We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

1

Tools for Watershed Planning – Development of a Statewide Source Water Protection System (SWPS)

Michael P. Strager

Division of Resource Management, West Virginia University,
USA

1. Introduction

A Surface Water Protection System (SWPS) was developed to bring spatial data and surface water modeling to the desktop of West Virginia Bureau of Public Health (WVBPH), Office of Environmental Health Services (OEHS), Environmental Engineering Division (EED). The SWPS integrates spatial data and associated information with the overall goal of helping to protect public drinking water supply systems.

The SWPS is a specialized GIS project interface, incorporating relevant data layers with customized Geographic Information Systems (GIS) functions. Data layers have been assembled for the entire state of West Virginia. Capabilities of the system include map display and query, zone of critical concern delineation, stream flow modeling, coordinate conversion, water quality modeling, and susceptibility ranking. The system was designed to help meet the goals of the Surface Water Assessment and Protection (SWAP) Program.

The goal of the SWAP program is to assess, preserve, and protect West Virginia's source waters that supply water for the state's public drinking water supply systems. Additionally, the program seeks to provide for long term availability of abundant, safe water in sufficient quality for present and future citizens of West Virginia. The SWPS was designed to help meet this goal by addressing the three major components of the SWAP program: delineating the source water protection area for surface and groundwater intakes, cataloging all potential contamination sources, and determining the public drinking water supply system's susceptibility to contamination.

This chapter outlines the functions and capabilities of the SWPS and discusses how it addresses the needs of the SWAP program. The following sections discuss the application components. The components consist of:

1. A customized interface for study area selection
2. Integration of the EPA WHAEM and MODFLOW models
3. Delineation of groundwater public supply systems
4. Watershed delineation and zone of critical concern delineation for surface water sites
5. Stream flow model from multivariate regression
6. The environmental database

7. UTM latitude/longitude conversion utility
8. Statewide map/GIS data layers
9. Water quality modeling capability
10. Groundwater and surface water susceptibility model

Component 1. A customized interface for study area selection

Using customized programming we were able to create a GIS interface to allow users to quickly find locations or define study areas for further analysis in the state. The locations may be selected in three ways: by geographical extent (e.g. county, watershed, 1:24,000 quad map, major river basin), by area name or code (e.g. abandoned mine land problem area description number, stream or river name, WV Division of Natural Resource (WVDNR) stream code, public water identification number or name), or by typing in the latitude and longitude coordinates. Once the study area is defined, the system zooms automatically to the extent of the selected feature and all available spatial data layers are then displayed. A discussion of the spatial data layers included is discussed in Component 8 of this document.

Component 2. Integrating EPA WHAEM and MODFLOW models

The SWPS application has the ability to read output from either EPA WHAEM or MODFLOW models. It does this by importing dxf file formats directly into SWPS from a pull down menu choice. Data can also be converted to shapefile format from SWPS to be read directly into WHAEM and MODFLOW. The data being read into SWPS needs to be in the UTM zone 17, NAD27 projection (with map units meters) for the new data to overlay on the current data existing within SWPS. Consequently, any data exported from SWPS will automatically be in the UTM zone 17 NAD27 coordinate system.

Component 3. Delineation of groundwater public supply systems

A fixed radius buffer zone was created around each groundwater supply site based on the pumping rate. If the pumping rate was less than or equal to 2,500 gpd, a radius of 500 feet was used. If the pumping rate was greater than 2,500 gpd but less than or equal to 5,000 gpd, a radius of 750 feet was used. If the pumping rate was greater than 5,000 gpd and less than or equal to 10,000 gpd, a radius of 1,000 feet was used. If the pumping rate was greater than 10,000 gpd and less than or equal to 25,000 gpd, then a radius of 1,500 feet was used.

There were two exceptions to this fixed radius buffer procedure. The first was for any groundwater site less than or equal to 25,000 gpd that was in a Karst or mine area. These locations regardless of their pumping rate less than 25,000 gpd were buffered 2,000 feet. The second exception was for sites over 25,000 gpd. For these sites, hydro geologic and/or analytical mapping delineations will be done by personnel at the Bureau of Public Health. These were only identified in SWPS as being a well location and are left to more sophisticated groundwater modeling software.

To perform buffers automatically, the user can use the GIS to create buffers dialog within SWPS susceptibility ranking menu option. The automatic fixed radius buffering requires knowledge about the pumping rate and fixed radius distance. This information is provided in a pulldown text information box within the susceptibility ranking menu option.

Component 4. Watershed delineation and zone of critical concern delineation for surface water sites

The ability to interactively delineate watersheds and zones of critical concern is built into SWPS. In this section, the watershed delineation tool is discussed, followed by the zone of critical concern delineation tool.

Watershed Delineation

SWPS allows the user to delineate a watershed for any mapped stream location in the state. The watershed is delineated based on the user-clicked point and it is added to the current view's table of contents as a new theme or map layer labeled "Subwatershed." The drainage area is reported back to the user as well. If only drainage area is requested, a separate tool allows for quick query of stream drainage area in acres and square miles, without waiting for the watershed boundary to be calculated.

The watershed delineation is driven by a hydrologically correct digital elevation model (DEM). The DEM is corrected using stream centerlines for all 1:24,000 streams. The stream centerlines are converted to raster cells and DEM values are calculated for each cell. All off-stream DEM cells are raised by a value of 20 meters to assure the DEM stream locations are the lowest cells in the DEM. This step is necessary to assure of more accurate watershed delineations especially at the mouth of the watersheds. After the DEM is filled of all spurious sinks, flow direction and flow accumulation grids are calculated. These grids help determine the direction of flow and the accumulated area for each cell in the landscape. These grids were necessary for watershed delineation to occur and are important inputs for finding the zones of critical concern for surface water intakes.

Surface Water Zones of Critical Concern

Stream velocity is the driving factor for determining a five-hour upstream delineation for each surface water intake in WV. Only with stream velocity calculated was it possible to include factors such as high bank-full flow, average flow, stream slope, and drainage area all at once. The velocity equation used in this study came from a report titled "Prediction of Travel Time and Longitudinal Dispersion in Rivers and Streams" (US Geological Survey, Water-Resources Investigations Report 96-4013, 1996). In this report, data were analyzed for over 980 subreaches or about 90 different rivers in the United States representing a wide range of river sizes, slopes, and geomorphic types. The authors found that four variables were available in sufficient quantities for a regression analysis. The variables included the drainage area (D_a), the reach slope (S), the mean annual river discharge (Q_a), and the discharge at the section at time of the measurement (Q). The report defines peak velocity as:

$$V'_p = V_p D_a / Q$$

The dimensionless drainage area as:

$$D'_a = D_a^{1.25} * sqrt(g) / Q_a$$

Where g is the acceleration of gravity. The dimensionless relative discharge is defined as:

$$Q'a = Q / Q_a$$

The equations are homogeneous, so any consistent system of units can be used in the dimensionless groups. The regression equation that follows has a constant term that has specific units, meters per second. The most convenient set of units for use with the equation are: velocity in meters per second, discharge in cubic meters per second, drainage area in square meters, acceleration of gravity in m/s2, and slope in meters per meter.

The equation derived in the report and the equation used in this study for peak velocity in meters per second was the following:

$$Vp = 0.094 + 0.0143 * (D'_a)^{0.919} * (Q'_a)^{-0.469} * S^{0.159} * Q/D_a$$

The standard error estimates of the constant and slope are 0.026 m/s and 0.0003, respectively. This prediction equation had an R^2 of 0.70 and a RMS error of 0.157 m/s.

Once a velocity grid was calculated as described above, it was used as an inverse weight grid in the flowlength ArcGIS (ESRI, 2010) command. The flowlength command calculates a stream length in meters. If velocity is in meters per second, the inverse velocity as a weight grid will return seconds in our output grid. This calculation of seconds would track how long water takes to move from every cell in the state where a stream is located to where it leaves the state. The higher values will exist in the headwater sections of a watershed. By querying the grid, it is possible to add the appropriate travel time to the cell value and this will the time of travel for an intake. All cells above an intake by 18,000 seconds (5 hours) will be the locations in which water would take to reach the intake.

To use this methodology, GIS data layers had to be calculated for drainage area, stream slope, annual average flow, and bank-full flow for all of WV. The sections below describe how each of these grids was created.

Drainage area

To obtain a drainage area calculation for every stream cell in the state required a hydrogically correct DEM. The process of creating a hydrologically correct DEM was covered in the watershed delineation component described earlier. Essentially, from the DEM the flow direction and flow accumulation values for each stream cell are derived. The output of the flow direction request is an integer grid whose values range from 1 to 255. The values for each direction from the center are:

$$
\begin{array}{ccc}
32 & 64 & 128 \\
16 & X & 1 \\
8 & 4 & 2
\end{array}
$$

For example, if the direction of steepest drop were to the left of the current processing cell, its flow direction would be coded as 16. If a cell is lower than its 8 neighbors, that cell is given the value of its lowest neighbor and flow is defined towards this cell (ESRI, 2010).

The accumulated flow is based upon the number of cells flowing into each cell in the output grid. The current processing cell is not considered in this accumulation. Output cells with a high flow accumulation are areas of concentrated flow and may be used to identify stream channels. Output cells with a flow accumulation of zero are local topographic highs and may be used to identify ridges. The equation to calculate drainage area from a 20-meter cell sized flow accumulation grid was:

(cell value of flow accumulation grid + 1) * 400 = drainage area in meters squared

Stream slope

Stream slope was calculated for each stream reach in the state. A stream reach is not necessarily an entire stream but only the section of a stream between junctions. The GIS command streamlink was first used to find all unique streams between stream intersections or junctions. For each of these reaches, the length was calculated from the flowlength GIS command. Having the original DEM allowed us to find the maximum and minimum values for each of the stream reaches. The difference in the maximum and minimum elevations for the stream reach divided by the total reach length gave us our stream reach slope in meters per meter.

Annual average flow

Annual average flow for each stream cell location was found based on a relationship between drainage area and gauged stream flow. For 88 gauging stations in WV, covering many different rainfall, geological, and elevation regions, we assembled a table of drainage area for the gauges versus the historic annual stream flow for the gauge. After fitting a linear regression line for this data set, we found the following equation for annual stream flow setting the y intercept to zero.

Annual stream flow in cfs = 2.05 * drainage area in square miles

This equation had a corrected R^2 of .9729. The XY plot and equation are shown in Figure 1.
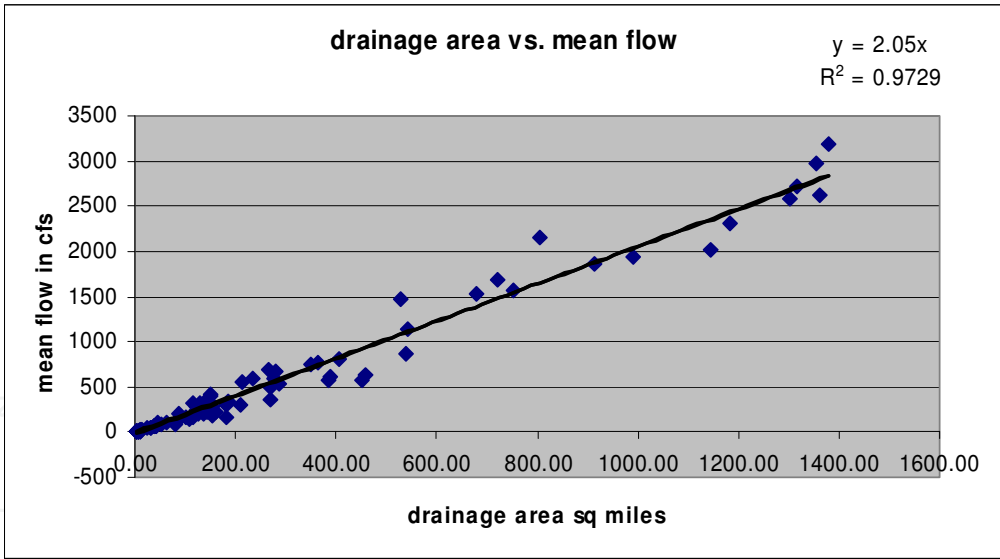

Fig. 1. Annual stream flow from gauged stations and drainage area at the gauges

Since drainage area is already calculated for each stream cell location, this equation incorporated the drainage area grid to compute a separate grid layer of annual stream flow. This would be another input for the velocity calculation.

Bank-full flow

The last input for the velocity equation was the bank-full flow measure. Just as with annual average flow, this required a modeled value for every raster stream cell in WV. Using the

same approach to regressing drainage area to gauged stream flow as performed to find an annual average flow equation, this equation was used to find bank-full flow. Bank-full flow as defined by the Bureau of Public Health, is 90% of the annual high flow. To find the 90% of high flow for each gauging station, all historic daily stream flow data was downloaded for each of the 88 gauging stations. This data was then sorted lowest to highest and then numbered lowest to highest after removing repeating values. The value of flow at the 90% of the data became the bank-full flow value for that gauge. These values were then regressed against drainage area at the gauge. The linear regression equation for bank-full stream flow setting the y intercept to zero is listed below.

Bank-full stream flow in cfs = 4.357 * drainage area in square miles

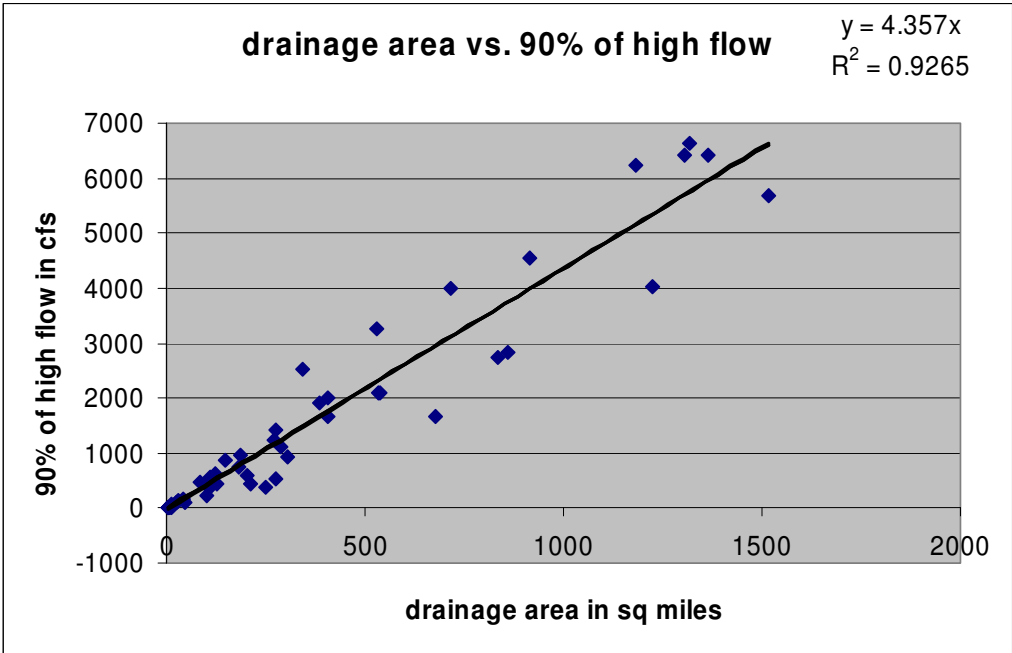This equation had a corrected R^2 of .9265. The XY plot and equation are shown in Figure 2.

