MAGNITUDE AND FREQUENCY OF FLOODS IN RURAL BASINS OF SOUTH CAROLINA, NORTH CAROLINA, AND GEORGIA

Toby D. Feaster[1], John C. Weaver[2], and Anthony J. Gotvald[3]

AUTHORS: [1]Hydrologist, U.S. Geological Survey, 405 College Ave., Clemson, SC, 29631; [2]Hydrologist, U.S. Geological Survey , 3916 Sunset Ridge Road, Raleigh, NC 27607; and [3]Hydrologist, U.S. Geological Survey , Peachtree Business Center, Suite 130, 3039 Amwiler Road, Atlanta, GA 30360
REFERENCE: *Proceedings of the 2008 South Carolina Water Resources Conference,* held October 14-15, 2008, at the Charleston Area Event Center.

Abstract. A multi-state approach is being used to update methods for estimating the magnitude and frequency of floods in rural ungaged basins in South Carolina, North Carolina, and Georgia that are not significantly affected by regulation, tidal backwater fluctuations, or urban development. Annual peak-flow data through September 2006 were analyzed for 943 streamflow gaging stations having 10 or more years of record on rural streams in South Carolina, North Carolina, Georgia and adjacent parts of Alabama, Florida, Tennessee, and Virginia. Flood-frequency estimates were computed for the 943 gage stations by fitting the logarithms of annual peak flows for each station to a Pearson Type III distribution. As part of the computation of flood-frequency estimates for gaged stations, new values for generalized skew coefficients were developed using a Bayesian generalized least-squares regression model. Additionally, basin characteristics for these gaged stations were computed by using a Graphical Information System and automated computer algorithms.

Regional regression analysis, using generalized least-squares regression, is being used to develop a set of predictive equations that can be used to estimate the 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence interval peak flows for rural ungaged basins in South Carolina, North Carolina, and Georgia. Exploratory regression analyses were done using ordinary least squares regression. Based on results from the exploratory analyses, flood-frequency estimates and basin characteristics for 828 of the 943 streamflow gaging stations were combined to form the final database that was used in the regional regression analysis. The provisional predictive equations are all functions of drainage area and the percentage of drainage basin within each of the five hydrologic regions.

INTRODUCTION

Reliable estimates of the magnitude and frequency of floods are required for the design of transportation and water-conveyance structures such as roads, bridges, culverts, dams, and levees. Federal, State, regional, and local officials also need these estimates for the effective planning and management of land use and water resources, to protect lives and property in flood-prone areas, and to determine flood-insurance rates.

Estimates of the magnitude and frequency of floods are needed at locations where streamflow gaging stations hereafter referred to as gaging or gaged stations continuously monitor streamflow as well as at ungaged sites, which have no streamflow information available for use as a basis in determining the estimates. Estimates of flood magnitude and frequency at ungaged sites is accomplished using a process known as regionalization, in which flood-frequency information determined for a group of gaged stations within a hydrologic region forms the basis for estimates at the ungaged sites. Historically, these hydrologic regions were determined individually by each state, which often led to differences in hydrologic regions at the state boundaries. These differences cause some discontinuity and confusion on which flood-frequency techniques and results are most appropriate for drainage basins near or crossing state boundaries. The U.S. Geological Survey (USGS) South Carolina, North Carolina, and Georgia Water Science Centers are currently cooperating on a rural flood-frequency investigation using a multi-state approach with hydrologic regions that cross state boundaries in order to maintain continuity at those boundaries.

