JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
FEBRUARY AMERICAN WATER RESOURCES ASSOCIATION 2005

AN ANALYSIS OF REGION-OF-INFLUENCE METHODS FOR FLOOD REGIONALIZATION IN THE GULF-ATLANTIC ROLLING PLAINS[1]

Ken Eng, Gary D. Tasker, and P.C.D. Milly[2]

ABSTRACT: Region-of-influence (RoI) approaches for estimating streamflow characteristics at ungaged sites were applied and evaluated in a case study of the 50-year peak discharge in the Gulf-Atlantic Rolling Plains of the southeastern United States. Linear regression against basin characteristics was performed for each ungaged site considered based on data from a region of influence containing the n closest gages in predictor variable (PRoI) or geographic (GRoI) space. Augmentation of this count based cutoff by a distance based cutoff also was considered. Prediction errors were evaluated for an independent (split-sampled) dataset. For the dataset and metrics considered here: (1) for either PRoI or GRoI, optimal results were found when the simpler count based cutoff, rather than the distance augmented cutoff, was used; (2) GRoI produced lower error than PRoI when applied indiscriminately over the entire study region; (3) PRoI performance improved considerably when RoI was restricted to predefined geographic subregions.

(KEY TERMS: floods; regional regression; region of influence; statistical analysis; surface water hydrology; streamflow characteristics.)

Eng, Ken, Gary D. Tasker, and P.C.D. Milly, 2005. An Analysis of Region-of-Influence Methods for Flood Regionalization in the Gulf-Atlantic Rolling Plains. Journal of the American Water Resources Association (JAWRA) 41(1):135-143.

INTRODUCTION

Hydrologists often need to estimate flood frequencies at sites with little or no flow information. One approach to calculate streamflow characteristics at ungaged sites is to use regional regression models. These models relate the basin characteristics, such as drainage area, to flows, such as the 50-year return peak discharge observed at a network of gaging stations in a region of interest. The 50-year return peak discharge is the annual peak flow that is expected to be exceeded on average in 1 out of 50 years, which is equivalent to the 0.2 percentile of the distribution of annual peak streamflows. Commonly, a region of interest is delimited by state boundaries or by physiographic boundaries within states. Such regions also can be identified by use of a variety of criteria, such as residuals from an overall regression (Wandle, 1977) or watershed boundaries (Neely, 1986). An alternative to defining subregions in geographic space is to define them in predictor variable space; the regression is performed on a subset of stations for which the basin characteristics are, by some overall measure, closest to those at the ungaged site of interest. In the typical application of this "region-of-influence" approach, furthermore, a unique "region" is defined for each ungaged site (Burn, 1990; Tasker and Slade, 1994; Tasker *et al.,* 1996; Pope *et al.,* 2001; Berenbrock, 2002; Feaster and Tasker, 2002). Until now, the region-of-influence approach has been applied to subdivisions contained within the boundaries of a State (i.e., Feaster and Tasker, 2002). This practice of performing a unique regression for each ungaged site of interest, however, need not be restricted to approaches that use predictor variable space to define hydrologic similarity. This practice can just as readily be applied to approaches where the domain is defined purely on a geographic basis. Herein, the term "region of influence" (RoI) is used generally to refer to any approach where a unique regression is developed for each ungaged site based on proximal gage sites. Further such approaches are qualified as predictor variable (PRoI) or geographic (GRoI), depending on which

[1]Paper No. 03173 of the *Journal of the American Water Resources Association* (JAWRA) (Copyright © 2005). **Discussions are open until August 1, 2005.**
[2]Respectively, Research Hydrologists, National Research Program, U.S. Geological Survey, 12201 Sunrise Valley Drive, Mail Stop 430, Reston, Virginia 20192; and Research Hydrologist, National Research Program, U.S. Geological Survey, Geophysical Fluid Dynamics Laboratory/NOAA, Route 1, Forrestal Campus, Princeton, New Jersey 08542 (E-Mail/Eng: Keng@usgs.gov).

space (predictor variable or geographic) is used to establish proximity.

When implementation of these general concepts is considered, a chain of questions naturally arises.

1. How should the region of influence be defined in practice? Burn (1990) defines the region of influence as containing a set of all stations closer than R (defined only in predictor variable space). In various more recent studies (i.e., Tasker and Slade, 1994; Tasker *et al.,* 1996; Pope *et al.,* 2001; Berenbrock, 2002; Feaster and Tasker, 2002), the RoI was defined as a set containing the n closest stations. In the most general case considered here, the RoI is defined as the set of all stations closer than a distance R (in predictor variable or geographic space) from the site or, if the number of stations in that set is smaller than some minimum allowable number n, the n closest stations. The inclusion of a minimum number of stations is motivated by the need for a sufficient number of degrees of freedom and for statistical robustness in the regression process. Note that the definition of the RoI reduces to that of Burn (1990) when n is set to 0 and to that of subsequent investigators when R is set to 0.

2. How might performance of the RoI approach depend on the values of R and n chosen? If either R or n is too large, hydrologically dissimilar stations will be used in the regression and the approach accuracy will decrease. If both R and n are too small, the dataset will be too small to perform a regression. An optimum pair of values must be located between these two extremes.

