depositional zones and minimum elevation in this region. Because minimum elevation is so highly negatively correlated with unit sedimentation rate due to other factors (discussed later), the effects of depositional zones are overwhelmed. When included as an independent variable in the stepwise regression model, it was included as the tenth variable when all reservoirs were considered together (without averaging), but made only a modest improvement (see Appendix E Table E.3) to the regression results. When regression was performed by physiographic region, it was only included in the model for the Pacific Mountains as the last variable included, and resulted in only a moderate improvement in the regression (Appendix E Table E.4).

4.4.3 Flatness Coefficient

The flatness coefficient does not correlate with sediment output as expected across all physiographic regions. While there are significant (at the 0.05 level) correlations between sediment accumulation and the flatness for all regions but the Atlantic Plains, the relationship is strongly positive for both the Appalachian Highlands and the Rocky Mountains. For both of these regions, flatness has a strong negative correlation with the occurrence of forestland. Additionally, agriculture and urban-land both show strong positive correlations with flatness in these regions – illustrating yet another difficulty in isolating anthropogenic influences on sediment yields without careful consideration of all watershed characteristics.

In all areas except the Pacific Mountains (and the Rocky Mountains for percent depositional zone), basin area shows a strong positive correlation with flatness and weak positive correlation with percent depositional zone, illustrating the generally held notion that increased in-basin storage of sediment with basin size is due at least in part to increasing opportunities for storage within low slope areas of the watershed. This relationship is not as easy to discern from the correlation between mean slope and watershed area, which do not show the consistent trends evident in the flatness coefficient. Thus, these variables do provide additional information about watershed topography that is not enumerated elsewhere.

4.4.4 Relief Ratio

Figure 4.5 shows a graph taken from Schumm (1956) illustrating a positive relationship between basin relief ratio and estimated sediment loss as determined by reservoir accumulation rates. Early research on basin sediment yield often reports a positive correlation between basin relief or relief ratio (total relief / stream length) and reservoir unit sedimentation rate (Schumm, 1956; Roehl, 1962; Maner, 1958). In subsequent studies it has been noted by more than one author (Renwick, 1996; Verstraeten and Poesen, 2001c) that this relationship does not always hold or is sometimes negative. This analysis has also shown that this is not a simple relationship and seems to invert (or, to be masked) in areas where man has altered the landscape, especially with agriculture. Indeed, Figure 4.6 (next page) shows that this watershed characteristic is often not dominant in governing sediment production and its relationship with unit

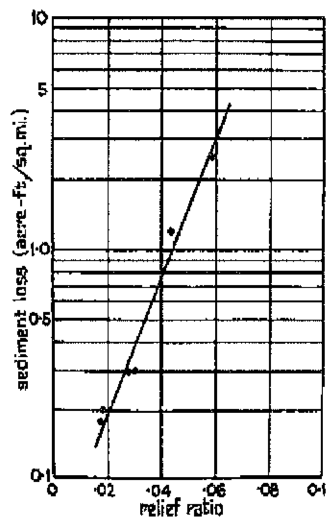


Figure 4.5 *– (From Schumm, 1956) A strong relationship between reservoir sed. rates (as basin sediment loss) and relief ratio as implied by a limited number of reservoir sed. rates. All reservoirs were from the Intermontane Plateaus region (compare to the results for Intermontane Plateaus reservoirs in this study shown in Figure 4.4)*