Fig. 2. Bank-full stream flow from gauged stations and drainage area at the gauges

This equation could be applied to the drainage area grid to calculate the bank-full flow for any stream cell in the state. It was the final input needed in the velocity calculation.

The interactive zone of critical concern ability of SWPS delineates the upstream contributing area for a surface water intake in the following way. First, the user locates the surface water intake and makes sure the intake is on the raster stream cell. A button on the interface then initiates the model. The model will query the time of travel value for the intake and then add 18,000 seconds (5 hours) to the queried value upper range. All cells which fit this range are identified and the stream order attribute retrieved for those cells. All cells that are on the main stem stream where the intake existed are buffered 1000 feet on each side of the stream. All tributaries to the main stem are buffered 500 feet on each side of the stream. Next, a watershed boundary for the location of the intake is delineated and used to clip any areas of the buffer that may extend beyond ridgelines. And lastly, the surface water intake is buffered 1000 feet and combined with the clipped buffer to include areas 1000 feet downstream of the intake.

This interactive ability allows zones of critical concern to be delineated for any river or stream in WV. Only large rivers which border WV, such as the Ohio, Tug, and Potomac can not be interactively delineated using this method. This is due to unknown drainage areas for these bordering rivers and unknown tributaries to these major rivers coming from the bordering states. This is the major limitation of this modeling approach for WV at this time and in the next version of this watershed tool will account for all outer drainage influences.

The Ohio River Sanitation Commission (ORSANCO) is responsible for delineating zones of critical concern for the intakes along the Ohio River. ORSANCO uses uniform 25-mile upstream distances for zones of critical concern for intakes along the Ohio River. This same approach could be applied to other rivers such as the Tug and Potomac in WV.

For reservoirs and lakes within the watershed delineation area, a set of standards was set by the Bureau of Public Health and was used in this study. For a reservoir, a buffer of 1000 feet on each bank and 500 feet on each bank of the tributaries that drain into the lake or reservoir was used. When a lake or reservoir is encountered within the five-hour time of travel, a specific delineation was used. If the length of the lake/reservoir was less than or equal to the five hour calculated time of travel distance from the intake, then the entire water body was included. If the length of the lake/reservoir was greater than the calculated five hour time of travel distance from the intake, then the section of water body within the five hour time of travel distance was used to establish the zone of critical concern.

Component 5. Stream flow model from multivariate regression

Overview

This project component for SWPS used multivariate techniques to evaluate stream flow estimation variables in West Virginia. The techniques included correlation analysis, multiple regression, cluster analysis, discriminant analysis and factor analysis. The major goal was to define watershed scale factors to estimate the stream flow at recorded USGS gauges. To do this, the contributing area upstream of each gauge was first delineated. Next, annual averages of precipitation and temperature and landscape based variables for the contributing upstream area were calculated and regressed against 30-year average annual flow at the USGS gauge. Results from the statistical analysis techniques found the most important variables to be upstream drainage area, 30-year annual maximum temperature, and stream slope. While this analysis was limited by the availability of data and assumptions to predict stream flow, the results indicate that stream flow can be modeled with reasonably good results.

The following sections include a review of the literature on stream flow estimation techniques, a description of the variables used in this study to predict stream flow, the multivariate statistical methods, and a discussion of results and limitations of the study.

Literature Review

The intent of this literature review was to determine variables that were used to estimate stream flow in other studies, identify different statistical procedures, and to find limitations in this study based on other papers.

The impact of land-use, climate change and groundwater abstraction on stream flow was examined by Qerner et al. (1997). They analyzed the effects of these factors using physical

models BILAN, HBVOR, MODFLOW and MODGROW. The models were used to simulate the impact of afforstation, climate warming by 2 and 4 degrees Celsius in combination with an adoption of the precipitation changes in groundwater recharge and groundwater abstractions on stream flow droughts. The authors found that all the physical models can be used to assess the impacts of human activities on stream flow. They also concluded that based on some climate change scenarios they followed out, that the deficit volume of water is very sensitive to both an increase in temperature and a change in precipitation. Even in basins with abundant precipitation, the warming of 2 degrees Celsius would result in a rise in the deficit volume of water by 20 percent. Their findings also acknowledge the importance of using precipitation, temperature, groundwater recharge and groundwater abstractions along with water storage holding capacity of watersheds.

Timofeyeva and Craig (1998) used Monte Carlo techniques to estimate month by month variability of temperature and precipitation for drainage basins delineated by a digital elevation model. They also used a runoff grid from the digital elevation model to estimate discharge at selected points and compared this to known gauge station data. The variance of temperature was modeled as the standard error of the regression from the canonical regression equation. For precipitation, they modeled the variance as the standard error of the prediction. This was done to achieve unbiased estimators. When comparing the climate and resulting runoff and stream flow estimators calculated by Monte Carlo estimation, to the observed flow, the simulated results were within the natural variability of the record (Timofeyeva and Craig, 1998).

Long-range stream flow forecasting using nonparametric regression procedures was developed by Smith, (1991). The forecasting procedures, which were based solely on daily stream flow data, utilized nonparametric regression to relate a forecast variable to a covariate variable. The techniques were adopted to develop long-term forecasts of minimum daily flow of the Potomac River at Washington, D.C. Smith's key finding was that to implement nonparametric regression requires the successful specification of "bandwidth parameters." The bandwidth parameters are chosen to minimize the integrated mean square error of forecasts. Basically, his stream flow technique focussed on examining past history of stream flow and making nonparametric regression forecasts based on what is likely to occur in the future. No additional variables besides historic flow were used to model future conditions.

Another nonparameteric approach to stream flow simulation was done by Sharma et al. (1997). They used kernal estimates of the joint and conditional probability density functions to generate synthetic stream flow sequences. Kernal density estimation includes a weighted moving average of the empirical frequency distribution of the data (Sharma, et al. 1997). The reason for this method is to estimate a multivariate density function. This is a nonparametric method for the synthesis of stream flow that is data driven and avoids prior assumptions as to the form of dependence (linear or non linear) and the form of the probability density function. The authors main finding was that the nonparametric method was more flexible for their study than the conventional models used in stochastic hydrology and is capable of reproducing both linear and nonlinear dependence. In addition, their results when applied to a river basin indicated that the nonparametric approach was a feasible alternative to parametric approaches used to model stream flow.

Garren (1992) noted that although multiple regression has been used to predict seasonal stream flow volumes, typical practice has not realized the maximum accuracy obtainable from regression. The forecasting methods he mentions which can help provide superior forecasting include: (1) Using only data known at forecast time; (2) principal components regression; (3) cross validation; and (4) systematically searching for optimal or near-optimal combinations of variables. Some of the variables he used included snow water equivalent, monthly precipitation, and stream flow. The testing of selection sites for a stream flow forecasting study, he feels should be based on data quality, correlation analyses, conceptual appropriateness, professional judgement, and trial and error. The use of principal components regression provides the most satisfactory and statistically rigorous way to deal with intercorrelation of variables. He concluded that the maximum forecast accuracy gain is obtained by proper selection of variables followed by the use of principal components regression and using only known data (no future variables).

The results of a multiple-input transfer function modeling for daily stream flow using nonlinear inputs was studied by Astatkie and Watt (1998). They argue that since the relationship between stream flow and its major inputs, precipitation and temperature, are nonlinear, the next best alternative is to use a multiple input transfer function model identification procedure. The transfer function model they use includes variables such as type of terrain, drainage area, watercourse, the rate of areal distribution of rainfall input, catchment retention, loss through evapotranspiration and infiltration into the groundwater, catchment storage, and melting snow. When comparing their modeling technique for stream flow to that of a nonlinear time series model, they found their transfer function model to be direct and relatively easy for modeling multiple inputs. They also found it more accurate in head to head tests against the nonlinear time series model.

Since stream flow modeling is an outcome of many runoff estimation models, the literature for deriving runoff grids is applicable to stream flow studies. Anderson and Lepisto (1998) examined the links between runoff generation, climate, and nitrate leaching from forested catchments. One of the things they sought out to prove in their study was that climate will influence the amount of nitrate that can be leached from the soil and the water flow that will transport it to the streams. They found that a negative correlation existed between stream flow and temperature. Significant positive correlation between modeled surface runoff and concentrations of nitrate was found when they considered periods of flow increases during cold periods. Their study identified the importance for identifying and calculating the surface runoff fraction, daily dynamics of soil moisture, groundwater levels, and extensions of saturated areas when doing a contaminant transport or flow estimation study.

In another study, Moore (1997) sought to provide an alternative to the matching strip, correlation, and parameter-averaging methods for deriving master recession characteristics from a set of recession segments. The author then choose to apply the method to stream flow recession segments for a small forested catchment in which baseflow is provided by drainage of the saturated zone in the shallow permeable soil. The plots indicated the recessions were non-linear and that the recessions did not follow a common single valued storage outflow relation. The final decision was a model with two linear reservoirs that provided substantially better fit than three single reservoir models, indicating that the form of the recession curve probably depends on not just the volume of subsurface storage, but also on its initial distribution among reservoirs.

Gabriele et al. (1997) developed a watershed specific model to quantify stream flow, suspended sediment, and metal transport. The model, which estimated stream flow, included the sum of three major components: quick storm flow, slow storm flow, and long-term base flow. Channel components were included to account for timing effects associated with waters, sediments, and metals coming from different areas. Because of relatively good results from the modeling process, the conceptualizations supported that the study area river was strongly influenced by three major components of flow: quick storm flow, slow storm flow, and long-term base flow. Therefore, sediment inputs can be associated with each of those stream flow components and assign metal pollution concentrations to each flow and sediment input.

From this review of other studies, variables were determined that have been used successfully in stream flow estimation. Examining the limitations of other studies has also provided insight into data layers that may not be able to include. Of the statistical techniques used, the multivariate approach, in which components are added or subtracted to achieve the best fit possible, is a sound statistical procedure. In addition to this approach, testing the correlations between variables is another way of finding a model for estimating stream flow in WV.

Methodology

The first step in assembling data for this study was to delineate the total upstream contributing area for each of thirteen USGS gauge stations in West Virginia. Figure 3 displays the location of each gauge and the defined upstream drainage area for that gauge.
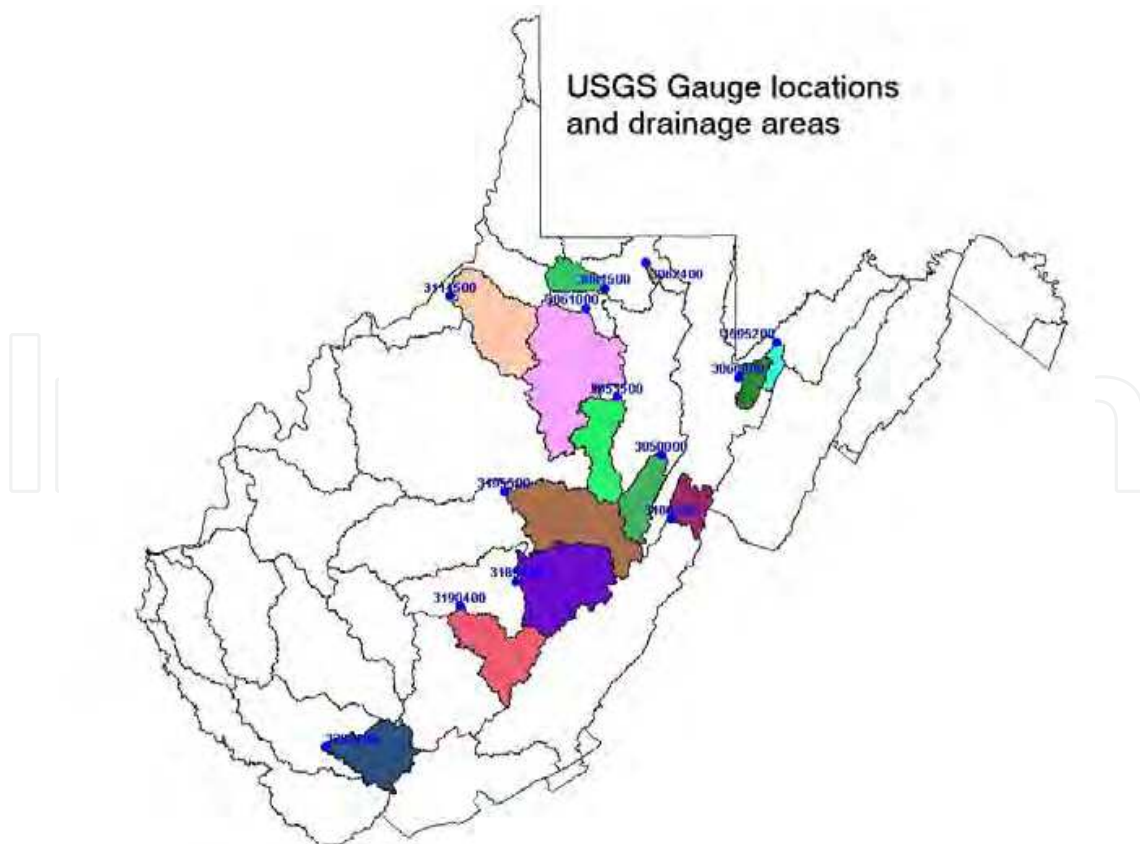

Fig. 3.

For every drainage area, the following criteria were calculated; total area, 30 year average annual precipitation, 30 year average annual maximum temperature, 30 year average annual minimum temperature, average drainage area slope, and stream slope. These variables were explanatory variables, which would be regressed against the dependent variable, the 30year average annual flow recorded at the gauge stations. The figures 4 to 7 show the distribution of 30-year precipitation, maximum temperature, minimum temperature, and elevation across the different areas. By using GIS techniques, it was possible to find the average value in the drainage areas along with drainage area slope and stream slope for each of the variables. The data for each gauge area and assembled variables is summarized in table 1.

id#	USGS Gauge name	Upstream drainage area (acres)	30yr annual precip ave (inches)	30yr annual ave temp max(F)	30yr annual ave temp min(F)	30yr annual Stream flow (cfs)	stream elevation drop max-min in (meters)	Watershed Slope average (degree)
g1	1595200	31296	52	54	35	99.68	418	5
g5	3050000	120352	50	57	37	379.37	607	15
g7	3053500	176708	46	60	38	613.56	643	11
g10	3061000	484507	43	62	39	1158.14	26	13
g11	3061500	74501	42	61	39	168.99	130	13
g12	3062400	7146	43	60	37	16.54	189	9
g13	3066000	55068	53	54	36	210.40	289	6
g17	3114500	289609	42	62	40	665.40	57	15
g19	3180500	85166	53	55	36	273.49	435	14
g21	3189100	338131	53	57	37	1445.61	743	13
g22	3190400	232990	50	59	38	750.36	830	11
g24	3195500	346231	48	59	38	1176.66	933	17
g26	3202400	196645	47	63	39	421.78	583	18

Table 1. Data used in study

The first step in analyzing the data in table 1 was to perform some basic statistics. The values across the different gauging station locations were investigated. The summarized statistical data is shown in table 2.