3. How can the optimum (R, n) be determined? Here, the optimum is defined as that set that yields the minimum error of prediction within an exhaustive search based on numerous combinations of these variables in conjunction with a test dataset. This is a generalized analog of the approach used in previous studies (i.e., Tasker and Slade, 1994; Tasker *et al.,* 1996; Pope *et al.,* 2001; Berenbrock, 2002; Feaster and Tasker, 2002), where R was set to zero *a priori* and n alone was optimized.

4. Can any generalizations be made about the optimal values of (R, n)? Undoubtedly, the results will depend on the dataset used. In the past studies, where R was held to 0, optimal values of n typically were in the range of 10 to 20.

5. Is either GRoI or PRoI superior to the other?

6. How does performance of PRoI depend upon the spatial scale of the analysis? (The spatial scale of GRoI is internalized within the approach and an analogous question does not arise.)

The purpose of this paper is to address Questions 4, 5, and 6.

STUDY AREA AND DATA

Estimates of the 50-year peak discharge (Q_{50}) and basin characteristics at 1,091 streamflow gaging stations in the southeastern United States were used in the regional regression models considered in this study (Figure 1). The period of records for the 1,091 gaged sites ranged from 10 to 103 years. These stations were selected because they were contained within the boundaries of a single physiographic region, the Gulf-Atlantic Rolling Plains (Hammond, 1964). Q_{50} values were estimated by the standard methods described in Bulletin 17B of the Hydrology Subcommittee of the Interagency Advisory Committee on Water Data (1982). Eight basin characteristics were available for this study: drainage area, main channel slope, main stream channel length, mean basin elevation, forested area, area of surface water bodies, mean minimum January temperature, and mean annual precipitation. A best subsets regression analysis was used to identify drainage area (A), main channel slope (S), and mean annual precipitation (P) as the most significant predictor variables. The A and S values were estimated from U.S. Geological Survey 1:24,000 scale topographic maps. S was calculated as the average channel slope (elevation difference divided by distance along the main channel) between points located 10 and 85 percent of the distance from the gaging station to the basin divide. Isohyetal maps (NOAA, 1976-1978) were used to obtain P.

METHODOLOGY

The regression model was based on a power law relation between the 50-year peak discharge (m^3/s) and the various basin characteristics given as

$$Q_{50} = 10^{b_0} A^{b_1} S^{b_2} P^{b_3} 10^{\varepsilon} \tag{1}$$

where ε is the random error (mean equal to zero and variance equal to σ^2). This model permits use of linear equations based on the logarithms (base 10) for the regression

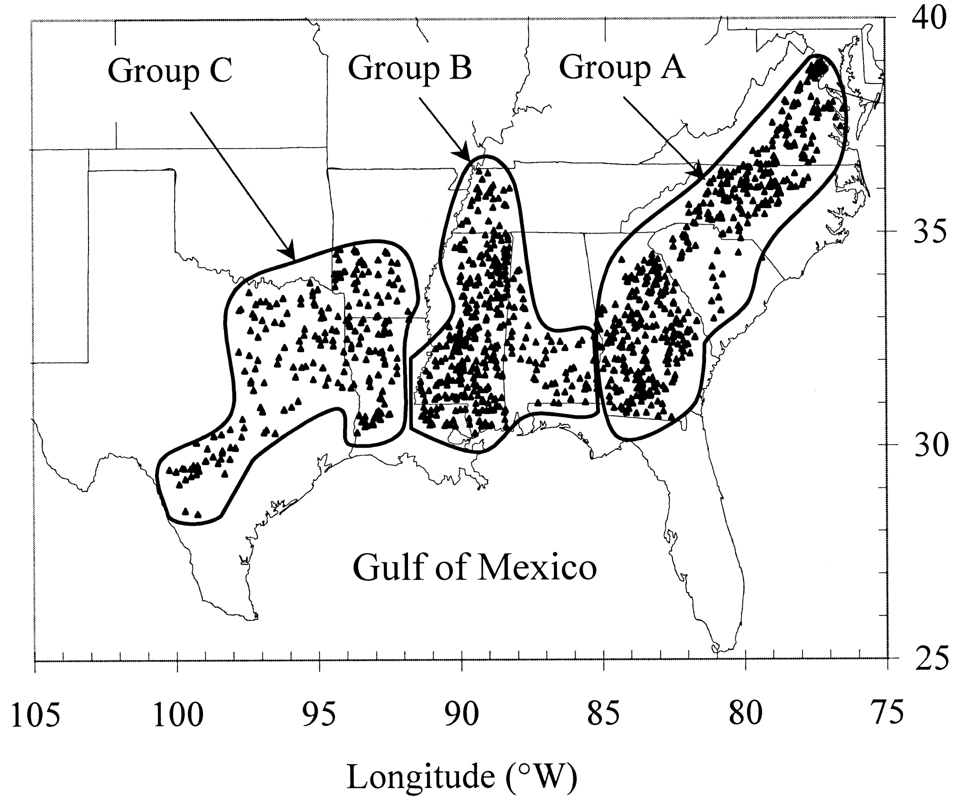

Figure 1. Southeastern United States. Triangles represent individual gaged sites. Groups A, B, and C are used for PRoI(0s) and RoI(∞s) analyses.

$$\log Q_{50} = b_0 + b_1 \log(A) + b_2 \log(S) + b_3 \log(P) + \varepsilon \quad (2)$$

where b_0, b_1, b_2, and b_3 are the constants to be determined with the regression. The variance inflation factor (*VIF*) is a metric that can be used to check for multicollinearity among predictor variables. The *VIF* is given as

$$VIF = \left(1 - R_{VIF}^2\right)^{-1} \quad (3)$$

where R_{VIF}^2 is the coefficient of determination obtained when the predictor variable of interest is regressed on the remaining predictor variables. A high correlation among the predictor variable of interest and the other variables indicate a large R_{VIF}^2 value resulting in a large *VIF* value, and vice versa for low correlations. A value of the *VIF* greater than 5 to 10 would be indicative of significant multicollinearity (Montgomery *et al.,* 2001). The average *VIF* values for $\log(A)$, $\log(S)$, and $\log(P)$ were 4.6, 4.6, and 1.2, respectively, so the multicollinearity was determined not to be significant.