PURPOSE AND SCOPE

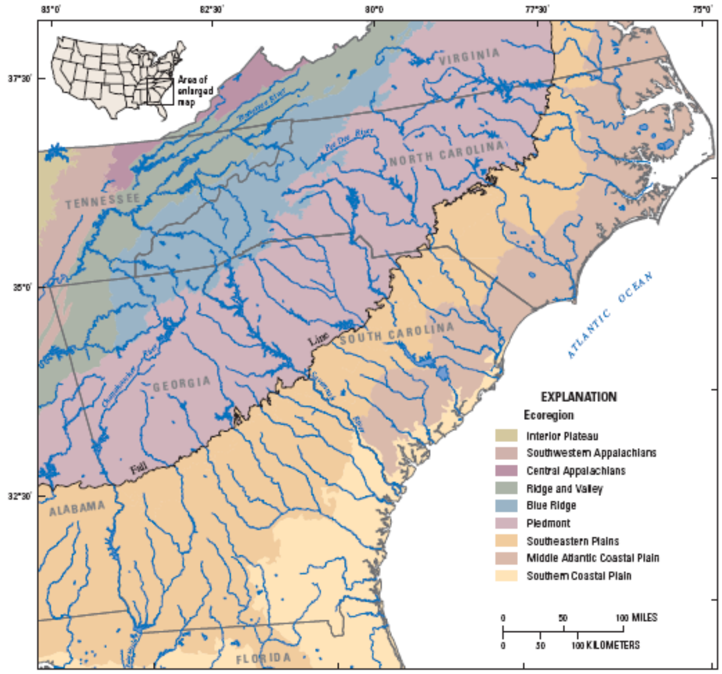
The purpose of this investigation, done in cooperation with the Departments of Transportation in South Carolina, North Carolina, and Georgia, and the North Carolina Floodplain Mapping Program, is to present methods for estimating the magnitude and frequency of floods on rural streams in South Carolina, North Carolina, and Georgia. This investigation is based on flood-frequency analyses of annual peak-flow data at streamflow gaging stations through September 2006. The investigation includes regional equations for estimating the magnitude and frequency of peak flows on rural, ungaged, non-regulated streams in South Carolina, North Carolina, and Georgia; will present estimates of the magnitude of floods at the 2- , 5- , 10- , 25- , 50- , 100- , 200- , and 500-year recurrence intervals for 828 gaging stations in the three states; will describe techniques used to develop regression equations for use in estimating the magnitude of floods for ungaged sites; will describe the accuracy and limitations of the equations; and will present example applications of the methods. The results will be documented using the USGS Scientific Investigations Report series in three separate volumes: North Carolina (volume 1), Georgia (volume 2) and South Carolina (volume 3). The regression techniques will be the same for all three states but the individual volumes also will present information that is unique to each state.

STUDY AREA

The study area includes all of South Carolina, North Carolina, and Georgia within seven U.S. Environmental Protection Agency (USEPA) Level III ecoregions—Southwestern Appalachians, Blue Ridge, Ridge and Valley, Piedmont, Southeastern Plains, Southern Coastal Plain, and Middle Atlantic Coastal Plain (fig. 1; U.S. Environmental Protection Agency, 2007). The ecoregions represent areas of general similarity in ecosystems and were determined from an analysis of the spatial patterns and the composition of biotic and abiotic phenomena that include geology, physiography, vegetation, climate, soils, land use, wildlife, and hydrology (Griffith and others, 2002). The Fall Line separates the higher elevation Southwestern Appalachians, Blue Ridge, Ridge and Valley, and Piedmont ecoregions from the low lying Southeastern Plains, Southern Coastal Plain, and Middle Atlantic Coastal Plain ecoregions.

PEAK-FLOW DATA

Streamflow gaging stations in South Carolina, North Carolina, and Georgia as well as adjacent parts of Alabama, Florida, Tennessee, and Virginia were examined for possible use in this investigation.



Base modified from U.S. Geological Survey 1:100,000-scale digital data
Ecoregions from U.S. Environmental Protection Agency 1:7,500,000-scale digital data (2002; revision of Omernik, J.M., 1987, Ecoregions of the conterminous United States: Annals of the Association of American Geographers 77(1):118–125)

Figure 1. Study area and ecoregions in Georgia, South Carolina, and North Carolina and the surrounding States.

Stations were not used in the analysis if less than 10 years of annual peak-flow data were available, or if peak flows at the stations were substantially affected by dam regulation, flood-retarding reservoirs, tides, or urbanization. The peak-flow records for rural gaged stations that met these criteria were then compiled and reviewed for quality assurance and quality control using techniques described by Feaster and Tasker (2002). The Kendall's tau statistic was chosen to assess the significance of time trends for each station (Helsel and Hirsch, 1992). The quality-assurance and quality-control analysis resulted in the selection of 943 gaging stations that were considered for use in this study. The 943 gaging stations are composed of 82 stations in South Carolina, 333 stations in North Carolina, 357 stations in Georgia, 35 stations in Alabama, 23 stations in Florida, 41 stations in Tennessee, and 72 stations in Virginia.