Variable	N	Mean	Median	TrMean	StDev	SE Mean	Minimum	Maximum	Q1	Q3
area	13	187565	176708	176972	144629	40113	7146	484507	64784	313870
precip	13	47.85	48	47.91	4.34	1.2	42	53	43	52.5
maxtemp	13	58.692	59	58.727	3.066	0.85	54	63	56	61.5
mintemp	13	37.615	38	37.636	1.446	0.401	35	40	36.5	39
flow	13	568	422	538	455	126	17	1446	190	954
strmslop	13	452.5	435	447.6	299.3	83	26	933	159.5	693
wsslope	13	12.31	13	12.45	3.88	1.08	5	18	10	15

Table 2. Basic statistics

30yr Annual Precipitation and drainage areas

Fig. 4.

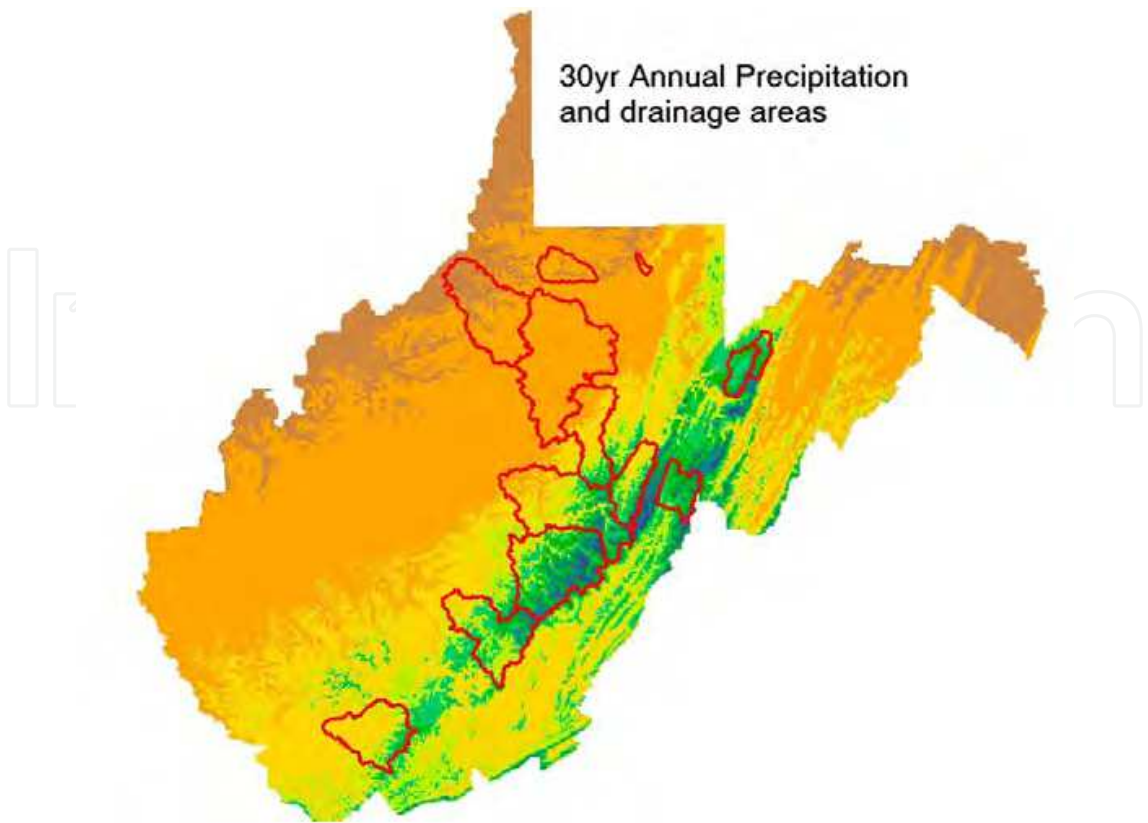
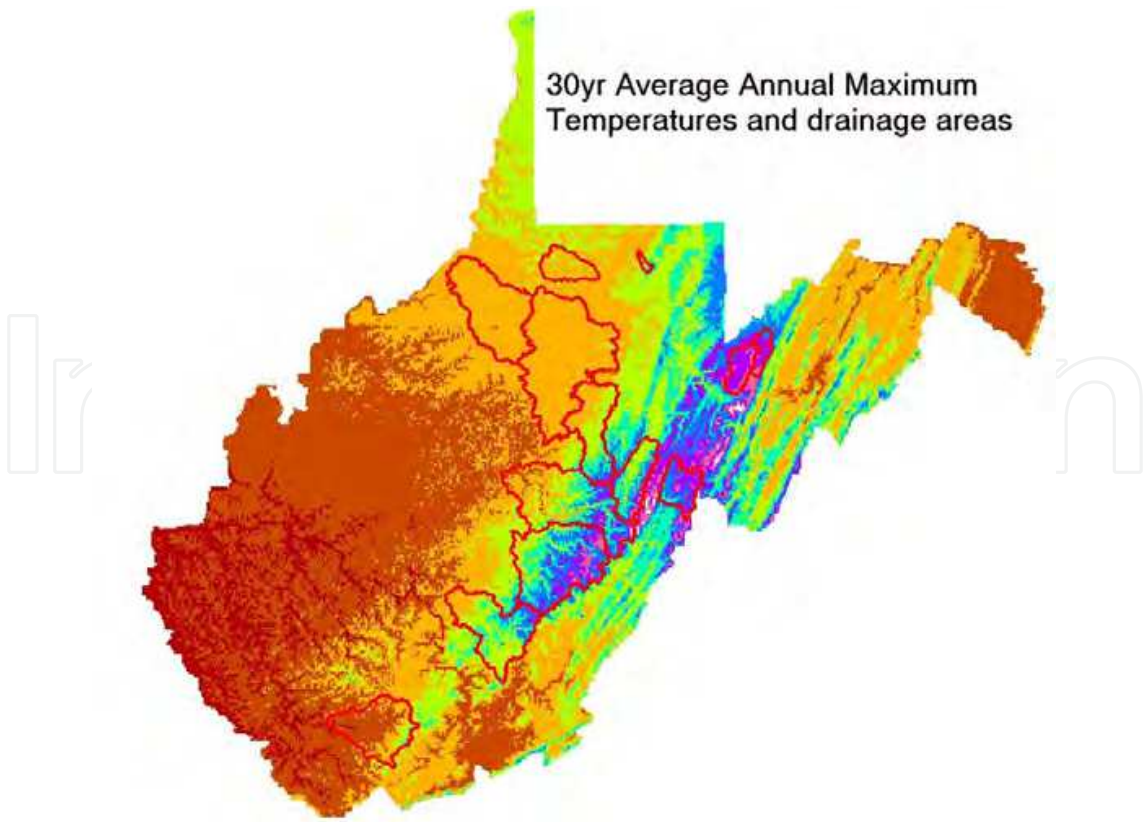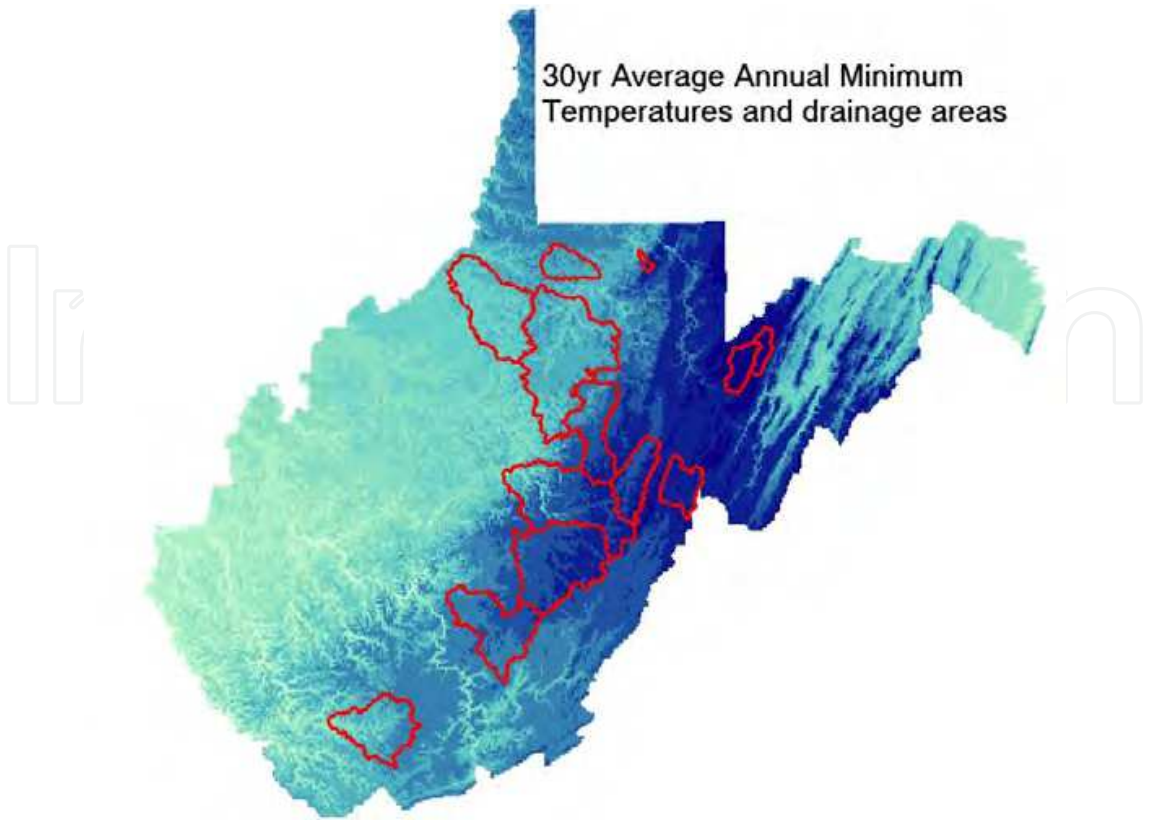30yr Average Annual Maximum Temperatures and drainage areas
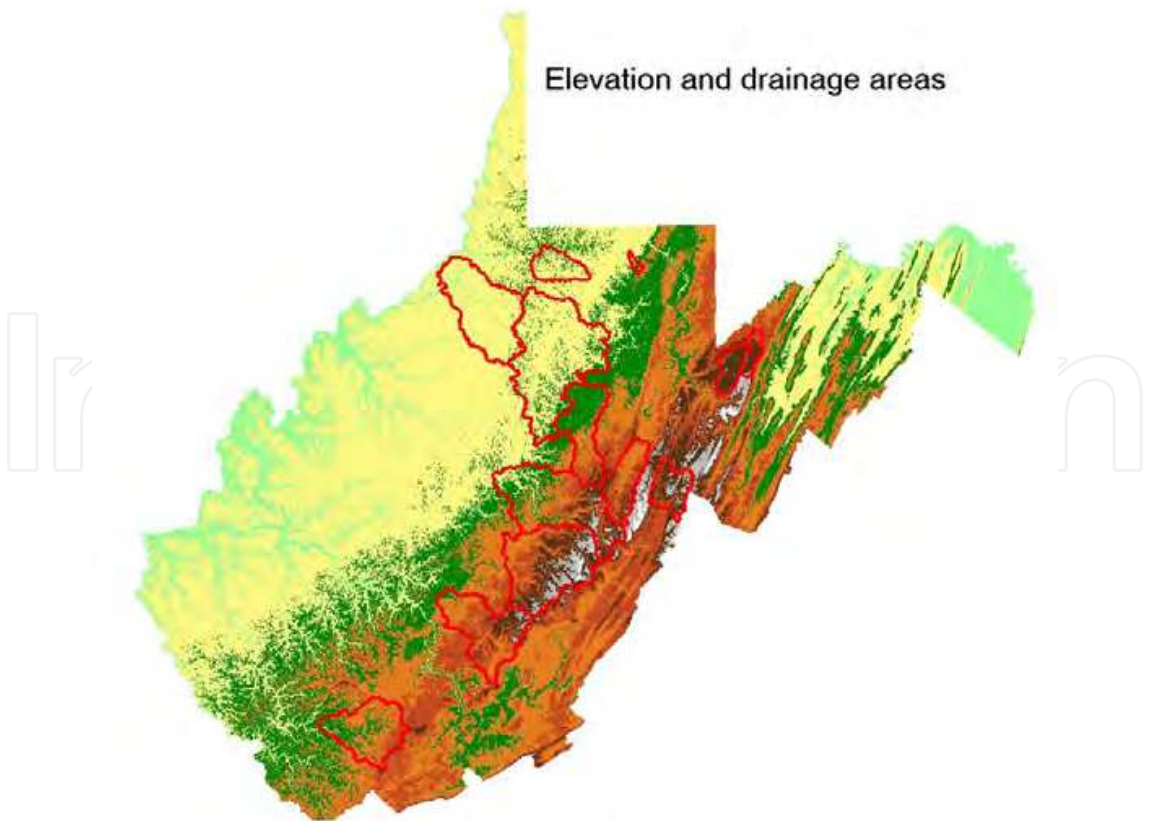
Fig. 5.

Fig. 6.

Fig. 7.

From table 2, it was noted which variables were closely grouped and which varied significantly among all the 13 different gauges. The area and flow variables have the highest standard deviation while the precipitation, maximum and minimum temperatures, and watershed slope have the lowest standard deviation. Other simple statistical graphs, which were used to gain insights into the data distribution and spreads, are shown in figures 8 to 14. The figures provided a graphical display of the distribution of values across the 13 gauges. Data exploration is important to determine trends and outliers in data that may bias results (Johnson, et al 2001). In addition, regression results may be impacted from large variations in data values. A common technique is to normalize data with a simple equation such as the value of interest minus the minimum value for that variable divided by the maximum minus minimum within the data range (Kachigan, 1986). However, in this study the values were not normalized due to the spatial nature of the information source. It was necessary to identify and incorporate the spatial variability across the entire study area at the statewide level. The end use of our regression relationship is the ability to query any raster stream cell and report all the unique information from the spatial analysis. Stream flow and water quality decisions for permitting may occur in high elevation cold headwater segments as well as large river systems with much accumulated drainage. Because the study area had unique topographic features that were to be regressed against representative stream flow information, the gauge driven delineated watersheds were chosen to represent this differentiation as best as possible as shown in Figure 3.

Fig. 8.

Fig. 9.

Fig. 10.

Descriptive Statistics

Variable: mintemp

Anderson-Darling Normality Test

A-Squared:	0.298
P-Value:	0.534
Mean	37.6154
StDev	1.4456
Variance	2.08974
Skewness	-1.7E-01
Kurtosis	-6.4E-01
N	13
Minimum	35.0000
1st Quartile	36.5000
Median	38.0000
3rd Quartile	39.0000
Maximum	40.0000

95% Confidence Interval for Mu

36.7418	38.4889

95% Confidence Interval for Sigma

1.0366	2.3863

95% Confidence Interval for Median

36.6846	39.0000

Fig. 11.