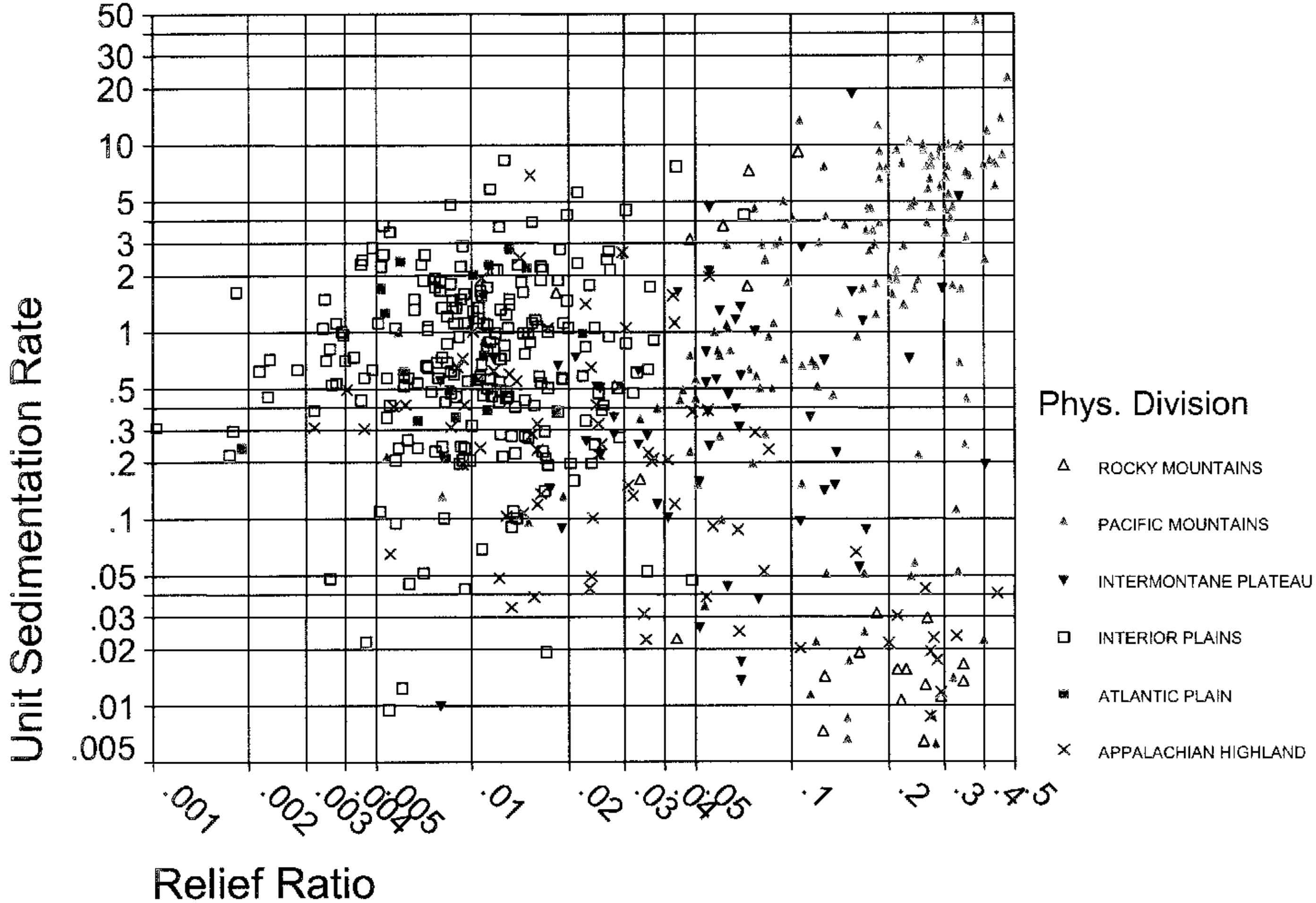


Figure 4.6 *– The relationship between relief ratio and sedimentation rate per unit area (m^3 km^{-2} yr^{-1}) for the reservoirs in this study.*

sedimentation rates seems to vary from region to region, especially when comparing plains-dominated to mountain-dominated divisions.

4.4.5 Elevation

This section briefly outlines the observed influence of watershed elevation seen in the regression analysis. Watershed elevation has no known direct effect on basin sediment yield. Nevertheless, several regression models find this variable to be the most influential in determining unit sedimentation rates. Analysis of correlation matrices and rotated factor matrices indicate that this influence is due mainly to the covariance of elevation with several erosionally important variables such as capacity-watershed ratio, hypsometric integral, percent grass, and relief.

In the Atlantic Plain region, both minimum and maximum watershed elevation show strong positive correlations with unit sedimentation rate and are entered as the exclusive variables (with capacity-watershed ratio) into the regression model for the 7 reservoirs with records longer than 10 years. When elevation variables are excluded and all reservoirs in this region were considered (n = 15. Analysis not shown.), maximum elevation is replaced by hypsometric integral and percent forest cover. For the reduced Atlantic Plain set (n = 7), when the elevation variables are excluded, the only variable included with the probability of inclusion set at 0.10 and exclusion at 0.20 is percent grass. However, when the entire set of Atlantic Plains reservoirs are considered at the same probability levels and with elevation variables excluded, the model is built on the capacity-watershed ratio, percent grass, hypsometric integral, and percent urban (which shows an unexpected negative influence on unit sedimentation rate in this region). While a larger dataset may not show the strong influence of elevation as it is shown here, it is suspected that at least a modest elevation influence will persist due to the general correlation between bedrock geology and elevation in this region. Most of the rocks at the surface in this region are of Cretaceous age, or younger, and the exposure of progressively older rocks is closely coupled with an increase in elevation. This relationship is hypothesized to influence soil characteristics, watershed topography, and (as a consequence) land cover. A more detailed analysis is required, however, before any definite conclusions can be drawn about the influence of elevation in this region.