PHYSICAL AND CLIMATIC BASIN CHARACTERISTICS

Peak-flow information can be transferred to ungaged sites through multiple regression analysis that relates streamflow characteristics (such as the 100-year peak flow) to selected physical and climatic basin characteristics for each gaged drainage basin. Geographical Information System (GIS) techniques were used in development of the drainage-basin boundaries and basin characteristics. Basin characteristics were selected for use as potential explanatory variables in the regression analyses on the basis of their conceptual relation to flood flows and the ability to measure the basin characteristics by using digital datasets and GIS technology. Twenty basin characteristics for each of the 943 rural gaging stations were computed and considered for this study: drainage area, basin perimeter, mean basin slope, basin shape factor, main channel length, main channel slope, minimum basin elevation, maximum basin elevation, mean basin elevation, percentage of basin that is impervious, percentage of basin that is forested, mean annual precipitation, 2-year 24-hour precipitation, 10-year 24-hour precipitation, 25-year 24-hour precipitation, 50-year 24-hour precipitation, 100-year 24-hour precipitation, soil drainage index, hydrologic soil index, and drainage density.

ESTIMATION OF FLOOD MAGNITUDE AND FREQUENCY AT GAGED STATIONS

A frequency analysis of peak-flow data at a gaged station on a stream provides an estimate of the flood magnitude and frequency at that specific station. Flood-frequency estimates are typically presented as a set of exceedance probabilities or, alternatively, recurrence intervals (T years) along with the associated flows. Exceedance probability (1/T years) is defined as the probability of exceeding a specified flow in a 1-year period and is expressed as decimal fractions less than 1.0 or as percentages less than 100. A flow with an exceedance probability of 0.01 has a 1-percent chance of being exceeded in any given year. Recurrence interval is defined as the number of years, on average, during which the specified flow is expected to be exceeded one time. A flow with a 100-year recurrence interval is one that, on average, will be exceeded once every 100 years.

Flood-frequency estimates for gaging stations are computed by fitting the series of annual peak flows to some known statistical distribution. For this study, flood-frequency estimates were computed by fitting logarithms (base 10) of the annual peak flows to a Pearson Type III distribution. This follows the guidelines and computational methods described in Bulletin 17B of the Hydrology Subcommittee of the Interagency Advisory Committee on Water Data (IACWD, 1982). The equation for fitting the log-Pearson Type III distribution to an observed series of water-year peak flows is as follows:

$$\log Q_T = X + KS \tag{1}$$

where

Q_T is the T-year recurrence interval peak flow, in cubic feet per second (ft^3/s);
X is the mean of the logarithms of the annual peak flows;
K is a factor based on the skew coefficient and the given recurrence interval, which can be obtained from Appendix 3 in Bulletin 17B; and
S is the standard deviation of the logarithms of the annual peak flows, which is a measure of the degree of variation of the annual values about the mean value.

A water year is the 12-month period from October 1 to September 30 and is designated by the calendar year in which it ends.

A generalized skew coefficient analysis was performed as part of this investigation. The initial dataset for the generalized skew coefficient analysis included 489 of the 943 gaging stations each having 25 or more years of annual peak-flow record. The station skew coefficients for these 489 gaging stations were plotted at the centroid of the watershed for each station in order to develop a skew isoline map. The map was then reviewed to determine if any geographic or topographic trends were apparent. No clearly definable patterns were found, so regression techniques were then used to develop a skew prediction equation using station skew as the response variable. A Bayesian Generalized Least Squares (GLS) regression model was used for the generalized skew coefficient regression analysis as suggested by Reis and others (2005) and Gruber and others (2007), both of which made use of the methods proposed by Martins and Stedinger (2002). This study also uses the methods from Martins and Stedinger (2002) except that this study used the distance between basin centroids instead of the distance between gaging stations.

Based upon the Bayesian GLS regression analysis, a single generalized skew value of -0.0186 was determined. More complicated Bayesian GLS models with additional explanatory variables were evaluated, but resulted in very modest improvements in accuracy. These modest improvements did not seem justified due to the increased complexity associated with the addition of explanatory variables. The mean square error (MSE) associated of the constant generalized-skew model is 0.083. This MSE is equivalent to 69 years of record length. This is a significant improvement over the Bulletin 17B skew map value with a mean square error of 0.302, which is equivalent to 17 years of record. Therefore, the generalized skew value of -0.0186 was used to compute the flood-frequency estimates for the 943 gaged stations according to methods recommended in Bulletin 17B.

ESTIMATION OF FLOOD MAGNITUDE AND FREQUENCY AT UNGAGED STATIONS

Regional regression analysis was used to develop a set of equations for use in estimating the magnitude and frequency of floods for rural, ungaged sites in South Carolina, North Carolina, and Georgia. These equations relate the 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year flood discharges computed from available records for gaged stations to measured physical and climatic characteristics of the drainage basins for the gaged stations.