For any site where streamflow characteristics were to be estimated, the constants in Equation (1) were determined by regression over all stations within the RoI of the site. The RoI was defined in terms of *R* and n as follows. The RoI includes the larger of two sets: (1) all stations at a distance smaller than *R* from the estimation site or (2) the n closest stations. For GRoI, distance was defined simply as geographic distance. For PRoI, the distance from an estimation site *i* to a station *j* was defined as

$$R_{ij} = \left[\sum_{m=1}^{M} \left(\frac{x_{im} - x_{jm}}{\sigma_m}\right)^2\right]^{1/2} \quad (4)$$

where $M = 3$ is the total number of basin characteristics being analyzed, x_{im} and x_{jm} are the values of the logarithm of basin characteristics m at sites i and j, and σ_m is the sample standard deviation (over the entire dataset) of the logarithm of basin characteristic m. The selection of *R* and n is explained below.

The log space metrics for performance evaluation were the root mean square error (RMSE) (Aitchison and Brown, 1957), expressed, in percent, as

$$RMSE = 100\left[(20.074 \times 10)^{\sigma_\varepsilon^2} - 1\right]^{1/2} \quad (5)$$

where

$$\sigma_\varepsilon = \left[\frac{\sum_{i=1}^{N} \left(\log Q_{50,i} - \log \hat{Q}_{50,i} \right)^2}{N} \right]^{1/2} \tag{6}$$

and the average error (*BIAS*) as

$$BIAS = \left[\frac{\sum_{i=1}^{N} \left(\log Q_{50,i} - \log \hat{Q}_{50,i} \right)}{N} \right] \tag{7}$$

where $Q_{50,i}$ is the estimated 50-year peak flow at site i based on streamflow records, $\hat{Q}_{50,i}$ is the regional regression estimate of the 50-year peak flow at site i, and N is the total number of sites in either the optimization or evaluation dataset.

Performance was evaluated by split sampling (Snee, 1977). The set of 1,091 stations was split into three equally sized subsets with similar statistical properties (Figure 2). Two of the three subsets then were combined and used in an optimization step to calculate *RMSE* values for various combinations of *R* and n, and the lowest resulting values of *RMSE* and the corresponding *R* and n values were noted. The third subset then was used to evaluate model performance by calculating the *RMSE* value associated with the optimal *R* and n determined in the previous step. All three possible combinations of subsets for this optimization evaluation procedure were applied, and an overall *RMSE* value then was computed as the root mean square value of the three individual values.

Some limiting cases of PRoI and GRoI also were considered, consistent with previous implementations of PRoI in the literature. Specifically, the PRoI(0) and GRoI(0) approaches were defined as special cases of PRoI and GRoI where the RoI always was defined simply to include the n closest stations, and these two cases were represented on Figures 3 and 4 as *R* equal to zero. Optimization and performance evaluation were carried out as described previously, with the exception that *R* was constrained to be zero.

A single, common limiting case of both PRoI and GRoI, here denoted as RoI(∞), was obtained when *R* was made arbitrarily large. In this case, all stations fall within the RoI of any ungaged site, so a single regression was formed on the basis of all data (for each optimization evaluation dataset). No optimization was done because the value of n was irrelevant when *R* was infinite.

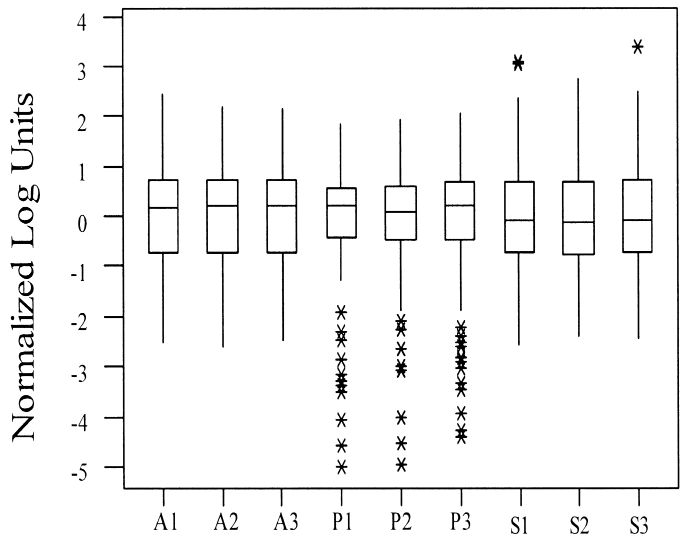

Figure 2. Box and Whisker Plots of the Three Split Sampled Data Subsets. For each box the middle horizontal lines represents the median, the upper line is the 75th percentile, and the lower line is the 25th percentile. The '*' represent the outliers. Each value was subtracted from the mean and normalized by the standard deviation.