The Appalachian Highlands area also exhibits a strong correlation of unit sedimentation rate with the minimum elevation variable, which is the first variable included in the step-wise regression for all models of this region. However, the correlation is negative in this region, meaning higher elevation reservoirs tend to have lower unit sedimentation rates. The correlation matrix for this area (see Appendix E, Table E.7) illustrates that minimum elevation is combining several variables that are correlated with unit sedimentation rate, including R-factor, relief ratio, and the capacity-watershed ratio. Removing elevation variables from the regression leads to the elimination of mean rainfall and percent shrubland as a determining variable and the inclusion of R-factor and soil bulk density in their place, but also leads to a lower R^2 and higher standard error of the estimate.

The Interior Plains is the only other physiographic division that includes an elevation variable, in this case maximum elevation, into the step-wise regression models. As in the Appalachian Highlands, reservoirs at lower elevations tend to show higher unit sedimentation rates. Maximum elevation was included as the second variable following the capacity-watershed ratio (see Appendix E, Table E.4). The correlation matrix reveals that, again, elevation is covariant with several important basin parameters in a similar manner to unit sedimentation rate, leading elevation to act as a sort of "super variable" summarizing a variety of variability with a single value. In this case, the variables most likely leading to this relationship are log of hypsometric integral, mean rainfall, and R-factor. It is interesting to note that the step-wise procedure, which produced nine separate models for the Interior Plains section, adds maximum elevation to the model during the second step and then removes it on the eighth step after the inclusion of percent agriculture, mean slope, percent forest, log of hypsometric integral, log of relief, and R-factor during the intervening steps. This suggests that the elevation value was serving as a proxy for many of these variables, and once they were added to the model, it offered little additional explanation of the variability of unit sedimentation rate itself.

4.4.6 Capacity-Watershed ratio

In all of regressions presented below for the entire US and for most of the regressions for individual physiographic divisions, the logarithmic capacity-watershed ratio (L_C/W) is shown to be the parameter most highly correlated with unit sedimentation rate and is typically the first variable entered into each step-wise regression model. The only exceptions to this are for the Atlantic Plain, where the number of reservoirs is inadequate for model development, and the Appalachian and

Rocky Mountains where the effects of elevation and rainfall intensity overwhelm the influence of the capacity-watershed ratio (though the capacity-watershed ratio is included in the models for all three of these regions as the second or third variable entered). This influence is due to the relationship between the capacity-watershed ratio and reservoir trap efficiency, as discussed in Section 2.2.3.

4.5 Discussion of Possible Sources of Unexplained Variation

"Just as every natural lake is an individual phenomenon, which does not appear again in an identical form on the earth's surface, every man-made lake is likewise a unique occurrence."

-Fels and Keller (1973)

Unit sedimentation rates (in $m^3/km^2/yr$) for the reservoirs in this study vary over 5 orders of magnitude. The models presented above can all predict unit sedimentation rates on average to within approximately one order of magnitude unless otherwise noted, though often not much better than this. The observed variations from predicted unit sedimentation rates have a variety of possible sources. A source of much of this unexplained variability is suspected to be the lack of information concerning reservoir operational practices. Without knowledge of typical reservoir storage capacity levels, trap efficiency cannot be accurately accounted for with the capacity-watershed ratio. Also, density current venting, bottom sediment dredging, and additional sediment-control measures are not explicitly considered by this model but can have very dramatic consequences on delivery and retention of sediment to reservoirs. Zhide and Xiaoqing (1997) report a reduction of reservoir trap efficiency for a Chinese reservoir from 92.9% to nearly 0% given the practice of actively venting muddy water and retaining clear water. While this practice is not likely to be widespread for the smaller reservoirs included in this study, it illustrates the important potential effects of reservoir management on unit sedimentation rates and the need to include at least qualitative information about reservoir management practices in future analyses.