Descriptive Statistics

Variable: precip

Anderson-Darling Normality Test

A-Squared:	0.523
P-Value:	0.148
Mean	47.8462
StDev	4.3368
Variance	18.8077
Skewness	-1.6E-01
Kurtosis	-1.62409
N	13
Minimum	42.0000
1st Quartile	43.0000
Median	48.0000
3rd Quartile	52.5000
Maximum	53.0000

95% Confidence Interval for Mu

45.2255	50.4668

95% Confidence Interval for Sigma

3.1098	7.1589

95% Confidence Interval for Median

43.0000	52.3154

Fig. 12.

Descriptive Statistics

Variable: strmslope

Anderson-Darling Normality Test

A-Squared:	0.228
P-Value:	0.764
Mean	452.538
StDev	299.299
Variance	89579.8
Skewness	2.30E-02
Kurtosis	-1.23275
N	13
Minimum	26.000
1st Quartile	159.500
Median	435.000
3rd Quartile	693.000
Maximum	933.000

95% Confidence Interval for Mu
271.674   633.403

95% Confidence Interval for Sigma
214.623   494.063

95% Confidence Interval for Median
170.393   674.537

Fig. 13.

Descriptive Statistics

Variable: wsslope

Anderson-Darling Normality Test

A-Squared:	0.319
P-Value:	0.493
Mean	12.3077
StDev	3.8813
Variance	15.0641
Skewness	-5.9E-01
Kurtosis	-1.5E-01
N	13
Minimum	5.0000
1st Quartile	10.0000
Median	13.0000
3rd Quartile	15.0000
Maximum	18.0000

95% Confidence Interval for Mu
9.9623   14.6531

95% Confidence Interval for Sigma
2.7832   6.4069

95% Confidence Interval for Median
10.3693   15.0000
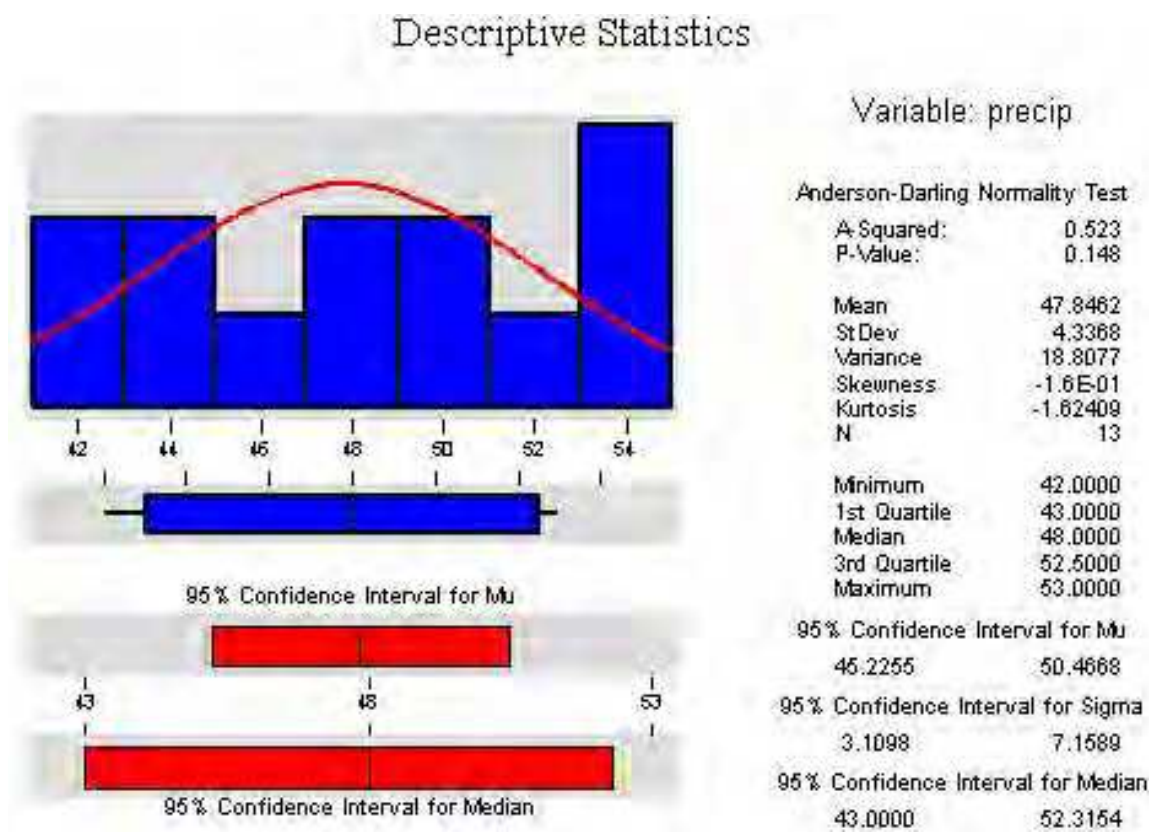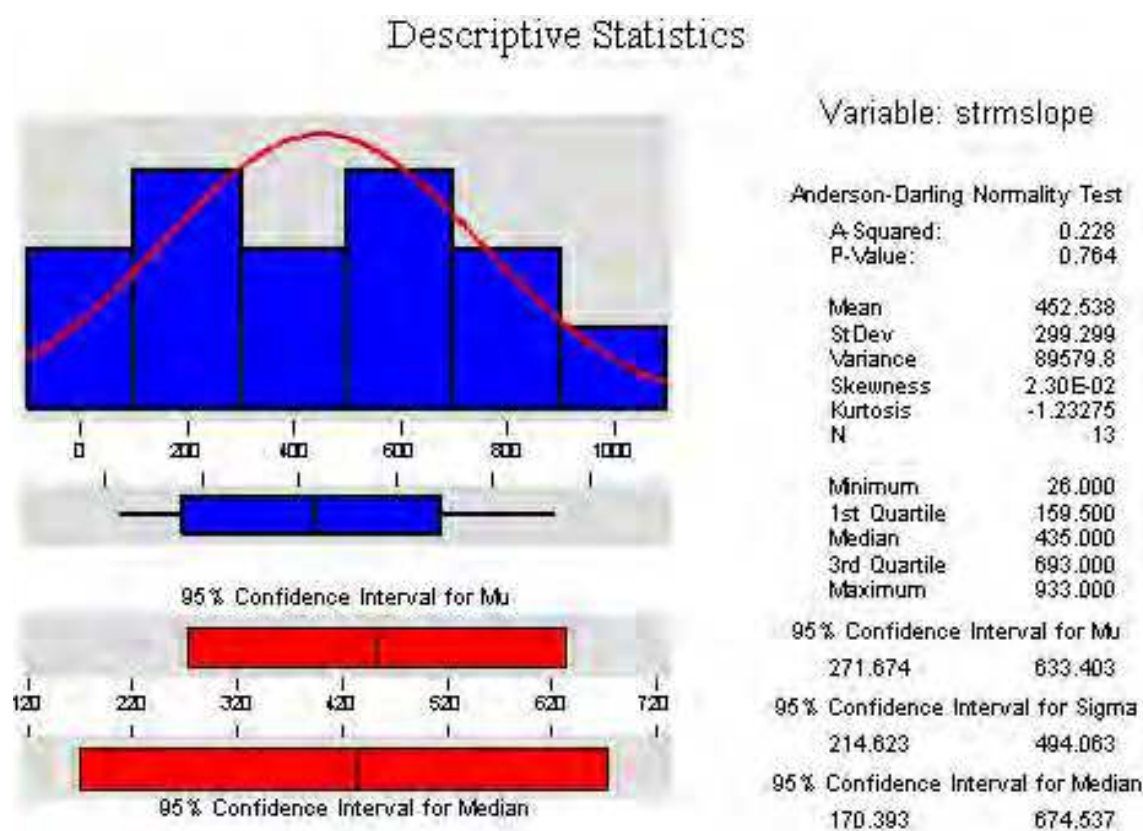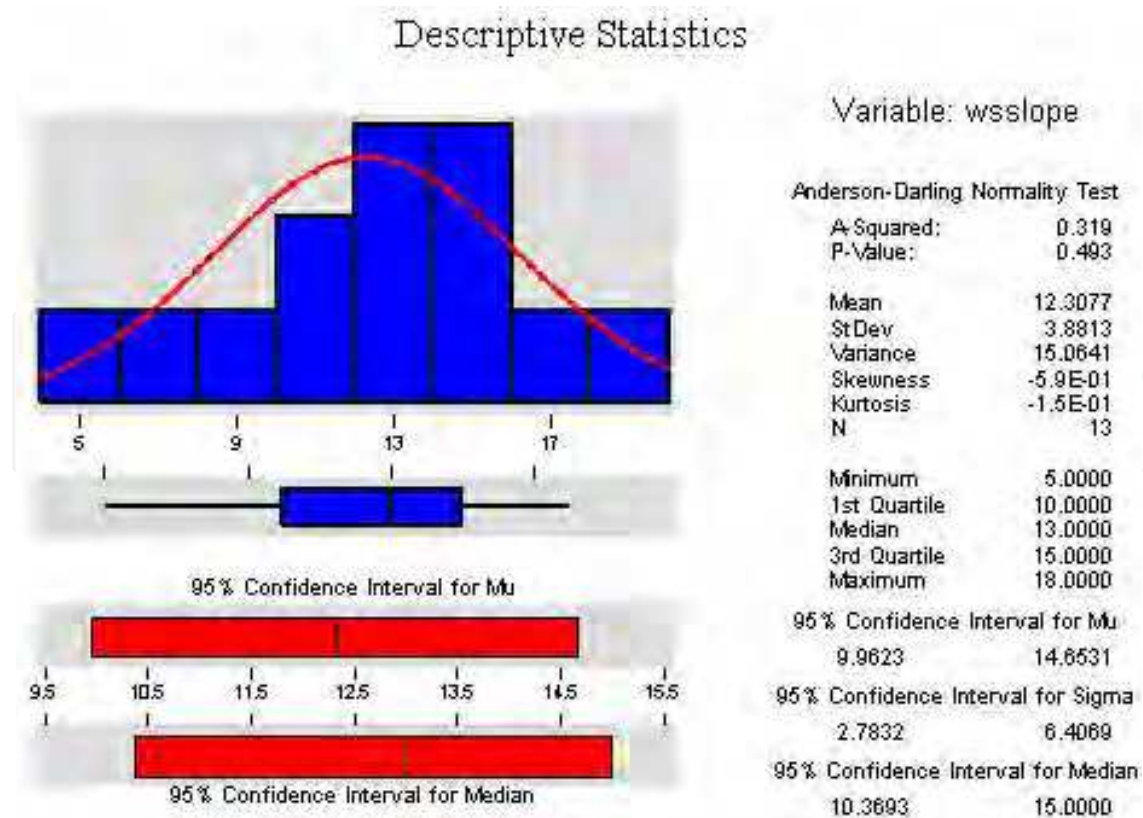
Fig. 14.

The next step in analyzing the data was to use generate a best-fit line plot for each of the independent variables in table 1 regressed against the dependent variable stream flow. These plots are shown in figures 15 to 20. From these best-fit line plots, the area, stream slope, and watershed slope variables had the best R squared values and positive linear relationship. The maximum and minimum temperature variables along with precipitation had the worst linear fit with stream flow. Their R squared values were very low with the precipitation variable looking very random in describing stream flow. At this point in the analysis it appeared that the area, stream slope and watershed slope will be the better variables to predict stream flow.

While the linear regression plots provided some idea of the extent of the relationship between two variables, the correlation coefficient gives a summary measure that communicates the extent of correlation between two variables in a single number (Kachigan, 1986). The higher the correlation coefficient, the more closely grouped are the data points representing each objects score on the respective variables. Some important assumptions of the correlation coefficient are that the data line in groupings that are linear in form. The other important assumptions include that the variables are random and measured on either an interval or a ratio scale. In addition, the last assumption for the use of the correlation coefficient is that the two variables have a bivariate normal distribution. The correlation matrix for the data used in this study is shown in table 3.

	area	precip	maxtemp	mintemp	flow	strslope	wsslope
area	1	-0.212	0.470	0.571	0.922	0.138	0.516
precip	-0.212	1	-0.850	-0.781	0.039	0.560	-0.279
maxtemp	0.470	-0.850	1	0.930	0.245	-0.226	0.590
mintemp	0.571	-0.781	0.930	1	0.356	-0.217	0.647
flow	0.922	0.039	0.245	0.356	1	0.392	0.435
strslope	0.138	0.560	-0.226	-0.217	0.392	1	0.245
wsslope	0.516	-0.279	0.590	0.647	0.435	0.245	1