Because the available predictor variables were not a complete description of the hydrologic character of a basin, it should not be expected that the PRoI approaches will perform well when applied over an area that is too large. For example, basin mean annual precipitation may not give a complete picture of the climatic conditions conducive to flood generation in the basin. For this reason, the PRoI(0) approach was applied not only to the entire study area, but also separately to each of three groups of gages associated with three subregions (Figure 1), with *RMSE* results aggregated to the entire region. This "subregional" variation of PRoI(0) was denoted by PRoI(0s). For comparison, the RoI(∞) approach also was carried out on the basis of the same subregions, and this case was denoted by RoI(∞s). An additional special analysis for the State of Georgia extended this analysis to a scale even smaller than the subregions. This analysis is explained in the RESULTS section.

Table 1 is presented to summarize all the different RoI approaches examined in this study.

RESULTS

Examples of the dependence of *RMSE* on *R* and n during the optimization step are shown in Figures 3

TABLE 1. Summary of Region-of-Influence Approaches Analyzed in This Study.

Approach	Set From Which Stations Are Selected	Distance Measure, *Rij*	Stations Selected From Set, as a Function of *R* and n	Approach
PRoI	All in study area	$\left[\sum_{m=1}^{M}\left(\frac{x_{im}-x_{jm}}{\sigma_m}\right)^2\right]^{1/2}$	All with $R_{ij} < R$, plus as many additional stations, closest first, as needed to obtain at least n stations	PRoI
PRoI(0)			Same as for PRoI. Because $R = 0$, this simply results in selection of the n closest stations	PRoI
PRoI(0s)	All in subregion of study area			PRoI(0s)
GRoI	All in study area	Ordinary Geographic Distance	All with $R_{ij} < R$, plus as many additional stations, closest first, as needed to obtain at least n stations	GRoI
GRoI(0)			Same as for GRoI. Because $R = 0$, this simply results in selection of the n closest stations	GRoI(0)
RoI(∞)	All in study area	Either of above (makes no difference)	All stations in set, because $R = \infty$	RoI(∞)
RoI(∞s)	All in subregion of study area		All stations in set, because $R = \infty$	RoI(∞s)

Notes: A parenthetic 0 or ∞ in the identifier of the approach refers to the cutoff value R. Note that there is no GRoI(0s), because the use of geographic distance to define the RoI obviates the need for subregionalization. The choice of distance measure is irrelevant for RoI(∞) and RoI(∞s), because with either measure, all stations in the set will be included, since R_{ij} is always finite, regardless of the definition of R_{ij}. In the rule for selection of station, "closest" means closest with respect to the distance measure in use, predictor variable (for the PRoI approaches) or geographic (for GRoI approaches).

and 4. Generally, such surfaces were smooth, and an unambiguous global minimum was found easily. For the PRoI example (Figure 3), *RMSE* was largest when R was large and smallest when R was zero and n was 20. For the GRoI example (Figure 4), the minimum error was found when n was 10 and R was 100 km.

In general, the PRoI approach and its PRoI(0) variation gave identical results (Table 2). That is, the PRoI optimum in the optimization step always was found with R equal to zero. Thus, no additional benefit was obtained by introducing the R parameter. In contrast, the GRoI approach generally produced best results in the optimization step when both R and n were nonzero. Counterintuitively, however, the GRoI(0) approach produced better overall results in the evaluation step than the GRoI approach (Figure 5).

Among the five approaches where predefined subregions were not used to produce the regression equations, the lowest RMSE was produced with GRoI(0) (Figure 5), followed by the general GRoI approach, the (equivalent) PRoI and PRoI(0) approaches, and the RoI(∞) approach. This ordering also was found when *RMSE* was evaluated separately across the three subregions. In the two subregional regression approaches, results were better than in the corresponding whole region approaches. RoI(∞s) results were approximately equal to the best results (GRoI) from the whole region approaches. PRoI(0s) results were improved over PRoI(0), but not as good as those from GRoI or RoI(∞s). In Figures 5 and 6, the bars for RoI(∞) and PRoI in Groups A, B, and C represent the errors for the stations contained in the corresponding group.

The whole region approaches were unbiased regionally (Figure 6). On a subregional basis, the lowest *BIAS* among the whole region approaches was associated with GRoI, followed by GRoI(0), the (equivalent) PRoI and PRoI(0) approaches, and the RoI(∞) approach. The subregional approaches were unbiased over individual subregions.