Other sources of variability are the effects of changing land-use and land-cover. Personal communication with staff at the National Resource Conservation Service in Mills County, Iowa indicated that a rapid increase in sedimentation rates was observed in the Mule Creek drainage reservoir system (3 Mule Creek reservoirs are included in this study) following an "okay" from the Secretary of Agriculture in 1973 for farmers to sell crops overseas and to farm from "fence row to fence row." This led to a switch for most farms in the Mills County area from a corn/oats/meadow/meadow annual crop rotation to a corn/bean rotation. The reservoirs had filled to nearly half capacity with sediment between their construction ca. 1953 and about 1973. Between the time of this policy change in 1973 to the mid 1980's the rest of the Mule Creek reservoirs' capacities were almost completely lost to sedimentation, despite the fact that reservoir sedimentation rates typically decline with age Lajczak (1996). This represents a nearly two-fold increase in sedimentation rates due to a change in crop rotations.

While investigating the role of weather patterns on reservoir sedimentation rates, Wilby, et al (1997) discovered that, though a period of strip mining occurred during the record of one of the reservoirs, the installation of a settling pond actually led to a reduction in unit sedimentation rate from the basin. Thus, not only is land-cover history significant in determining reservoir infilling rates, but details such as crop type and erosion-prevention practices, which are not quantified by the NLCD (or any other current large-scale database), can also play an important role in determining these rates, especially for small watersheds.

Another role that land-use can play in sediment delivery to reservoirs has been described by several authors. "Poaching" by cattle, whereby stream or reservoir banks are trampled by livestock, usually leads to bank failure and can in some cases account for a significant portion of the sediment delivery to the system where stocking densities are high and cattle are allowed unrestricted access to banks. Poaching by cattle delivers sediment directly from the pasture to the stream and breaks down the soil structure, making it more erodible (Wilby, et. al. 1997). Furthermore, "the significant hydrological effect of cattle trampling on both infiltration capacities and storm hydrograph response may have a direct impact on soil erosion from the fields and on channel erosion consequent upon increasing flow magnitudes" (Foster, 1995). Lloyd, et al (1998) have shown that shoreline erosion due to livestock access to small farm ponds can account for a significant portion of the sediment delivered to ponds located near hillcrests (up to 85%). They also point out that cattle tend to produce higher rates of bank erosion than other livestock due to their weight and tendency to return to the same place again and again. Foster and Walling (1994) have also found a correlation between increasing sedimentation rates in one English reservoir with increased livestock numbers in the watershed. Finally, Heathwaite, et al. (1990) report an increase in sediment production from heavily grazed fields of 80 times that of a similar field without grazing and note the increased effects if cattle are given unlimited access to stream banks.

Although these sources of variability may have adversely affected the results of the models developed for this study, the information provided here should help explain possible sources of error for those who may use these models in a predictive capacity. Also, this information should be noted when future researchers use this GIS method for creating new models with improved datasets.

Chapter 5 – Discussion and Conclusions

The preceding chapters describe a method for reservoir geolocation and watershed characterization using GIS as well as a series of statistical results and models for predicting reservoir sedimentation.

It can be concluded from visual analysis of the delineated watersheds and the results of the statistical analysis of the data that the method used for geolocation and characterization performs this task reliably for most regions. However, the accuracy of this method is reduced with increasing terrain complexity and decreasing basin size. This may be important in future analyses utilizing the RESIS database, which contains a large number of small debris basins with high sedimentation rates in complex terrains, as errors in these regions may significantly effect model results. However, for databases such as the NID and NASQUAN (the National Stream Water Quality Accounting Network), where most basins have larger drainage areas (many of the RESIS reservoirs with exceptionally small watersheds are not present in the NID) and the density of watershed spillpoints in topographically complex areas is not disproportionately high, this method should prove to be reliable.

The methods presented in this paper allow an unprecedented level of information about a large number of watersheds to be collected compared to previous studies. Although the current methodology can be improved, the foundations presented here should allow future researchers to describe large numbers of watersheds with reasonable accuracy for large-scale studies. This method can be used with reasonable success for the majority of the United States. It should be particularly useful in the delineation of and characterization of the NID dams and NASQUAN water quality gauging sites. This method could be greatly improved with the use of a larger, seamless DEM like the National Elevation Dataset (NED). While similar extrapolation to the world should be approached with caution, this is due mainly to the great variations of GIS data quality from place to place. Topography generated by missions such as the National Shuttle Radar Topography Mission (SRTM) will soon provide global datasets with accuracy similar to, if not better than, the data used in this analysis. Future missions may produce similar datasets describing land-cover, climate, and soil. That said, with data of similar quality to that used in this paper it should be possible to perform similar analyses (i.e. with large numbers of reservoirs, gauging stations, or field chemistry and sedimentology sites where general watershed or positional characteristics are known, but precise location is not registered to a DEM) for much of the world. As the results have shown, some quantification of geographic and topographic complexity is recommended as this provides a general idea of the probability of and types of errors that may be encountered.