Table 3. Correlation matrix

The variables with significant correlations (R > .7) are shaded in table 3. The variables listed in order of highest correlation to lowest significance are mintemp and maxtemp, flow and area, precip and maxtemp, and precip and mintemp. The correlations between the weather data were expected. In areas of higher precipitation, the temperature will be cooler (the annual averages for maximum temperature will be lower and the annual average for minimum temperatures will be lower) hence the high negative correlation. The other high positive correlated variables indicate that the variation in one variable will lead to variation in the other variable. For regression analysis the variables should be independent. Collinearity refers to linear relationships within the variables. The amount of multicollinearity across variables can be examined with principal component analysis of a sample correlation matrix (Sundberg, 2002) among other methods to remove dependence. This study examined the smallest eigenvalue and eliminated variables with values less than 0.05 as an indication of substantial collinearity (Hocking, 1996). As expected the precipitation variables were not independent to the elevation data and therefore removed.
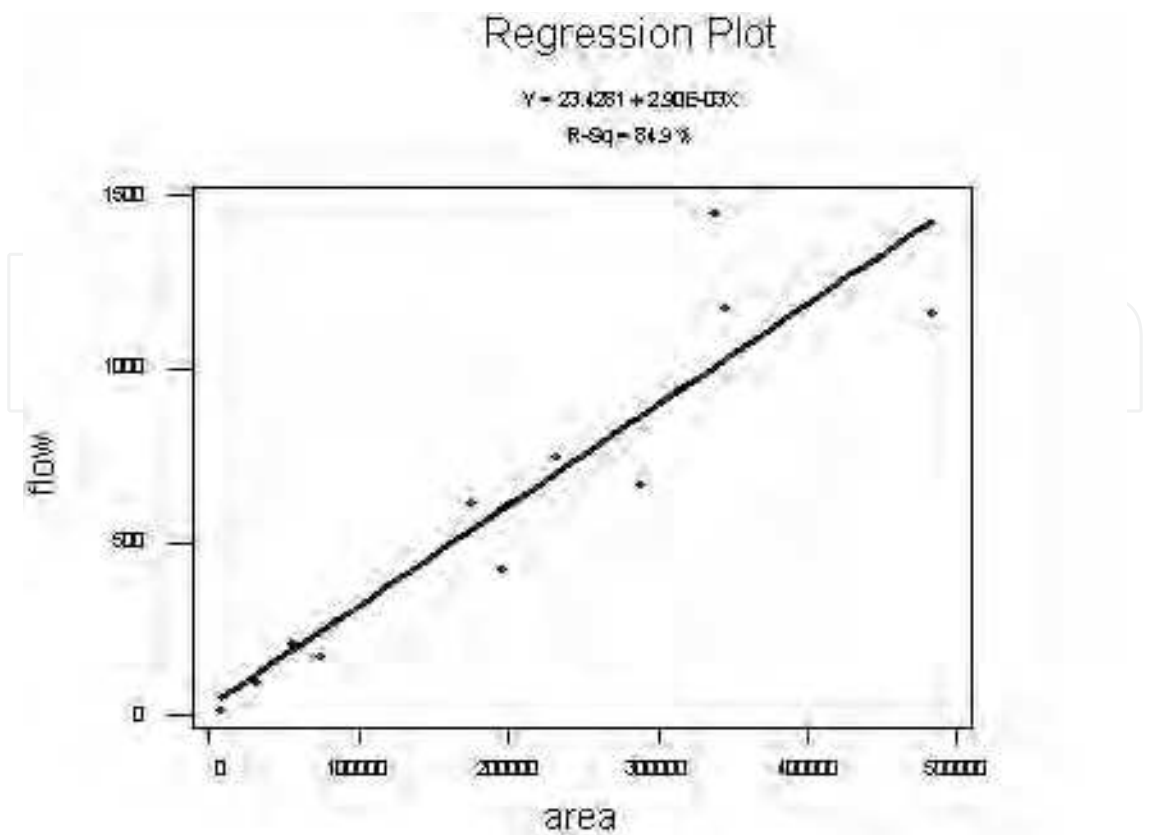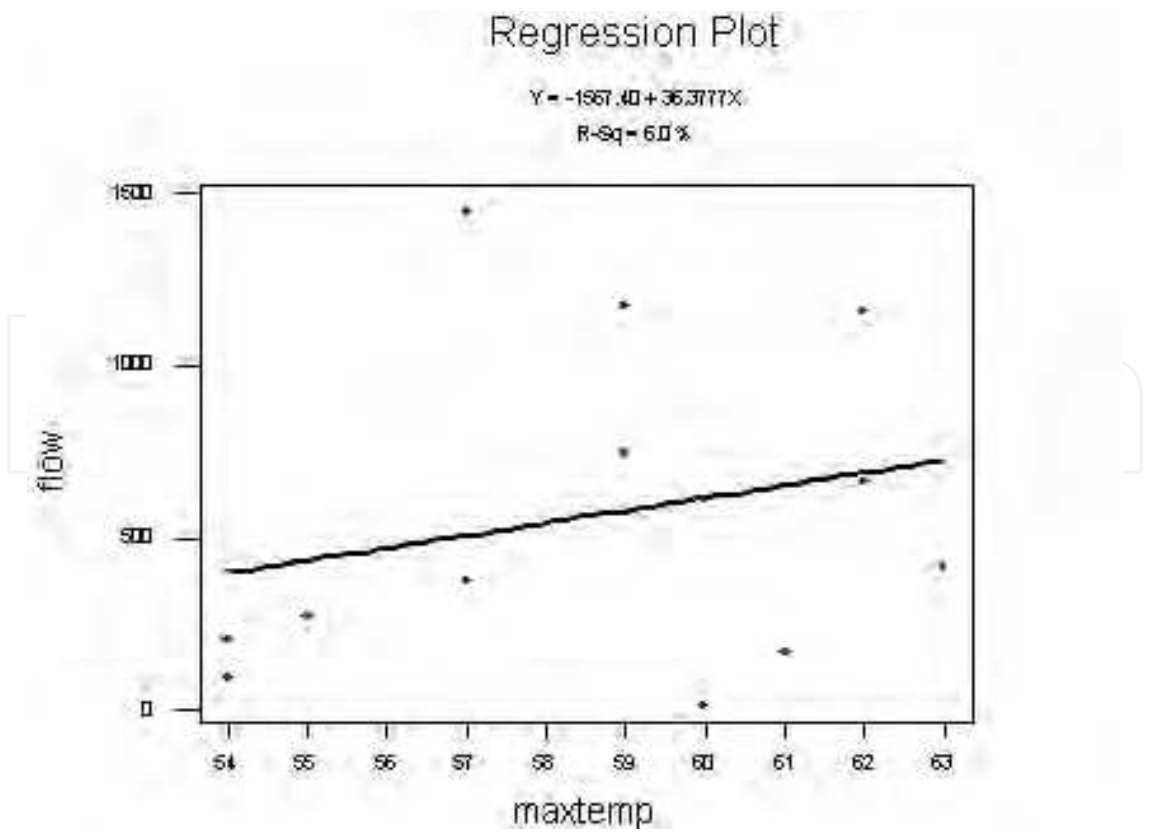
Regression Plot

$Y = 23.4281 + 2.90E\text{-}03X$

R-Sq = 84.9 %

Fig. 15.

Regression Plot

$Y = -1567.40 + 36.3777X$

R-Sq = 6.0 %

Fig. 16.

Regression Plot

Y = -3647.07 + 112.049X
R-Sq = 12.7 %

Fig. 17.

Regression Plot

Y = 374.266 + 4.04263X
R-Sq = 0.1 %
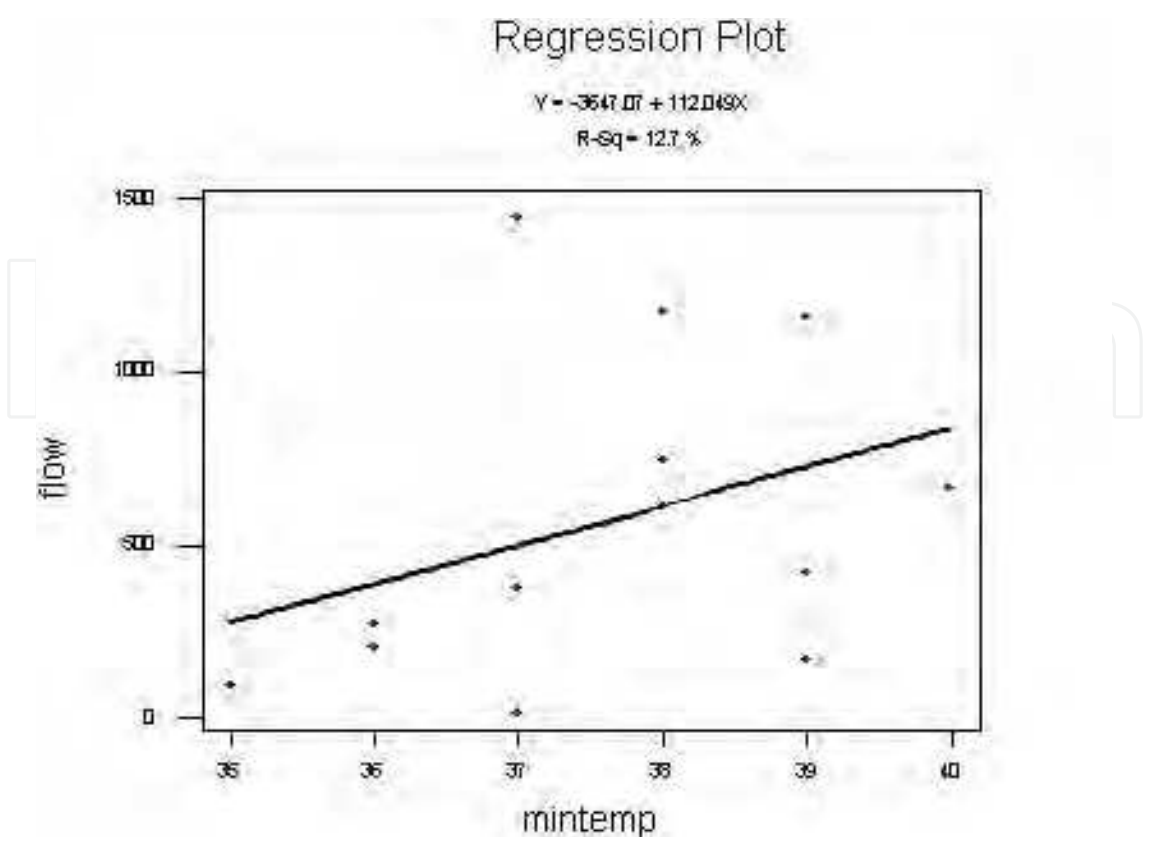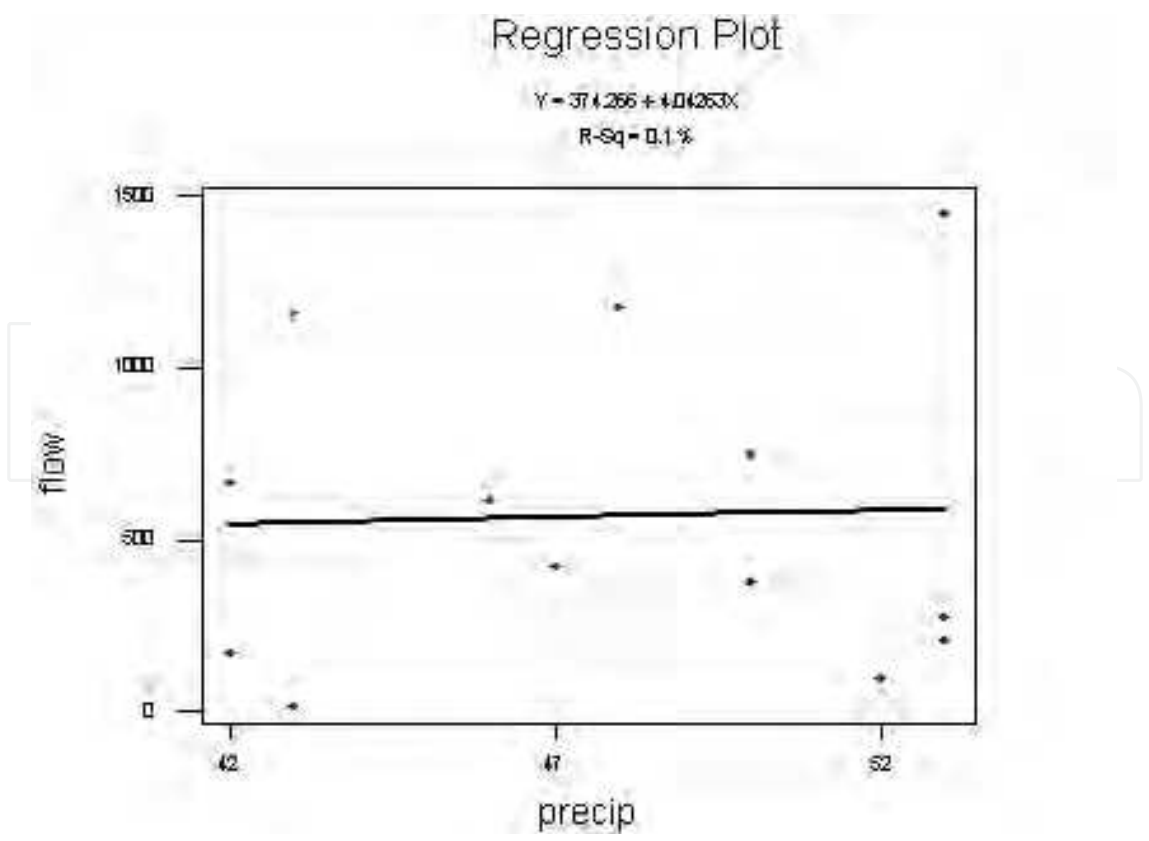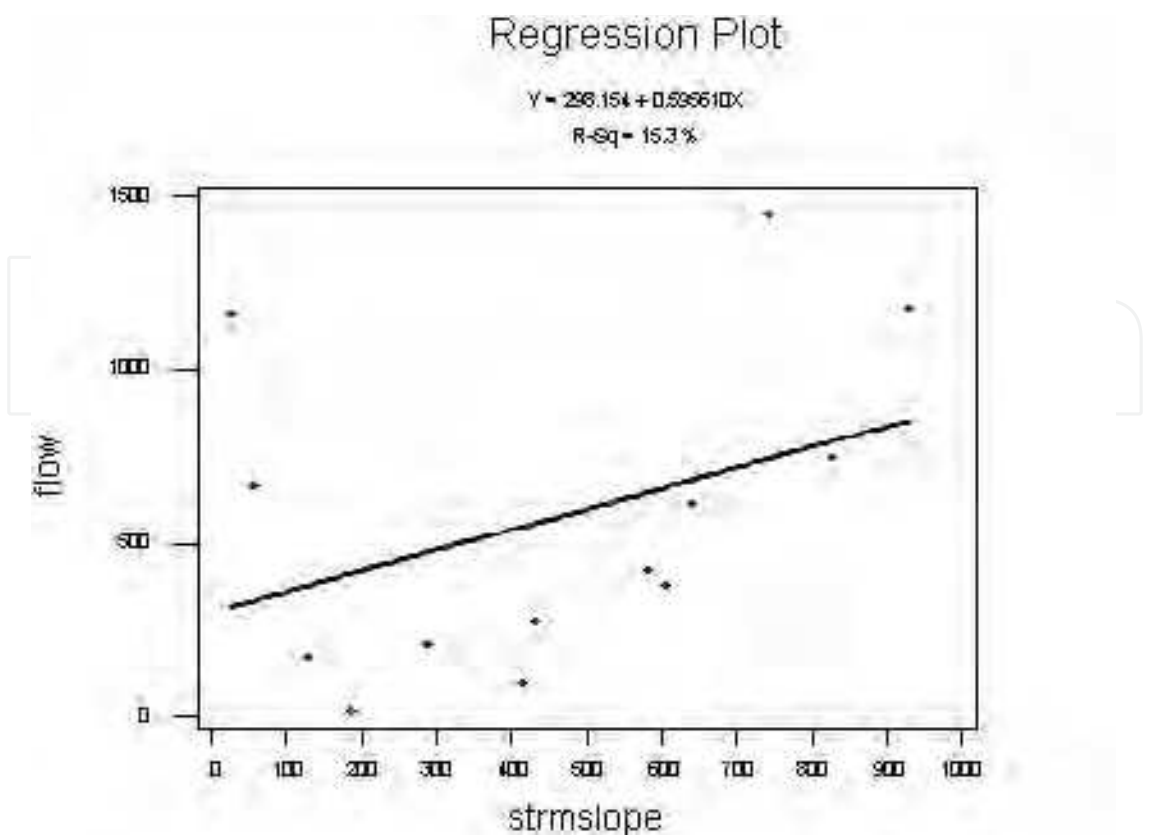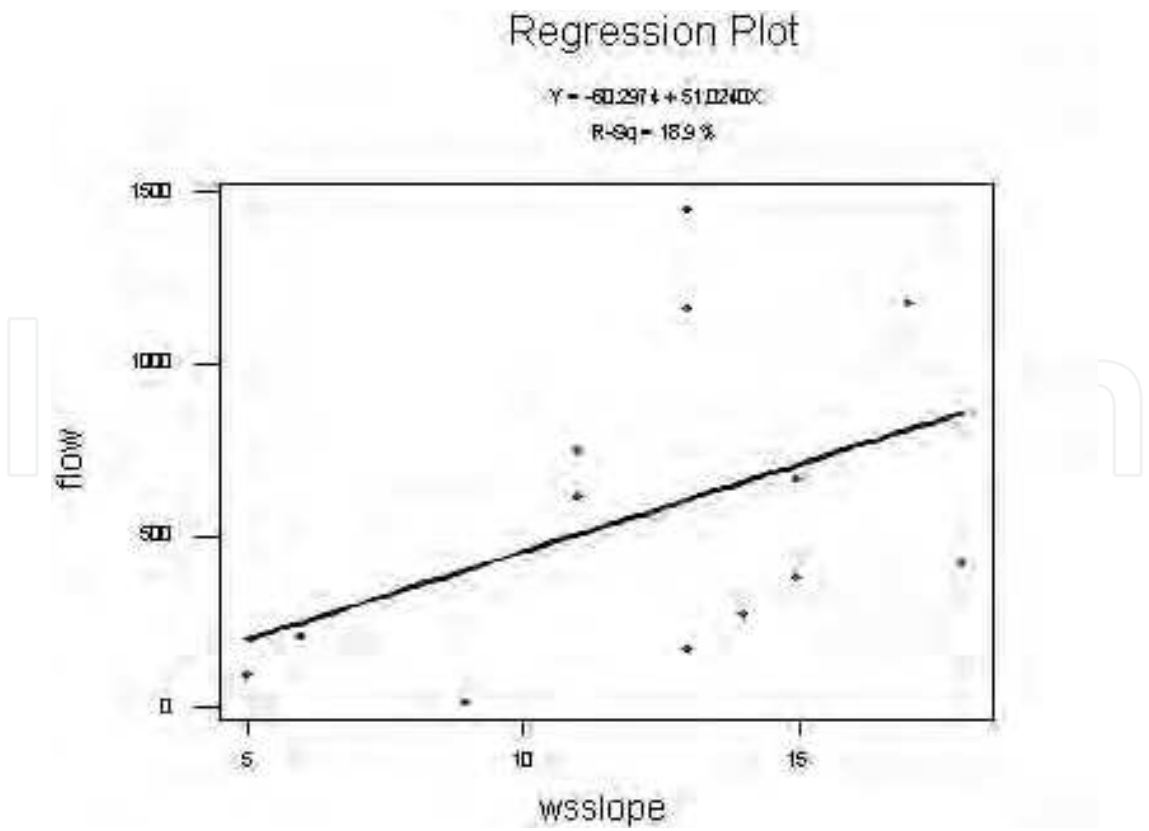