Until recently, PRoI only has been applied on physiographic regions contained within the boundary of a State. The performance of the PRoI approaches depends on the geographic scope of the analysis (Figures 5 and 6). To explore this dependence further and to provide a basis for comparison with past analyses (e.g., Tasker *et al.,* 1996), the PRoI(0) approach was applied also using only data from the portion of the Gulf-Atlantic Rolling Plains physiographic region in Georgia (one of the states with an especially high density of gaging stations). The resulting *RMSE* and *BIAS* were compared with those (for the same physiographic region in Georgia) from the same approach applied on subregions and the entire region. Results also were compared with those for the RoI(0) approach applied only to the Georgia gaging stations (Figures 7 and 8). Application of PRoI(0) over the

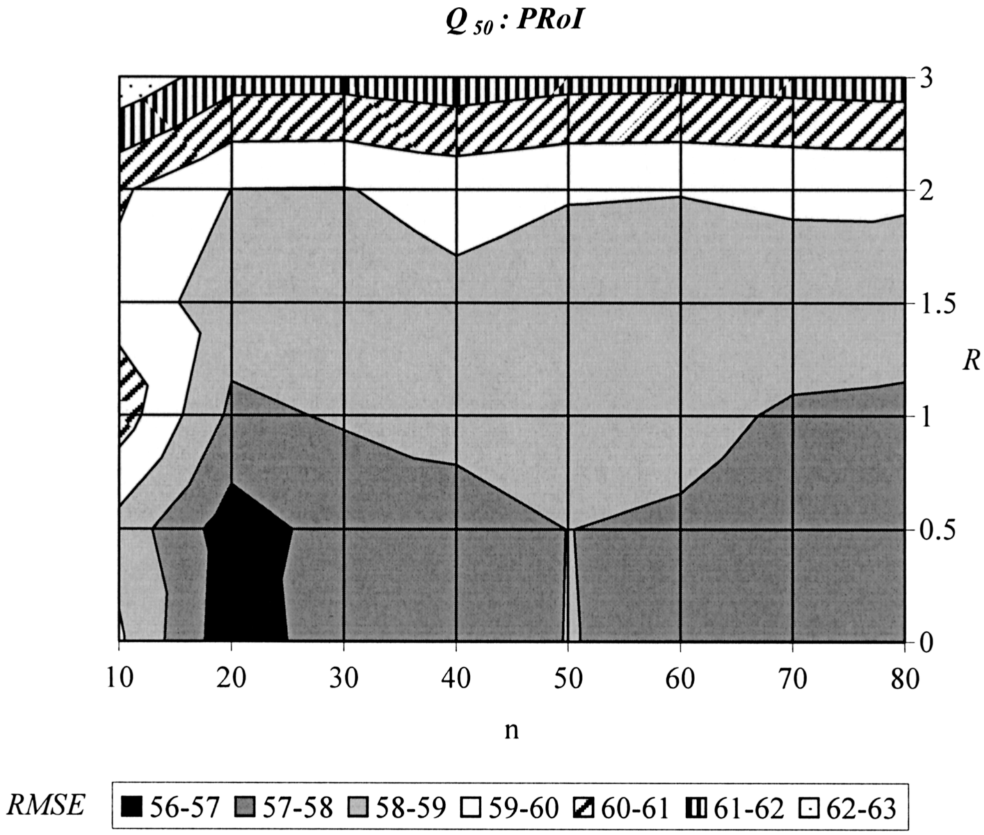

Figure 3. Optimization Set *RMSE*, in Percent, as a Function of *R* and n for the PRoI Approach, for One of the Three Optimization Evaluation Sets. The combination of *R* and n that result in the lowest *RMSE* values are associated with the black region. Results are interpolated from computations with *R* and n increments of 0.25 and 5, respectively.

portion of the physiographic region within Georgia produced the lowest RMSE and BIAS, comparable to those from larger study areas such as RoI(∞s). These results indicate that PRoI approach performs better when optimization is limited to small geographic scales, such as the size of a state, than when it is not. This is consistent with satisfactory performance having been found in other studies when optimization was limited to physiographic provinces within states.

DISCUSSION

When the simple RoI(∞) approach was applied over subjectively chosen groups of States within the Gulf-Atlantic Rolling Plains, the resulting accuracy essentially equaled or exceeded that of all other approaches applied to the regional dataset. However, the GRoI approach may provide a competitive, objective alternative to the subjective definition of subregions for regression. When applied to the regional dataset, the GRoI approach performed about as well as the RoI(∞s) approach.

The GRoI approach is an automated way of defining geographic subregions. The estimated streamflow characteristics have spatial correlation as a result of correlation in the underlying basin and climate characteristics. This correlation is exploited in the GRoI approach.

The PRoI approach potentially is more powerful than the GRoI approach because the PRoI approach makes more explicit the dependence of streamflow characteristics on basin and climatic characteristics. If the full set of relevant basin and climatic characteristics could be known and measured, then PRoI might outperform GRoI on a regional scale. The set of basin and climatic characteristics considered here was limited, and it can be expected that performance of PRoI may improve relative to that of GRoI if a more complete set of predictands is used. These results should not be seen as a failure of the PRoI approach. Results from the present study may be considered a baseline to use for comparison of future results.