While the resolution of most of the databases was thought to be sufficient for watershed characterization, 30-meter DEMs may not adequately characterize complex topographic characteristics (such as depositional zones or hypsometric integral) for smaller watersheds (less than ~ 0.5 km^2). However, until more accurate datasets are made publicly available, large-scale studies such as this one must rely on the data with the greatest global coverage, despite the possibility of errors.

The models developed for this thesis are of varying quality, but the models for the Interior Plains, Appalachian Highlands, and Pacific Mountains physiographic regions can give reasonably accurate unit sedimentation rate estimates for small reservoirs located in the regions for which they were developed. Also, the model predicting unit sedimentation rates for small reservoirs for the entire United States (Model Set 2a) could be applied to attain order-of-magnitude estimates of unit sedimentation rates for large-scale studies.

Above all, this study illustrates that caution must be exercised when using descriptive watershed parameters, especially land-cover parameters, to predict reservoir unit sedimentation rates over broad regions. These parameters do not show consistent influences from region to region. The most likely explanation for the variable influence of land-cover parameters on unit sedimentation rates is the inconsistent manner in which humans have carried out land-cover change. The effects of changing land-cover from one type to another may be offset by erosion control practices if they are properly employed. Thus, a simple characterization of modern land-cover is probably not sufficient for predicting accurate unit sedimentation rates. Until a more comprehensive method can be developed that takes into account both land-cover change and the estimated extent of erosion control practices (a system such as that developed for RUSLE) as well as reservoir maintenance practices, multiple regression methods will only provide approximate estimates of reservoir unit sedimentation rates.

One important observation in this analysis that has been mentioned in previous literature, though never as conclusively shown, is that as the geographic scope of the regression analysis is reduced (along with the number of reservoirs considered), the reliability of the models is typically improved. It is perhaps counterintuitive to think of a model being improved by reducing the number of cases considered. However, this trend is fairly evident from the data presented above and is mainly the result of two factors.

First, reservoir systems are quite complex and the rates of sedimentation in two similar and adjacent systems can vary quite markedly if only one or two factors such as reservoir operation or land-use are different. These two variables have been shown to consistently thwart previous attempts to create reliable, large-scale reservoir sedimentation models. Thus, the construction of a model that does not specifically address these factors is going to be subject to greater errors as greater numbers of reservoirs are considered.

Second, *spatial autocorrelation*, which refers to the tendency for most spatially varying parameters to show greater variation as the scale observed is broadened, means that models developed for smaller regions are more accurate. Because these regression models use a limited number of parameters (typically from 3 to 10) to predict unit sedimentation rates, those parameters not considered will add to the observed error. As one limits the geographic range of the analysis, the variation of parameters not included in the regression is reduced, improving the reliability of the model. As the geographic scope of an analysis is broadened, the number of parameters that must be considered must also increase, adding to the complexity of the model. While all of the reservoir and watershed parameters considered in this work and in the other analyses discussed in the introduction are contributing in some way to the rates of sedimentation for every reservoir, the relative contribution of each parameter is a function of how variable that parameter is on the geographic scale of the analysis. Broader analyses will require more variables. Narrower analyses will require fewer.

In this light, it seems unlikely that any reliable model for reservoir sedimentation over broad geographic areas will be developed using multiple regression analysis unless that analysis includes an immense number of variables, including those describing reservoir operation and the temporal variation of land-use and land-cover. While physical models likely hold more promise for broad applicability across geographic regions, the required input data for these models will prevent their use for large-scale studies until more detailed data become available describing the reservoirs themselves and the land surface and its history. Until then, focus should be put on broadening our knowledge of reservoir operational histories. Specifically it would be advantageous for future studies of this type to collect and incorporate greater details of typical reservoir storage capacities, flushing operations, and drawdown severity and frequency.

Until greater details about land-cover and land-cover change as well as reservoir type and operation are known for the reservoirs being modeled, the methods outlined in this work should aid in the development of regression models for small geographic areas. The ideal geographic range for such regression analysis is variable but is shown to be a function of the terrain complexity of the area being modeled. The key to this type of model development to insure that any such model includes descriptions of any hydrologically important parameter that is suspected to significantly vary over the range of the analysis.

Bibliography

Ackerman, C. W. and R. L. Corinth, 1962. An empirical equation for reservoir sedimentation. IASH, Pub. 59, p. 359 – 366.