Fig. 18.

Fig. 19.

Fig. 20.

Performing regression analysis on the data was the next step in formulating a relationship and model to predict and estimate stream flow. Using the technique by Garren (1992) a regression equation with all the remaining variables was created, evaluate the P values of each variable, and eliminate variables until the highest adjusted R square is found. The first run with the regression analysis indicates that the variables area, strmslop and maxtemp will have the most influence on flow because of their low P values. Table 4 shows the regression analysis including all the variables.

```
The regression equation is
flow = 2325 + 0.00310 area - 12.0 precip - 37.3 maxtemp + 8 mintemp
         + 0.423 strmslope - 4.8 wsslope


Predictor         Coef       StDev          T          P
Constant          2325        4370       0.53      0.614
area         0.0030987   0.0004281       7.24      0.000
precip          -12.02       30.84      -0.39      0.710
maxtemp         -37.31       53.50      -0.70      0.512
mintemp            7.8       100.7       0.08      0.941
strmslop        0.4235      0.2281       1.86      0.113
wsslope          -4.83       18.54      -0.26      0.803


S = 156.3       R-Sq = 94.1%     R-Sq(adj) = 88.2%
```

Table 4. Regression analysis including all variables

By systematically removing the variables with a high P value and noting the R squared adjusted value, it was possible to arrive at a final set of variables to use in a regression equation to estimate stream flow. Table 5 shows the regression analysis results after removing the variable with the highest P value (mintemp).

```
The regression equation is
flow = 2492 + 0.00311 area - 12.3 precip - 35.0 maxtemp + 0.421
strmslope
         - 4.3 wsslope

Predictor         Coef       StDev          T          P
Constant          2492        3522       0.71      0.502
area         0.0031121   0.0003628       8.58      0.000
precip          -12.35       28.30      -0.44      0.676
maxtemp         -35.00       41.23      -0.85      0.424
strmslop        0.4208      0.2089       2.01      0.084
wsslope          -4.33       16.11      -0.27      0.796

S = 144.7       R-Sq = 94.1%     R-Sq(adj) = 89.9%
```

Table 5. Regression analysis with mintemp removed

The R squared adjusted improved slightly to 89.9% with mintemp removed. This process of removing the current highest P value variable and re-running of the model was repeated six times. The associated R squared values were noted and table 6 was created from the results.

As table 6 indicates, the combination of variables that provided the highest R squared adjusted value were area, maxtemp, and strslope. The associated regression equation with the optimal set of variables is:

$$flow = 1232 + 0.00304\ area - 23.6\ maxtemp + 0.338\ strmslope$$

Variables included in the regression	R squared adjusted
Area, mintemp, maxtemp, strslope, wsslope	88.2
Area, maxtemp, strslope, wsslope	89.9
Area, maxtemp, strslope	91.1
Area, maxtemp, strslope	91.8
Area, strslope	90.5
Area	83.6

Table 6. Multiple regression results

The next procedure used in the analysis was discriminant analysis. This technique was used to identify relationships between qualitative criterion variables and the quantitative predictor variables in the dataset. The objective was to identify boundaries between the groups of watersheds that the gauges were associated. The boundaries between the groups are the characteristics that distinguish or discriminate the objects in the respective groups. Discriminant analysis allows the user to classify the given objects into groups – or equivalently, to assign them a qualitative label – based on information on various predictor or classification variables (Kachigan, 1992).

The gauge station dataset was assigned a qualitative variable based on which major drainage basin in West Virginia the area was located. The major basins used were the Monongahela (m), Gauley (g) and Other (x). The class "other" was assigned to gauges that did not fall in the Monongahela or Gauley drainage basins. Running the discriminant analysis in Minitab produced the results shown in table 7.

Only gauge one and gauge five were reclassified from the discriminant analysis results. It should be noted however that the discriminant function should be validated by testing its efficacy with a fresh sample of analytical objects. Kachigan (1986) notes that the observed accuracy of prediction on the sample upon which the function was developed will always be spuriously high, because we will have capitalized on chance relationships. The true discriminatory power of the function will be found when tested with a completely separate sample.

By using discriminant analysis, it enabled the investigation of how the given groups differ. In the next analysis step, cluster analysis, the goal is to find whether a given group can be partitioned into subgroups that differ. The advantage of the approach is in providing a better feel of how the clusters are formed and which particular objects are most similar to one another.

The cluster analysis was performed with distance measures of Pearson and Average and link methods of single and Euclidean. The Average and Euclidean choices worked the best in identifying clusters. Figure 21 shows the dendrogram results and table 8 lists the computation results.

```
Linear Method for Response:    class

Predictors:    area   precip   maxtemp   mintemp   flow   strslope   wsslope

Group          g          m          x
Count          2          5          6

Summary of Classification
Put into        ....True Group....
Group                   g          m          x
g                       2          0          0
m                       0          4          1
x                       0          1          5
Total N                 2          5          6
N Correct               2          4          5
Proportion    1.000     0.800      0.833

N =   13      N Correct =   11      Proportion Correct = 0.846
Squared Distance Between Groups
                  g          m          x
g            0.0000    14.5434    17.0393
m           14.5434     0.0000     4.5539
x           17.0393     4.5539     0.0000

Linear Discriminant Function for Group
                  g          m          x
Constant    -7379.7    -7053.9    -7003.2
area            -0.0       -0.0       -0.0
precip          85.6       83.5       83.6
maxtemp         86.5       84.2       84.3
mintemp        157.5      155.0      153.0
flow             0.8        0.7        0.7
strslope        -0.3       -0.3       -0.3
wsslope        -38.0      -37.0      -36.6

Summary of Misclassified Observations
Observation       True      Pred      Group      Squared     Probability
                  Group     Group                Distance
    1 **            x         m          g        20.956         0.000
                                         m         5.163         0.578
                                         x         5.796         0.421
    2 **            m         x          g        23.906         0.000
                                         m         5.223         0.229
                                         x         2.790         0.771

gauge id    majshed              class    FITS1
g1          NorthBranch          x        m
g5          Tygart               m        x
g7          Tygart               m        m
g10         WestFork             x        x
g11         MonRiver             m        m
g12         MonRiver             m        m
g13         Cheat                m        m
g17         MiddleOhio           x        x
g19         Greenbrier           x        x
g21         Gauley               g        g
g22         Gauley               g        g
g24         Elk                  x        x
g26         UpGuyandotte         x        x
```

Table 7. Discriminant analysis

```
Standardized Variables, Euclidean Distance, Average Linkage
Amalgamation Steps
Step Number of Similarity  Distance  Clusters    New    Number of obs.
     clusters    level       level    joined   cluster  in new cluster
  1     12       85.24       0.933    1    7      1            2
  2     11       80.06       1.261    3   11      3            2
  3     10       78.24       1.376    2    9      2            2
  4      9       70.39       1.872    5    6      5            2
  5      8       69.18       1.948    4    8      4            2
  6      7       68.16       2.013   10   12     10            2
  7      6       61.42       2.439    3   10      3            4
  8      5       56.79       2.732    1    2      1            4
  9      4       54.13       2.900    3   13      3            5
 10      3       45.06       3.473    4    5      4            4
 11      2       41.04       3.727    3    4      3            9
 12      1       35.49       4.078    1    3      1           13
```

```
Final Partition
Number of clusters:   2
            Number of   Within cluster  Average distance Maximum distance
          observations  sum of squares   from centroid    from centroid
Cluster1       4            8.340           1.414            1.816
Cluster2       9           46.098           2.188            2.934
```

```
Cluster Centroids
Variable        Cluster1     Cluster2   Grand centrd
area            -0.7923       0.3522      -0.0000
precip           0.9578      -0.4257      -0.0000
maxtemp         -1.2045       0.5353      -0.0000
mintemp         -1.1175       0.4966       0.0000
flow            -0.7181       0.3191      -0.0000
strslope        -0.0511       0.0227      -0.0000
wsslope         -0.5946       0.2643      -0.0000
```

```
Distances Between Cluster Centroids

              Cluster1     Cluster2
Cluster1       0.0000       3.2672
Cluster2       3.2672       0.0000
```

Table 8. Hierarchical cluster analysis of observations

From the clustered results, gauges 1 and 7 (g1 and g13) are the most alike and merge into a cluster at around 85 on the similarity scale. Gauges 3 and 11 (g7 and g22) are the next most similar at the 78 level. However, these objects do not form the same cluster until a lower level of similarity around the 35 level. By clustering the objects, we were able to identify groups that are alike and because of the small dataset, it was easy to examine the data table and discover values that make the objects similar.