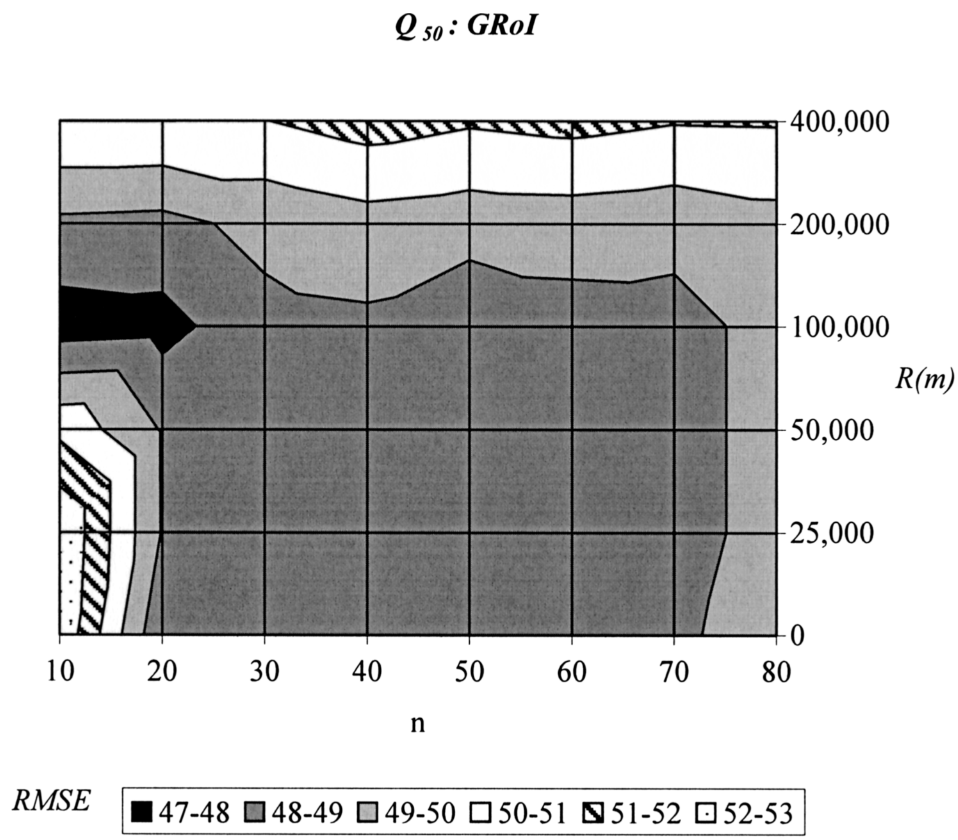

Figure 4. Optimization Set *RMSE*, in Percent, as a Function of *R* and n for the GRoI Approach, for One of the Three Optimization Evaluation Sets. The combination of *R* and n that result in the lowest *RMSE* values are associated with the black region. Results are interpolated from computations with *R* and n increments of 25,000 meters and 5, respectively.

TABLE 2. Optimized Values of n and *R*.

Approach	n	*R*
PRoI	20, 10, 15	0, 0, 0
GRoI	20, 20, 10	100, 100, 100
PRoI(0)	20, 10, 15	(0, 0, 0)
GRoI(0)	45, 45, 45	(0, 0, 0)
PRoI(0s) – Subregion (Group) A	50, 50, 60	(0, 0, 0)
PRoI(0s) – Subregion (Group) B	50, 50, 50	(0, 0, 0)
PRoI(0s) – Subregion (Group) C	60, 30, 10	(0, 0, 0)

Notes: In each entry, the three values pertain to estimation prediction sets 1, 2, and 3, respectively. Parenthetic values are assigned *a priori* with the approach. For GRoI, *R* is given in km; PRoI, *R* is dimensionless.

An attempted generalization of the PRoI approach to consider not only using the closest n gaging stations, but also seeking to maintain a constant value of similarity (as measured by *R*) across all basins, failed to produce any improvement in performance when the approach was applied over the whole study region. That is, the PRoI approach, as defined here, performed no better than the PRoI(0) approach, which was equivalent to the implementations of Tasker and Slade (1994), Tasker *et al.* (1996), Pope *et al.* (2001), and Feaster and Tasker (2002).

Analogously, the more general GRoI approach introduced here can be considered inferior to the simpler GRoI(0) approach. When selecting stations for regression on the basis of geographic distance, it appeared advantageous simply to choose the n closest stations, rather than to consider any expansion of the estimation set when nearby stations are available.

Results presented here may be biased by the neglect of cross-correlation of flow records among nearby stations. This model bias could result in deceptively low error statistics. The size of this bias might tend to be larger for GRoI than for PRoI, because stations used for regression with GRoI tend to be closer to the estimation site, on average, than stations used with PRoI. The use of a generalized