Exploratory Regional Regressions. Ordinary least-squares (OLS) regression techniques were used in the exploratory analysis to determine the best regression models for all combinations of the basin characteristics as well as in the development of hydrologic regions that define the study area. Selection of the explanatory variables for each hydrologic region was based on all-possible-subsets (APS) regression methods (Neter and others, 1985). The final explanatory variables for each hydrologic region were selected on the basis of several factors, including: standard error of the estimate, Mallow's Cp statistic, statistical significance of the explanatory variables, coefficient of determination (r^2), and ease of measurement of explanatory variables. Multicollinearity in the candidate exploratory variables also was assessed by the variance inflation factor (VIF) and the correlation between explanatory variables. Regression residuals were plotted at the centroid of their respected drainage basin in order to determine geographical patterns of bias. The USEPA Level III ecoregions (fig. 1) were used as the initial hydrologic regions for the regression analysis. For each ecoregion, stations that had at least 75 percent or more of the basin area in one ecoregion were grouped together. The APS regression methods were then conducted on each of the six groups of stations in order to determine the candidate explanatory variables for each ecoregion. The APS analysis found that the addition of other variables with drainage area did not reduce the standard error of estimate by more than 3 percent. This small reduction did not warrant the use of additional explanatory variables in the model, so drainage area was selected as the only basin characteristic used for further analysis. Based on the OLS analysis, the USEPA ecoregions were combined as warranted obtaining 5 hydrologic regions that were used to develop the regional flood-frequency regression equations (fig. 2). As a result of the OLS analysis, 828 gaged stations were used to develop regional regression equations.

Regional Regression Equations. When grouping stations into hydrologic regions, stations that drain from more than one region are typically omitted from the regional regression analysis. For this investigation, 83 of the 828 gaged stations have significant drainage from more than one region. Griffis and Stedinger (2007) showed that a pooled regression model approach that combines data across regions increases the number of stations and information available for model estimation. A similar approach was used by

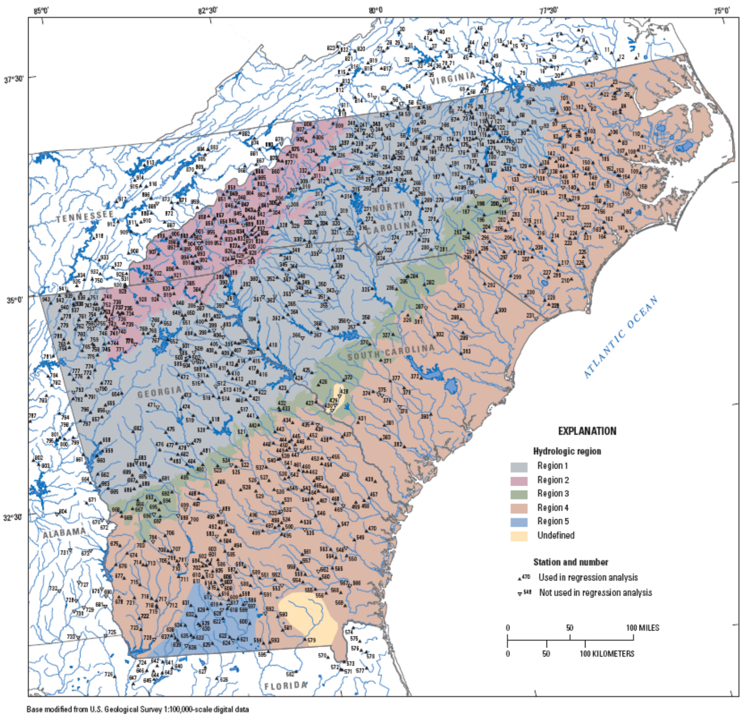


Figure 2. Hydrologic regions in the study area as well as the location of the streamflow gaging stations used in the regional regression analysis.

Feaster and Tasker (2002) to distinguish between states for two of the hydrologic regions included in their investigation. In order to incorporate stations draining from multiple regions in the regional regression analysis, a modified version of Griffis and Stedinger's pooled regional regression approach was used for this study.

This modified approach uses the percentage of the basin within the hydrologic regions as explanatory variables instead of the qualitative indicator variables. With this approach, all the stations in the study region are pooled in the regional regression analysis, so no stations are omitted due to multiple regions within a basin. The hydrologic model for this modified approach has the following linear form for *n* number of hydrologic regions:

$$\textbf{Log } Q_T = a_1(PCT_1) + a2(PCT_2) + \ldots + a_n(PCT_n) + b_0(\textbf{log } DA) \qquad (2)$$

where

a_1, a_2 , ... a_n, and b_0 are regression coefficients;
PCT_1, PCT_2,…, and PCT_n are the basin percentages in hydrologic regions *1*, *2*, …., and *n*, in percent; and
DA is drainage area, in square miles.