Anderson, Henry W., 1949. Flood frequencies and sedimentation from forest watersheds. Transactions American Geophysical Union, Volume 30, p. 567 – 583.

Bahr, David B. and James P. Syvitski, Rivers versus creeks: The structure of noise in power-law relationships characterizing fluvial systems. *Unpublished.*

Bechteler, W. and M. Nujic, 1998. Predicting reservoir sedimentation with 2D model FLOODSIM, International Journal of Sediment Research, Vol. 13, No. 1, p. 68 – 76.

Brown, C.B., 1943. Discussion of *Sedimentation in reservoirs*, J. Witzig. *Proceedings of the American Society of Civil Engineers*, Vol. 69, p. 1493-1500

Brune, Gunnar M., 1953. Trap efficiency of reservoirs, Transactions American Geophysical Union, Vol. 34, p. 407 – 418.

Churchill, M. A., 1948. Discussion of *Analyses and use of reservoir sedimentation data* by L.C. Gottschalk. *In: Proceedings of the federal interagency sedimentation conference, Denver, Colorado*, Washington, DC: US Geological Survey, 139-140.

Crowder, B. M., 1987. The economic cost of reservoir sedimentation: a regional approach to estimating cropland erosion damages. Journal of Soil and Water Conservation, Vol. 42, p. 194 – 197.

Dean, Walter E. and Eville Gorham, 1998. Magnitude and significance of carbon burial in lakes, reservoirs, and peatlands. Geology, Vol. 26(6), p. 535-538.

Dendy, F. E., W. A. Champion, and R. B. Wilson, 1973. Reservoir Sedimentation Surveys in the United States, Proceedings of the Symposium on Man Made Lakes, Knoxville, Tenn., May 3-7, 1971.

Dertine, K. K. and M. F. Mendeck, 1978. Industrialization of New Haven, Conn., as recorded in reservoir sediments.

Eakin, Henry M. and Carl B. Brown, 1939. Silting of Reservoirs, U.S. Department of Agriculture, Technical Bulletin 524, 186 p.

Einsele, G. and M. Hinderer, 1995. Terrestrial sediment yield and the lifetimes of reservoirs, lakes and larger basins, Geol Rundsch, Vol. 86, p. 288 – 310.

Farnham, C. W., C. E. Beer, and H. G. Heinemann, 1966. Evaluation of factors affecting reservoir sediment deposition, *In: Hydrology of Lakes and Reservoirs, Symposium of Garda (Oct. 9-15, 1966)*, IAHS Publication No. 71, p. 747-758.

Fels, E. and R. Keller, 1973. World Register on Man-Made Lakes, Geophysical Monograph, Vol. 17, p. 43 – 50.

Fenneman, N. M., and Johnson, D. W. , 1946. Physical division of the United States, U.S. Geological Survey map, scale 1:7,000,000.

Flaxman, Elliott M., 1966. Some variables which influence rates of reservoir sedimentation in western United States, *In: Hydrology of Lakes and Reservoirs, Symposium of Garda (Oct. 9-15, 1966)*, IAHS Pub. 71, p. 824 – 838.

Foster, I. D. L. and D. E. Walling, 1994. Using reservoir deposits to reconstruct changing sediment yields and sources in the catchment of the Old Mill Reservoir, South Devon, UK, over the past 50 years, Hydrological Sciences Journal, Vol. 39, No. 4, p. 347 – 368.

Foster, Ian D. L., 1995. Lake and reservoir bottom sediments as a source of soil erosion and sediment transport data in the UK. *In: Sediment and Water Quality in River Catchments,* eds. Foster, I.D.L., A.M. Gurnell, and G.W. Webb, p. 265-283.

Gill, M.A., 1979. Sedimentation and useful life of reservoirs. Journal of Hydrology, Vol. 44, p. 89 – 95.

Glymph, Louis M. Jr, 1954. Studies of sediment yields from watersheds. IASH, General Assembly of Rome, Pub. 36 (1), p. 178 – 191.

Glymph, Louis M., Jr., H. G. Heinemann, and Victor O. Kohler, 1951. Unpublished study from the files of the U.S. Soil Conservation Service, Lincoln, Nebraska.

Gottschalk, Louis C., 1946. Silting of stock ponds in land utilization project area, SD-LU-2, Pierre, South Dakota, U.S. Department of Agriculture, Special Report No. 9, Washington D.C.

Gottschalk, L. C. and G. M. Brune, 1950. Sediment design criteria for the Missouri Basin Loess Hills, U.S. Dept. of Agriculture, SCS-TP-97, Milwaukee, Wisconsin.