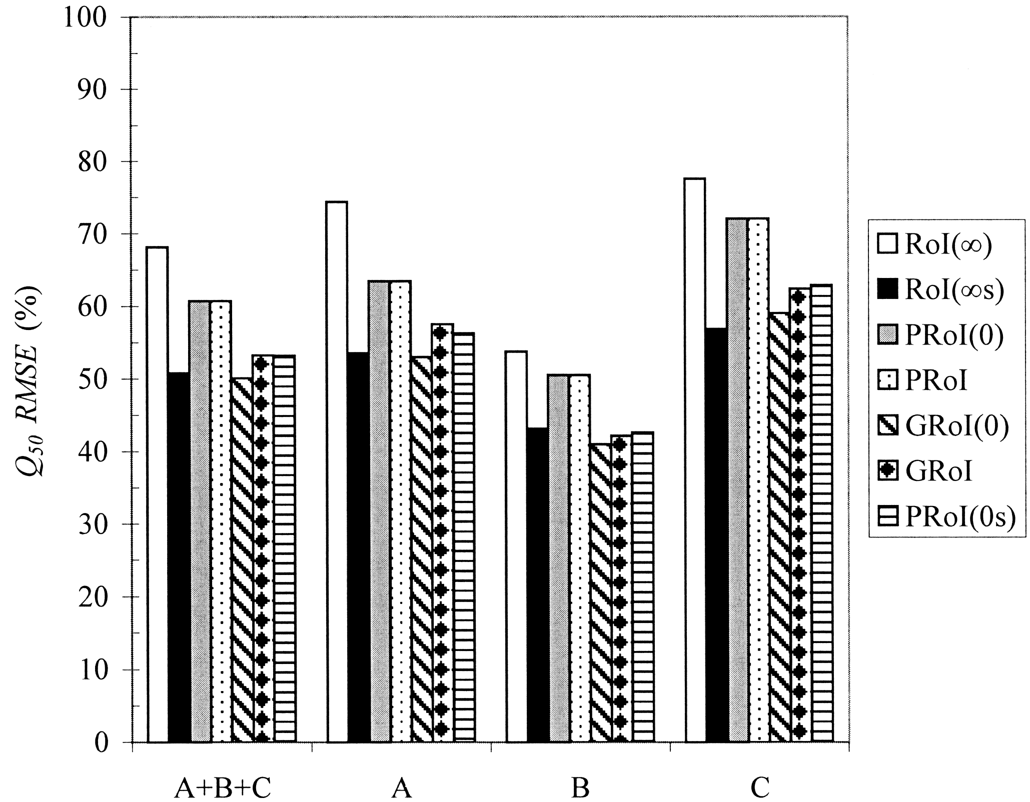

Figure 5. *RMSE*, in Percent, Evaluated Over the Evaluation Dataset. 'A,' 'B,' and 'C' represent the groups on Figure 1, and 'A+B+C' is for the entire study area.

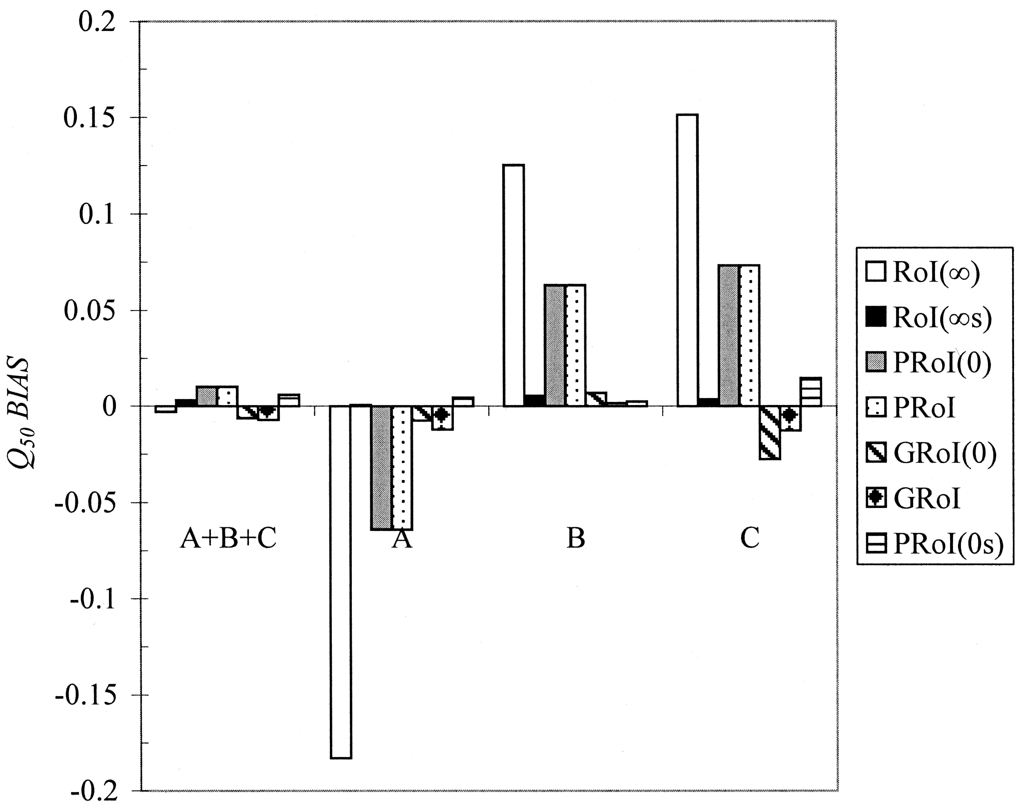

Figure 6. The Average Differences (*BIAS*) Evaluated Over the Evaluation Dataset. 'A,' 'B,' and 'C' represent the groups on Figure 1, and 'A+B+C' is for the entire study area.

least squares regression would remove this bias (Stedinger and Tasker, 1985). However, the data for such an analysis were not available at the time of this study.

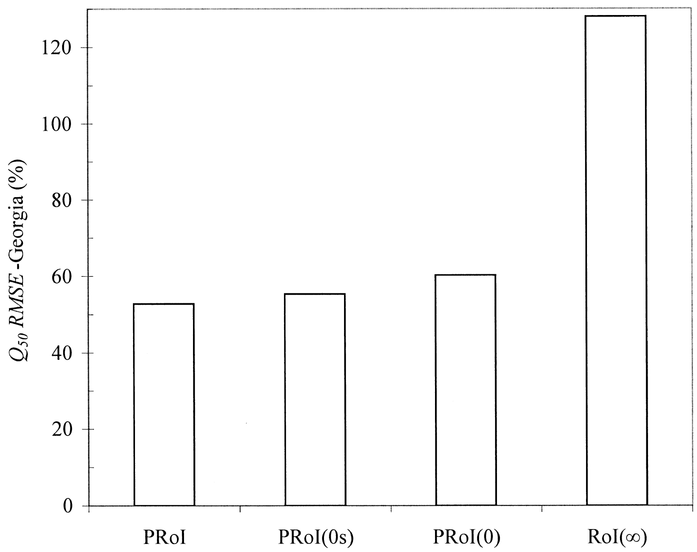

Figure 7. *RMSE*, in Percent, Comparison of Predictor Variable Region-of-Influence (PRoI) Approaches in Georgia.