After cluster analysis, the choice was made to perform a factor analysis as an aid in data reduction. Although there were only seven variables, the possibility existed to gain insight into removing the duplicated information from among the set of variables. The results were assembled as a loading plot – figure 22, a score plot – figure 23, and a scree plot – figure 24. The output session data is listed in table 9.
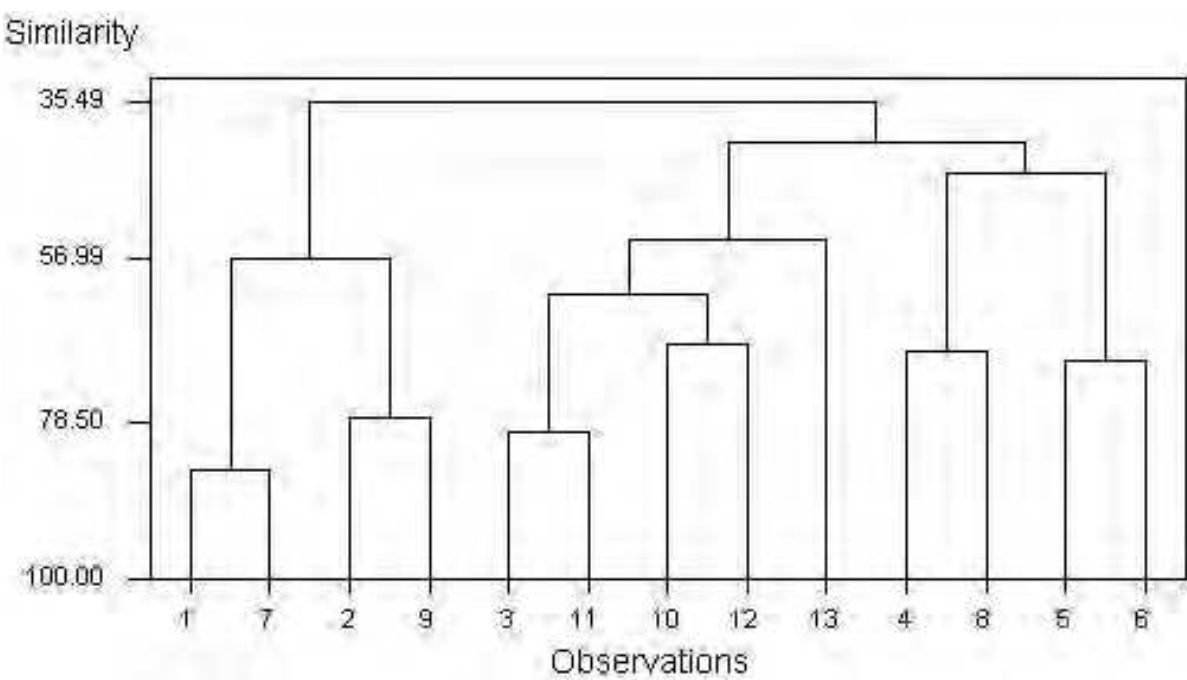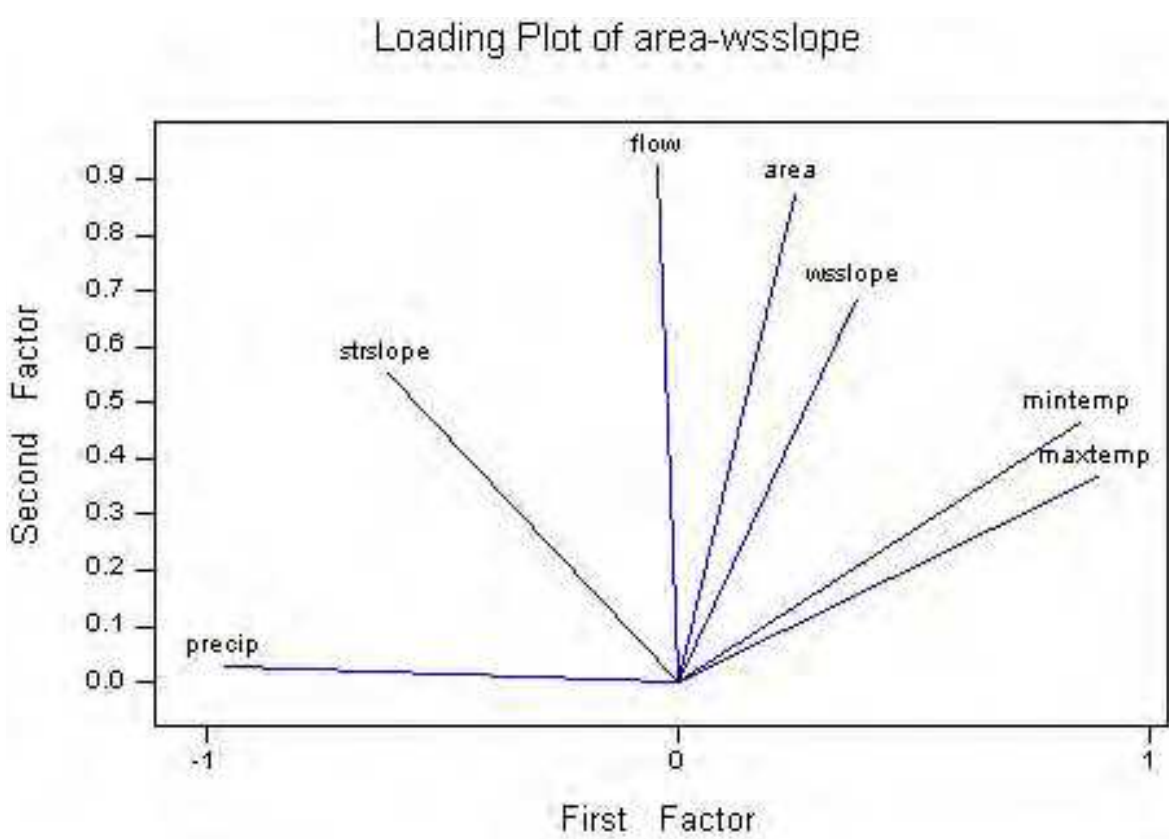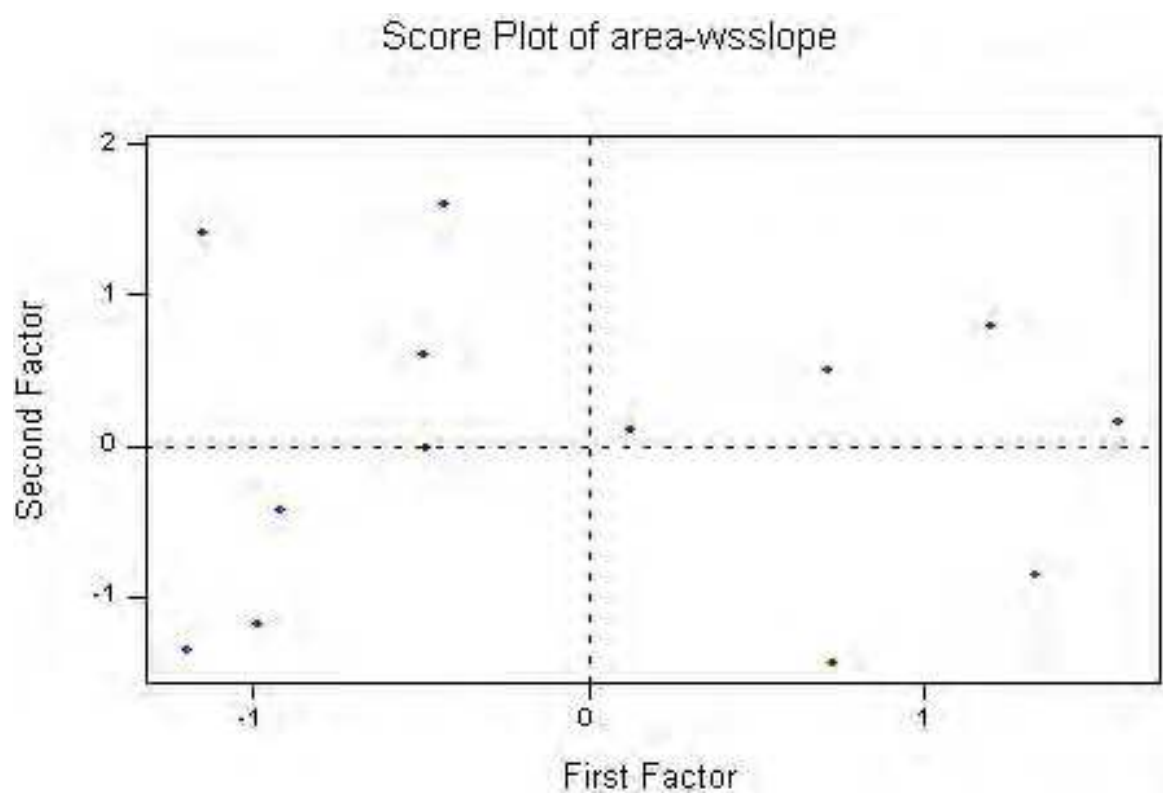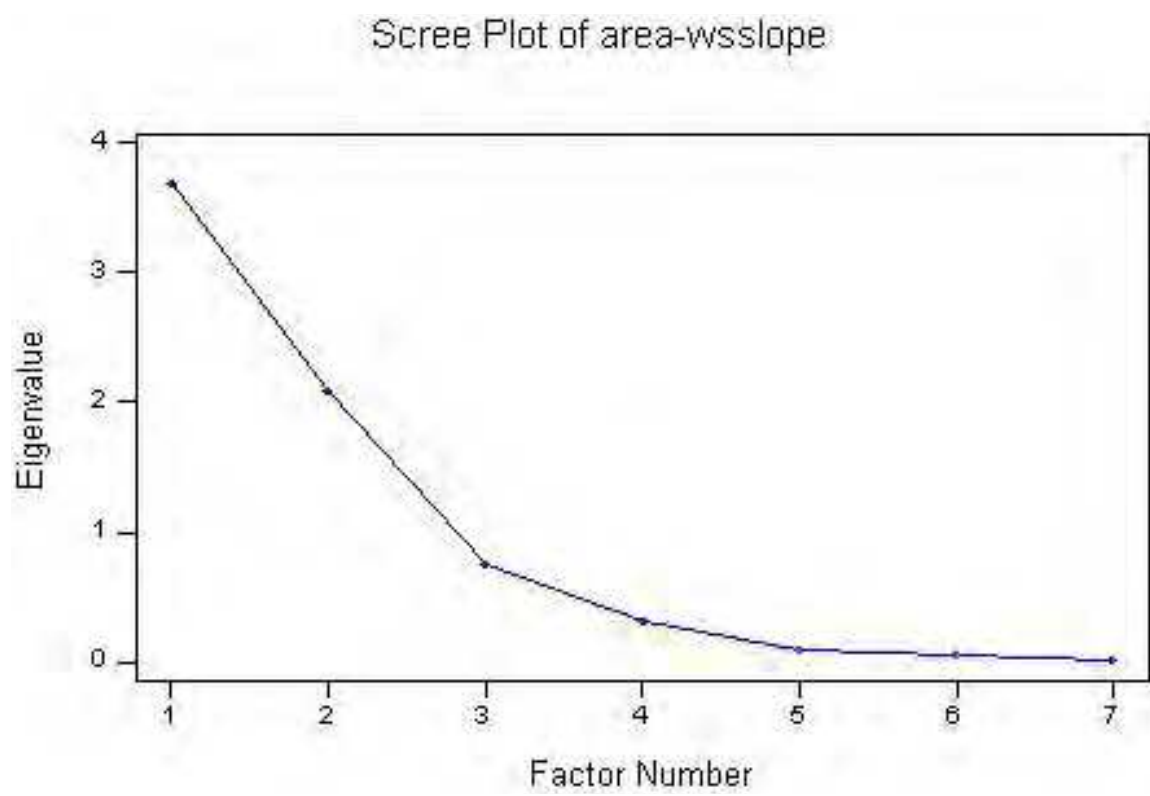
Fig. 21.

Fig. 22.

Fig. 23.

Fig. 24.

```
Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable       Factor1       Factor2 Communality
area           -0.748        -0.512        0.822
precip          0.720        -0.639        0.927
maxtemp        -0.913         0.291        0.919
mintemp        -0.949         0.192        0.937
flow           -0.559        -0.737        0.855
strslope        0.117        -0.821        0.687
wsslope        -0.731        -0.286        0.616


Variance        3.6732        2.0898        5.7630
% Var           0.525         0.299         0.823


Rotated Factor Loadings and Communalities
Varimax Rotation

Variable       Factor1       Factor2 Communality
area            0.243         0.874        0.822
precip         -0.962         0.025        0.927
maxtemp         0.886         0.365        0.919
mintemp         0.849         0.465        0.937
flow           -0.047         0.924        0.855
strslope       -0.618         0.552        0.687
wsslope         0.375         0.689        0.616


Variance        3.0164        2.7466        5.7630
% Var           0.431         0.392         0.823


Factor Score Coefficients

Variable    Factor1    Factor2
area        -0.002      0.319
precip      -0.347      0.108
maxtemp      0.280      0.054
mintemp      0.257      0.096
flow        -0.111      0.368
strslope    -0.277      0.280
wsslope      0.064      0.233
```

Table 9. Factor analysis

From these results, the variables high in loadings on a particular factor would be those which are highly correlated with one another, but which have little or no correlation with the variables loading highly on the other factors. The negative loading variable has a meaning opposite to that of the factor. The size of the loading is an indication of the extent to which the variable correlates with the factor.

Limitations, and Discussion of Results

The limitations with this study can be attributed to the number of gauges used and the variables used to predict stream flow. With more complete data over the state, it would have been able to assemble more gauges for this component of the project. Also, if possible it would have been good to include variables used to describe interception, evapotranspiration, infiltration, interflow, saturated overland flow, and baseflow from groundwater. The rate and areal distribution of rainfall input would have been helpful in establishing the catchment retention.

Other issues with the data collection make the estimation of stream flow difficult. First, there is very high variability in recording stream flow data. The stream flow variable exhibited the highest standard deviation and variation across the year. Second, taking yearly annual averages was a crude method in which to characterize the varying conditions that occur across seasons, months, weeks, and even days. Third, the precipitation and temperature data used in the study needed to be better allocated to the gauge drainage areas (as compared to using the drainage area average for the variable) because of the amount of variability that is present in the entire watershed for the precipitation and temperature data. Overall, the choice of variables to analyze were appropriate based on the success other studies found. In the study the results of the multivariate regression indicated that stream flow could best be estimated using area, stream slope and 30 year annual average maximum temperature. Other data analysis techniques revealed the correlation present between the two temperature variables, flow and area, and precipitation to the two temperature variables.

The last important summary from the tests came from the cluster analysis that grouped the gauge station objects based on similarity. The grouped gauges shared the same ecoregions. Ecoregions are defined as "regions of relative homogeneity in ecological systems or in relationships between organisms and their environments" (Omernik 1987). Omernik (1987) mapped the ecoregions of the conterminous United States, based on regional patterns in individual maps of land use, land surface form, potential natural vegetation, and soils. A discriminant analysis using the ecoregion of each gauge station catchment area would have been a better choice than the using the major river basins used in this study. The similar gauge station catchment areas identified by the cluster analysis and the associated ecoregion borders in West Virginia are displayed in figure 25.

Component 6. The environmental database

An environmental database of point data was included within SWPS. These points are found in the shapefiles directory of SWPS and are loaded for viewing when a user defines a study area location in the state. A brief listing of some of the files in the environmental database follows:

- National pollution discharge elimination system sites
- Landfills
- Superfund sites (CERCLIS)
- Hazardous and solid waste sites (RCRIS)
- Toxic release inventory sites
- Coal dams

- Abandoned mine land locations
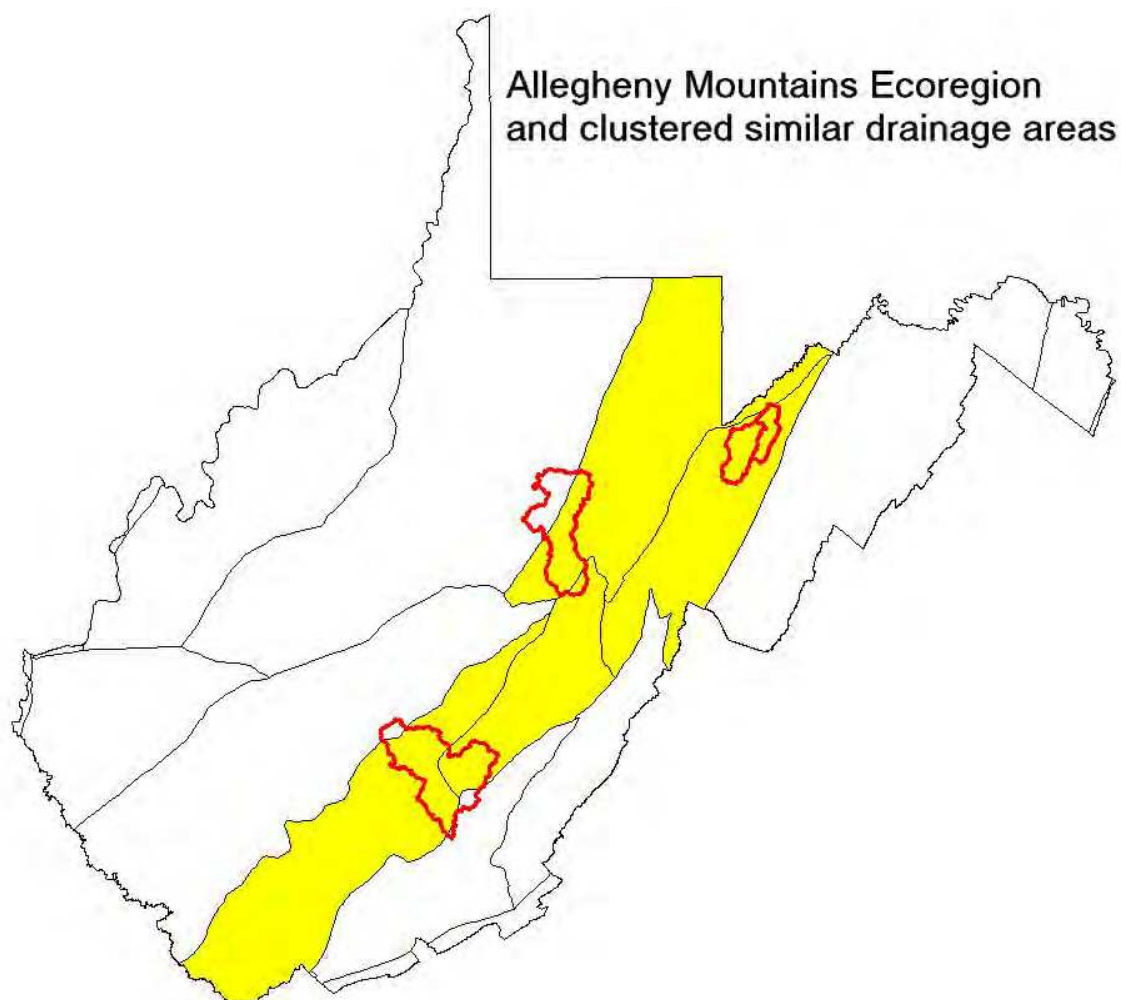- Animal feed lots
- Major highways
- Railroads

Fig. 25.