Using this model assumes that the slope or b_0 coefficient is constant for every region. In order to test the significance of the differences of slopes for each region, cross products of the explanatory variables were added to the equation. For *n* number of hydrologic regions, the hydrologic model has the following linear form:

$$\textbf{Log } Q_T = a_1(PCT_1) + a_2(PCT_2) + \ldots a_n(PCT_n) + b_0(\textbf{log } DA) + b_1 [(\textbf{log } DA)(PCT_1)] + b_2 [(\textbf{log } DA)(PCT_2)] + \ldots b_n[(\textbf{log } DA)(PCT_n)] \qquad (3)$$

where

b_1, b_2 , … b_n are regression coefficients; and other variables are previously defined.

To test the significance of the slope difference for each region, the coefficient of one of the cross products is set to zero, and the significance of the differences in slopes (b_1, b_2 , ... b_n coefficients) are tested at the 95-percent probability level.

Generalized least-squares (GLS) regression methods, as described by Stedinger and Tasker (1985), are being used to determine the final regional T-year peak-flow regression equations, using the USGS computer program GLSNET (G.D. Tasker, K.M. Flynn, A.M. Lumb, and W.O. Thomas, Jr., written commun., 1995). The slope differences for hydrologic regions 2 and 3 were found to be significant at the 95-percent probability level for the Q_2 through Q_{500} discharges and therefore, a slope adjustment factor was included in the regression equations for hydrologic regions 2 and 3. The results of the analyses are a single T-year peak-flow estimate equation for the entire study area, which is composed of 5 hydrologic regions (fig. 2). The final form of the regional regression equations for the 2- through 500- year recurrence interval peak flows are expected to be as follows:

$$\textbf{Log } Q_T = a_1(PCT_1) + a_2(PCT_2) + a_3(PCT_3) + a_4(PCT_4) + a_5(PCT_5) + b_0 \textbf{ log } DA + b_1 [(\textbf{log } DA)(PCT_2)] + b_2 [(\textbf{log } DA)(PCT_3)] \qquad (4)$$

All variables are previously defined.

LITERATURE CITED

Feaster, T.D., and Tasker, G.D., 2002, Techniques for estimating the magnitude and frequency of floods in rural basins of South Carolina, 1999: U.S. Geological Survey Water Resources Investigations Report 02-4140, 34p.

Griffith, G.E., Omernik, J.M., Comstock, J.A., Schafale, M.P., McNab, W.H., Lenat, D.R., MacPherson, T.F., Glover, J.B., and Shelburne, V.B., 2002, Ecoregions of North Carolina and South Carolina. Map (scale 1:1,500,00). Reston, Virginia, U.S. Geological Survey.

Griffis, V.W., and Stedinger, J.R., 2007, The use of GLS regression in regional hydrologic analyses: Journal of Hydrology, v. 344, p. 82-95.

Gruber, A. M., Reis, D.S., and Stedinger, J.R., 2007, Models of Regional Skew Based on Bayesian GLS Regression, Paper 40927-3285, World Environmental & Water Resources Conference - Restoring our Natural Habitat, K.C. Kabbes editor, Tampa, Florida.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier, p. 326.

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency : Hydrology Subcommittee Bulletin 17B, 28p., 14 appendixes, 1 plate.

Martins, E.S., and Stedinger, J.R., 2002, Cross-correlation among estimators of shape, Water Resources Research, 38(11), doi: 10.1029/2002WR001589, 26.

Neter, John, Wasserman, William, and Kutner, M.H., 1985, Applied linear statistical models: Homewood, Illinois, Richard D. Irwin, 1,127 p.

Reis, D. S., Jr., J. R. Stedinger, and E. S. Martins, 2005, Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation, Water Resources Research, 41, W10419, doi:10.1029/2004WR003445.

Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis, 1, ordinary, weighted and generalized least squares compared: Water Resources Research, v. 21, no. 9, p. 1,421-1,432. [with correction, 1986, Water Resources Research, vol. 22, no. 5, p. 844.]

U.S. Environmental Protection Agency, 2007, Level 3 Ecoregions of the United States: accessed on July 24, 2008 at http://www.epa.gov/bioindicators/html/lv3-eco.html.