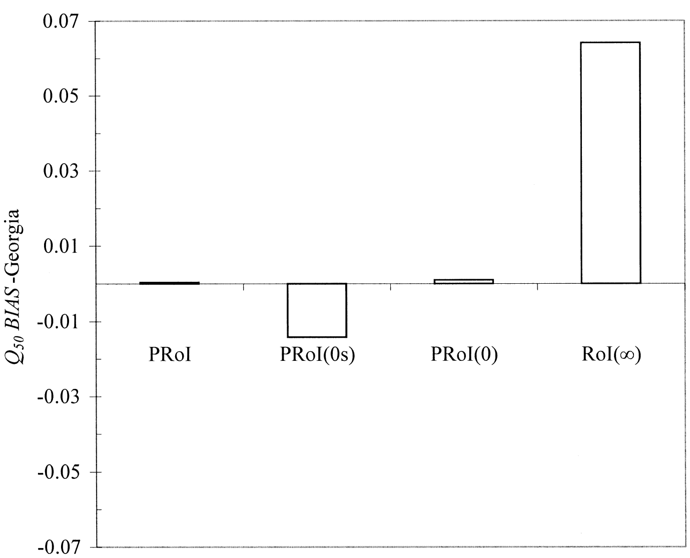

Figure 8. *BIAS* Comparison of Predictor Variable Region-of-Influence (PRoI) Approaches in Georgia.

ACKNOWLEDGMENTS

The authors wish to thank Wladmir Guimaraes, U.S. Geological Survey, Columbia South Carolina, for providing the estimates of 50-year flood peaks and basin characteristics. The authors also thank Krista Dunne, U.S. Geological Survey, Princeton New Jersey, for providing geographic datasets used in this study.

LITERATURE CITED

Aitchison, J. and J.A.C. Brown, 1957. The Lognormal Distribution. Cambridge University Press, Cambridge, Massachusetts, 176 pp.

Berenbrock, C., 2002. Estimating the Magnitude of Peak Flows at Selected Recurrence Intervals for Streams in Idaho. U.S. Geological Survey Water-Resources Investigations Report 02-4170, 59 pp.

Burn, D.H., 1990. Evaluation of Regional Flood Frequency Analysis With a Region of Influence Approach. Water Resour. Res. 26(10):2257-2265.

Feaster, T.D. and G.D. Tasker, 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, 41 pp.

Hammond, E.H., 1964. Analysis of Properties in Land Form Geography: An Application to Broad-Scale Land Form Mapping. Annals Assoc. American Geophys. 54:11-23.

Hydrology Subcommittee of the Interagency Advisory Committee on Water Data, 1982. Guidelines for Determining Flood Flow Frequency. Bulletin 17B of the Hydrology Subcommittee, Office of Water Data Coordination, U.S. Geological Survey, Reston, Virginia, 99 pp.

Montgomery, D.C., E.A. Peck, and G.G. Vining, 2001. Introduction to Linear Regression Analysis (Third Edition). John Wiley and Sons, Inc., New York, New York, 641 pp.

Neely, B., 1986. Magnitude and Frequency of Floods in Arkansas. U.S. Geological Survey Water Resources Investigations Report 86-4335, 51 pp.

Pope, B.F., G.D. Tasker, and J.C. Robbins, 2001. Estimating the Magnitude and Frequency of Floods in Rural Basins of North Carolina – Revised. U.S. Geological Survey Water Resources Investigations Report 01-4207, 44 pp.

Snee, R.D., 1977. Validation of Regression Models: Models and Examples. Technometrics 19:415-428.

Stedinger, J.R. and G.D. Tasker, 1985. Regional Hydrologic Analysis 1 – Ordinary, Weighted, and Generalized Least Squares Compared. Water Resour. Res. 21(9):1421-1432.

Tasker, G.D., S.A. Hodge, and C.S. Barks, 1996. Region of Influence Regression for Estimating the 50-year Flood at Ungaged Sites. Water Resources Bulletin 32(1):163-170.

Tasker, G.D. and R.M. Slade, Jr., 1994. An Interactive Regional Regression Approach to Estimating Flood Quantiles. *In:* Water Policy and Management: Solving the Problems, D.G. Fontane and H.N. Tuvel (Editors). ASCE Proceedings of the 21st Annual Conference of the Water Resources Planning and Management Division, pp. 782-785.

NOAA (National Oceanic and Atmospheric Administration), 1976-1978. Climates of the United States, Climatology of the United States. No. 60, Parts 1-52, U.S. Department of Commerce, Washington, D.C.

Wandle, S.W., 1977. Estimating the Magnitude and Frequency of Floods on Natural-Flow Streams in Massachusetts. U.S. Geological Survey Water Resources Investigations Reports 77-39, 27 pp.