Component 7. UTM latitude/longitude conversion utility

This capability of SWPS allows the user to map coordinates in degrees, minutes and seconds by using an input dialog screen. The user's points are then mapped in the UTM zone 17 projection. Points may be added to an existing point feature theme or a new point theme can be created. The ability to type coordinates and have the points reprojected saves the user many extra steps. In addition to mapping points from user input, a point can be queried for its x and y locations in UTM, stateplane, or latitude and longitude coordinates. The user can identify locations quickly by clicking anywhere in the display to report this information.

Component 8. Statewide map/GIS data layers

All GIS data is organized in the shapefiles and grids directories of SWPS. This data is listed below. These datasets are provided in addition to the data listed in the environmental database discussed in component 6.

- Coded hydrology
- Watersheds
- Lakes/impounds
- Counties
- 1:24K quads
- Major river basins
- Abandoned mine lands
- Watersheds with fish collection data
- Stream orders
- Major Rivers
- WV GAP land use/ cover
- Bond forfeiture sites
- Expected mean concentration grid
- Public wells
- Groundwater wells
- Surface water wells

- Roads (1:100K scale)
- Cities
- USGS gauging stations
- Public wells
- Runoff grid
- Flow direction
- Flow accumulation
- Cumulative runoff
- Coal Geology
- Override 7Q10 streams
- Landfills
- NW Wetlands
- Springs over 500gpm
- 1950 land use/cover
- Shreve stream orders
- Stream length from mouth

- EPA MRLC land use/cover
- Digital Raster Graphics
- SPOT imagery
- Digital Elevation Model
- Hillshaded relief
- Elevation TIN
- 303-D listed streams
- DRAFT 14 digit HUCS
- Wet weather streams
- Surface mine inventory sites
- Public wells
- Strahler stream orders
- Stream slope
- Max and min stream elevations
- Surface water zones of critical concern
- Streamflow

Component 9. Water quality modeling capability

The water quality modeling capabilities of SWPS are built using a landscape driven approach that uses a predefined runoff and cumulative runoff grid to drive the analysis. It is essentially a weighted mass balance approach that will show changing concentrations and loadings based on changing flow conditions only. The runoff grid is based on a relationship between rainfall and stream flow. It is the main factor that directs flow directions to the stream or steepest path direction and estimates the stream flow.

The assumptions/limitations of this water quality modeling approach are the following:

1. Streams have the same **hydrogeometric properties** (stream slope, roughness, width, and depth).
2. Also assumed are that the streams have the same **ecological rate constants** (reareation rates, pollution decay rates and sediment oxygen demand rate).
3. Transport of pollutants is considered to be conservative (values get averaged over changing flow conditions only) -> no loss or decay of pollutants is considered Does not consider infiltration, or ground water flow additions
5. Does not include atmospheric conditions such as evapotranspiration

The water quality model in SWPS can be used in two different ways. The first is when the user has collected point locations of water quality data and wants to associate the sampled data to instream concentrations and loadings downstream of the sampling points. This is essentially a weighted mass balance approach using the stream flow and sampled locations to associate the point location information to stream condition. The input data using this method needs to be in Mg/L. The resultant modeled levels are reported back as stream values in Mg/L for concentration and Kg/Yr. for loading. The advantage of this first method of using sampled data is that it allows the user to see how the data location information can be used to estimate downstream conditions away from the sampling site.

The second way the water quality model in SWPS can be used is in estimating total nitrogen, phosphorous and total suspended solids as concentrations and loadings in the stream based on expected mean concentrations from land use/cover classes. This method does not require any sampled water quality but uses the cover classes from a land use/cover grid (30meter-cell size). The thirteen classes for West Virginia from this data set were aggregated to six general classes because loading values for nitrogen, phosphorous and total suspended solids were only available for those six classes. The aggregated classes and the corresponding classes included:

- Urban (low intensity developed, high intensity developed, residential)
- Open/Brush (hay, pasture grass, mixed pasture, other grasses)
- Agriculture (row crops)
- Woodland (conifer forest, mixed forest, deciduous forest)
- Barren (quarry areas, barren transitional areas)
- Wetland (emergent and woody wetlands)

The classes are associated with expected loadings based on the acreage size of the class. The loadings are annual averages and when used with the modeled stream flow can give concentration and loading results for the stream. The cover classes and associated expected mean concentrations levels used in the model are shown below.

	Total Nitrogen	Total Phosphorous	Total Suspended Solids
Urban	1.89	0.009	166
Open/Brush	2.19	0.13	70
Agriculture	3.41	0.24	201
Woodland	0.79	0.006	39
Barren	3.90	0.10	2200
Wetland	0.79	0.006	39

The nutrient export coefficients above are multiplied by the amount (area) of a given land cover type. It is used as a simulation to estimate the probability of increased nutrient loads from land cover composition. It should be noted that there are factors other than land cover that contribute to nutrient export and these are rarely known with certainty. Some of the factors that may vary across watersheds and may change the expected mean concentration results include:

• year to year changes in precipitation
• soil type
• slope and slope morphology (convex, concave)
• geology
• cropping practices
• timing of fertilizer application relative to precipitation events
• density of impervious surface

The loading and concentration results in consideration of these assumptions however can still give insight in comparing expected pollutant values for watersheds. The results should be thought of in most cases as the worst case scenarios for stream water quality levels.

Component 10. Groundwater and surface water susceptibility model

The susceptibility ranking model within SWPS was constructed using the steps defined within the West Virginia Surface Water Assessment Program (SWAP) document. The susceptibility ranking for ground water systems was based on the physical integrity of the well and spring infrastructure; hydrologic setting; inventory of potential contaminant sources and land uses, and water quality. The susceptibility ranking for surface water systems was based on water quality and the inventory of potential contaminant sources. A more detailed explanation of the ground and surface water susceptibility models follows.

Groundwater susceptibility model

To determine the groundwater susceptibility for a site, the physical barrier effectiveness is first calculated. Physical barrier effectiveness is the Tier 1 assessment. It is used to note if there is a known impact on water quality, evaluate the source integrity as low or high, and to find the aquifer vulnerability. Based on these results, the physical barrier effectiveness can be determined as having high, moderate, or low potential susceptibility. If there is a known impact on water quality, then the model automatically goes to the Tier Two assessment and sets the groundwater susceptibility as being high. If there is no known impact on water quality then the source integrity and aquifer vulnerability set the physical barrier effectiveness for the Tier Two assessment. The aquifer vulnerability is determined from the different scenarios listed below;

If	**Then**
All springs	High Aquifer Sensitivity
Alluvial Valleys	
Unconfined	High Aquifer Sensitivity
Confined	Moderate Aquifer Sensitivity
Appalachian Plateau Province (fracture)	Moderate Aquifer Sensitivity
Folded Plateau Area (fracture)	Moderate Aquifer Sensitivity
Karst Areas	High Aquifer Sensitivity
Valley and Ridge Province (fracture)	Moderate Aquifer Sensitivity
Coal Mine Areas	High Sensitivity

From the above scenarios, an aquifer vulnerability was determined. Using this with the source integrity rating can provide physical barrier effectiveness. Physical barrier effectiveness is

- High if there is low source integrity and high aquifer sensitivity.
- Low if there is high source integrity and moderate aquifer sensitivity
- Moderate if there is high source integrity and high aquifer sensitivity or low source integrity and moderate aquifer sensitivity

Again, if there is no known impact on water quality, this method will determine the physical barrier effectiveness as being high, moderate or low susceptibility. If there is a known water quality impact then the final groundwater susceptibility is high.

Using the physical barrier effectiveness with the land use concern level determines the final groundwater susceptibility. The land use concern level is determined from the percentage of land use in the buffered groundwater site. The percentage of land use was found for every buffered location. In cases where the groundwater site had a pumping rate over 25,000 gpd, no buffer was created. For these groundwater sites no land use percentage was calculated.

The final groundwater susceptibility is rated as:

- High if the physical barrier effectiveness is high
- High if the land use concern is high and the physical barrier effectiveness is moderate
- Moderate if the land use concern is medium or low and the physical barrier effectiveness is moderate
- Moderate if the land use concern is high or medium and the physical barrier effectiveness is low
- Low if the land use concern is low and the physical barrier effectiveness is low

The percentage of land use is reported to the user before the tier one assessment appears. He or she needs to know the associated concern levels with the percentage of land uses that are reported for each groundwater site. By not hard coding in the land use concern levels for each buffer, the user has the ability to perform "what if" type scenarios if existing land use changes or is different than what currently exists.

Surface water susceptibility model

The surface water susceptibility model is slightly less complicated than the groundwater model. For the surface water susceptibility determination, the percent of land use was calculated for each of the zone of critical concern. The land use concern level, and if there is a known water quality impact, are the two factors which are used to determine the surface water susceptibility. As in the groundwater model, the percent land use is presented to the user before the model is run so the user can make a determination and perform "what if" type scenarios with differing land use within the zone of critical concern. If there is a known water quality impact, then the surface water susceptibility is automatically high. If there is no known water quality impact, then the final surface water susceptibility is:

- High if land use concern level is high
- High if land use concern level is medium
- Low if land use concern level is moderate

Summary and Conclusion

Drinking water is a critical resource that continues to need protection and management to assure safe supplies for the public. Since agencies to protect water resources operate at mostly state jurisdictions, it is important to implement a system at a statewide level. This chapter discussed watershed tools that integrate spatially explicit data and decision support to assist managers with both surface and ground water resources. It has three major components which include; an ability to delineate source water protection areas upstream of supply water, an inventory of potential contamination sources within various zones of critical concern, and the determination of the public drinking water supply systems

susceptibility to contamination. The current system provides the ability to assess, preserve, and protect the states source waters for public drinking.

2. References

Anderson, L. and A. Lepisto. 1998. "Links Between Runoff Generation, Climate and Nitrate-N Leaching From Forested Catchments." Water, Air, and Soil Pollution. 105: pp227-237.

Astatkie, T. and W. E. Watt. 1998. "Multiple-Input Transfer Function Modeling of Daily Stream flow Series Using Nonlinear Inputs." Water Resources Research. Vol. 34, No.10, pp2217-2725.

Environmental Systems Research Institute (ESRI). 2010. ArcInfo ArcMap. Redlands, CA.

Gabriele, H. M., and F. E. Perkins. 1997. "Watershed-Specific Model for Stream flow, Sediment, and Metal Transport." Journal of Environmental Engineering. pp61-69.

Garren, D. C. 1992. "Improved Techniques in Regression-Based Stream flow Volume Forecasting." Journal of Water Resources Planning and Management. Vol. 118, No. 6. pp654-669.

Hocking, R. R. 1996. Methods and Applications of Linear Models. Wiley and Sons, New York, NY.

Johnston, K., J. M. Ver Hoef, K. Krivoruchko, and N. Lucas. 2001. Using ArcGis Geostatistical Analyst. Environmental System Research Institute, Redlands, CA.

Kachigan, S. K. 1986. Statistical Analysis. Radius Press, New York, NY.

Moore, R. D. 1997. "Storage-Outflow Modeling of Stream flow Recessions, with Application to Shallow-Soil Forested Catchments." Journal of Hydrology. 198(1997) pp260-270.

Omernik, J. M. 1987. "Ecoregions of the conterminous United States." Annals

of the Association of American Geographers Vol. 77, pp118-125.

Querner, E. P., Tallaksen, L. M., Kasparek, L., and H. A. J. Van Lanen. 1997. "Impact of Land-Use, Climate Change and Groundwater Abstraction on Stream flow Droughts Using Physically-Based Models." Regional Hydrology: Concepts and Models for Sustainable Water Resource Management. IAHS Publ. No. 246, pp171-179.

Sharma, A., D. G. Tarboton, and U. Lall. 1997. "Stream flow Simulation: A Nonparametric Approach." Water Resources Research. Vol. 33, No. 2, pp291-308.

Smith, J. A. 1991. "Long-Range Stream flow Forecasting Using Nonparametric Regression." Water Resources Bulletin. Vol. 27, No. 1, pp39-46.

Sundberg, R. 2002. Collinearity. Encyclopedia of Environmetrics, Vol. 1. John Wiley and Sons, Ltd, Chichester, West Sussex UK.

Timofeyeva, M. and R. Craig. 1998. "Using Monte Carlo Technique for Modelling the Natural Variability of Stream flow in Headwaters of the Sierra Nevada, USA." Hydrology, Water Resources and Ecology in Headwaters. 1998. No. 248, pp59-65.

U. S. Geological Survey. 1996. "Prediction of Travel Time and Longitudinal Dispersion in Rivers and Streams." Water-Resources Investigations Report 96-4013.

West Virginia Department of Health and Human Resources. August 1, 1999. "State of West Virginia Source Water Assessment and Protection Program." Office of Environmental Health Services, Charleston, WV.
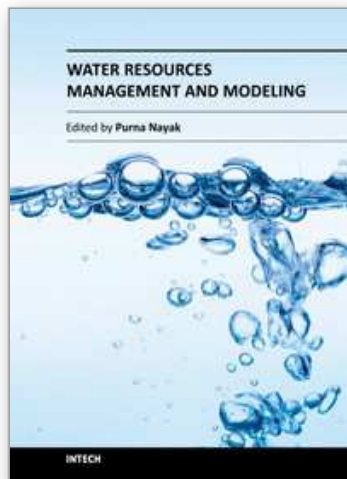
Water Resources Management and Modeling

Edited by Dr. Purna Nayak

Hydrology is the science that deals with the processes governing the depletion and replenishment of water resources of the earth's land areas. The purpose of this book is to put together recent developments on hydrology and water resources engineering. First section covers surface water modeling and second section deals with groundwater modeling. The aim of this book is to focus attention on the management of surface water and groundwater resources. Meeting the challenges and the impact of climate change on water resources is also discussed in the book. Most chapters give insights into the interpretation of field information, development of models, the use of computational models based on analytical and numerical techniques, assessment of model performance and the use of these models for predictive purposes. It is written for the practicing professionals and students, mathematical modelers, hydrogeologists and water resources specialists.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Michael P. Strager (2012). Tools for Watershed Planning - Development of a Statewide Source Water Protection System (SWPS), Water Resources Management and Modeling, Dr. Purna Nayak (Ed.), ISBN: 978-953-51-0246-5, InTech, Available from: http://www.intechopen.com/books/water-resources-management-and-modeling/tools-for-watershed-planning

INTECH
open science | open minds