Heathwaite, A. L., T. P. Burt, and S. T. Trudgill, 1990. Land-use controls on sediment production in a lowland catchment, South-west England, *In:* Soil Erosion on Agricultural Land, eds. J. Boardman, I. D. L. Foster, and J. A. Dearing, p. 69 – 86.

Heinemann, H. G., 1984. Reservoir trap efficiency. *In:* Erosion and sediment yield: some methods of measurement and modeling, eds. R.F. Hadley and D.E. Walling, p. 201-218.

Kattelmann, Richard, 1997. Sedimentation in Sierra Nevada Reservoirs, *In:* Energy and Water: Sustainable Development: Proceedings of them D: the 27th Congress of Int'l Assoc. for Hydraulic Research: San Francisco, Cal., Aug 10-15, 1997, theme editors John S. Gulliver and Pierre-Louis Violett, p. 72 - 77.

Kinner, David A., 1999. Hydrology of tropical catchments with variable topography and land use: an application of TOPMODEL. M.S., University of Colorado.

Kohler, Victor O. Personal communication cited in Glymph, 1954 (above), Soil Conservation Service, Lincoln, Nebraska.

Lajczak, Adam, 1996. Modelling the long-term course of non-flushed reservoir sedimentation and estimating the life of dams. Earth Surface Processes and Landforms, Vol. 21, p. 1091-1107.

Land use and climate impacts on fluvial systems during the period of agriculture – Recommendations for a research project and its implementation, ed. Robert J. Wasson, PAGES Workshop Report, Series 96-2.

Lloyd, Sara D., Paul Bishop, and Ivars Reinfelds, 1998. Shoreline erosion: a cautionary note in using small farm dams to determine catchment erosion rates, Earth Surface Processes and Landforms, Vol. 23, p. 905 – 912.

Maner, Sam B. and Leland H. Barnes, 1953. Suggested criteria for estimating gross sheet erosion and sediment delivery tares for the Blackland Prairies problem area in soil conservation. U.S. Soil Conservation Service, Fort Worth, Texas.

Maner, Sam B., 1958. Factors affecting sediment delivery rates in the Red hills physiographic area. Transactions American Geophysical Union, Vol. 39, p. 669 –675.

McHenry, J. Roger, 1974, Reservoir sedimentation, Water Resources Bulletin, Vol. 10, no. 2, p. 329 – 337.

McLean, R. I., J. K. Summers, C. R. Olsen, S. L. Domotor, I. L. Larsen, H. Wilson, 1991, Sediment Accumulation rates in Conwingo Reservoir as determined by man-made and natural radionuclides, Estuaries, Vol. 14, No. 2, p. 148 – 156.

Meade, R.H., T.R. Yuzyk, and T.J. Day, 1990. Movement and storage of sediment in rivers of the United States and Canada, *In: Surface Water Hydrology*, eds. M.G. Wolman and H.C. Riggs. Geological Society of America, The Geology of North America, Vol. O-1, p. 255-280.

Meade, Robert H. (editor), 1995, Contaminants in the Mississippi River, 1987-92, U.S. Geological Survey Circular 1133, 140 p.

Mulholland, P. J. and J. W. Elwood, 1982. The role of lake and reservoir sediments as sinks in the perturbed global carbon cycle. Tellus, Vol. 34, p. 490-499.

Norusis, Marija J., 1999. SPSS 6.1 Guide to Data Analysis. Prentice-Hall, Englewood Cliffs, New Jersey, 582 p.

Paulet, Manuel, 1971. An interpretation of reservoir sedimentation as a function of watershed characteristics, Ph.D., Purdue University.

Rausch, D. L. and H. G. Heinemann, 1984. Measurement of reservoir sedimentation, *In:* Erosion and Sediment Yield: Some Methods of Measurement and Modelling, p. 179 – 200.

Renard, K. G., G. R. Foster, F. A. Weesies, D. K. McCool, and D. C. Yoder, coordinators, 1997. Predicting Soil Erosion by Water: A Guide to Conservation Planning With the Revised Universal Soil Loss Equation (RUSLE). U.S. Department of Agriculture, Agriculture Handbook No. 703, 404 p.

Renwick, William H., 1996. Continental-scale reservoir sedimentation patterns in the United States, Erosion and Sediment Yield: Global and Regional Perspectives (Proceedings of the Exeter Symposium, July 1996), IAHS Publ. no. 236, p. 513 – 522.

Ritchie, Jerry C., 1989. Carbon content of sediments of small reservoirs, Water Resources Bulletin, Vol. 25, No. 2, p. 301-308.

Roehl, John W., 1962. Sediment source areas, delivery ratios and influencing morphological factors, *In: Commission on Land Erosion,* IAHS Publication No. 59, p. 202 – 213.

Roehl, John W., 1966. Sediment design requirement for small reservoirs, *In: Hydrology of Lakes and Reservoirs, Symposium of Garda (Oct. 9-15, 1966),* IAHS Publication No. 71, p. 795 – 803.

Schumm, Stanley, 1956, The relation of drainage basin relief to sediment loss, *In*: Int'l Union Geodesy and Geophys., 10^{th} General Asscembly (Rome), Trans., vol. 1, p. 216-219.

Singh, Krishan P. and Ali Durgunoğlu, 1989, A new method for estimating future reservoir storage capacities, Water Resources Bulletin, vol 25(2), p. 263-274.

Smith, S. V., W. H. Renwick, R. W. Buddemeier, and C. J. Crossland, 2001. Budgets of soil erosion and deposition for sediments and sedimentary organic carbon across the conterminous United States, Global Biogeochemical Cycles, Vol. 15, No. 3, p. 697 – 707.

Spomer, R. G. and R.L. Mahurin, 1987. Landform changes determined photogrammetrically ondeep loess soils. Trans. ASAE 30(1), p. 153 – 157.

Spomer, R. G., J. R. McHenry, and R. F. Piest, 1985. Sediment movement and deposition using Cesium-137 tracer. Trans. ASAE 28, p. 767-772.

Stall, J. B. and L. J. Bartelli, 1959. Correlation of reservoir sedimentation and watershed factors. Illinois State Water Survey, Report of Investigation 37. Urbana, Illinois.

Stallard, Robert F., 1998. Terrestrial sedimentation and the carbon cycle: coupling weathering and erosion to carbon burial, Global Biogeochemical Cycles, Vol. 12, p. 231 – 257.

Trimble, 1999, Decreased rates of alluvial sediment storage in the Coon Creek Basin, Wisconsin, 1975 - 93, Science, Vol. 285, p. 1244 - 1246.

Tobler, W. 1979. Cellular Geography. *In: Philosophy in Geography,* eds. Gale, S. and Olsson, G., Dordrecht Reidel, p. 379-386.

Verstraeten, Gert and Jean Poesen, 2000, Estimating trap efficiency of small reservoirs and ponds: methods and implications for the assessment of sediment yield, Progress in Physical Geography, vol. 24 (2), p. 219-251.

Verstraeten, Gert and Jean Poesen, 2001(a), Variability of dry sediment bulk density between and within retention ponds and its impact on the calculation of sediment yields, Earth Surface Processes and Landforms, 26, p. 375-394.

Verstraeten, Gert and Jean Poesen, 2001(b). Modelling the long-term sediment trap efficiency of small ponds, Hydrological Processes, Vol. 15, p. 2797 – 2819.

Verstraeten, Gert and Jean Poesen, 2001(c), Factors controlling sediment yield from small intensively cultivated catchments in a temperate humid climate. Geomorphology, vol. 40, p. 123-144.

Vörösmarty, Charles J., Michel Meybeck, Balázs Fekete, Keshav Sharma, Pamela Green, and James Syvitski, *in press,* Anthropogenic sediment retention: major global-scale impact from the population of registered impoundments, Global and Planetary Change.

Wilby, R.L., H.Y. Dalgleish, and I.D.L. Foster, 1997. The impact of weather patterns on historic and contemporary catchment sediment yields. Earth Surface Processes and Landforms, Vol 22, p. 353-363.

Wolman, M. Gordon, 1989. Erosion control and reservoir deposition, *In:* Taming the Yellow River: Silt and Floods, eds. Brush, L. M. et al., p. 139 – 161.

Wolock, D. M., 1997. STATSGO soil characteristics for the conterminous United States, U.S. Geological Survey Open-File Report 97-656, 17 p.

Woodburn, R. 1955. Sediment production in small watersheds. Agricultural Engineering, Vol. 36, p. 467 – 470.

Zhide, Zhou, and Yang Xiaoqing, 1997. Reservoir sedimentation in China – Experiences and lessons, *In:* Energy and Water: Sustainable Development: Proceedings of them D: the 27th Congress of Int'l Assoc. for Hydraulic Research: San Francisco, Cal., Aug 10-15, 1997, theme editors John S. Gulliver and Pierre-Louis Violett, p. 72 - 77.