Information Needs for the Proper Application of Hydrologic Regional Regression Equations

JOHN OWEN HURD

Regional regression equations are used by many state and local agencies to estimate flood characteristics used in the design of drainage structures. The regression equations are preferred by many users because of their relative accuracy, simplicity (i.e., ease of application), and replicability. However, the experience of the Ohio Department of Transportation indicates that some problems do exist in applying the equations. The problems are in part caused by the disparity in background of report writers and users and in part to limitations of the equations. The experience of ODOT is used to demonstrate some of the potential misapplications of regional regression equations, and some of the information that has been given to designers to clarify the application of the equations is presented.

There are numerous methods of estimating flood characteristics (such as flood peaks and hydrographs) to be used in the design of drainage facilities for ungauged drainage basins. These include the rational method (*1*), SCS method (*2*), index flood method (*3*), regional regression equations (*4*), and so on. Regional regression equations are used by many state and local transportation agencies to estimate flood characteristics used in the design of drainage structures. The most commonly used equations are those used to estimate flood peaks. These equations generally relate flood characteristics to certain drainage basin characteristics, climatic variables, and channel characteristics. The equations are developed by using multiple-regression analyses to relate flood characteristics determined from flood records at gauged sites to the basin characteristics of these sites. There has been much deliberation over the correct means of developing the flood characteristics from the gauged streamflow data (*5*). However, for the purpose of this discussion it is assumed that the methods used to determine the flood characteristics at the gauged site are accurate.

The regression equations are preferred by many users because of their relative accuracy, simplicity (i.e., ease of application), and replicability. The regional equations may more accurately estimate flood characteristics within their range of use because they are developed from flood data for a specific region, whereas methods based on a larger broader base of flood data may be less accurate. Flood peak estimates for 25- and 100-year recurrence intervals using current Ohio equations (*6*) provide the same accuracy as log-Pearson Type III estimates from station records of 7 and 9 years, respectively. Most regional regression equations are easy to use. They represent flood characteristics as simple log-linear functions of a few easily determined drainage basin parameters such as drainage area, main channel slope, and average annual precipitation. Generally, little subjective judgment is required by the user in determining values for these parameters. Flood characteristics estimated by use of the equations are usually insensitive to small errors in parameter estimation. Therefore, nearly all practitioners will arrive at the same results by using the equations to estimate flood characteristics at a selected ungauged site.

THE PROBLEM

It appears that a problem could not exist if this method is so accurate, easy to use, and gives such consistent results. However, the experience of the Ohio Department of Transportation (ODOT) indicates that problems do exist in applying the equations. Part of the reason for these problems involves the individuals performing the design of drainage facilities for various governmental agencies or private consultants. These individuals are rarely hydrologists and their knowledge of the statistical methods used in development of the regional regression equations is often limited at best. This lack may be true even in large agencies and consulting firms. In some instances, drainage designers are not registered engineers, but technicians with limited educational background in the theory behind the hydraulic design methods they are using. In other instances, individuals doing drainage design are not specialists, but general highway designers performing drainage design, structural design, highway design, and so on. The limitation may be especially common within small agencies or consulting firms. These people are not normally afforded the time to become familiar with the intricacies and background of all the design methods being used.

Another reason for the problems involves the limitations of the equations. The equations are generally accurate only within the specific geographic area for which they were developed. Their use is further limited to the particular type of basins from which data used in their development were gathered (e.g., natural unregulated rural basins). The application of the equations should be limited to sites with basin characteristics within the ranges of the basin characteristics in the original data set used to develop the equations.

Even then the equations may give biased estimates of flood characteristics for basins with certain basin features. In Ohio, use of the rural flood peak equations for steep basins with main channel orientation aligned with prevailing storm move-

---

Ohio Department of Transportation, 25 South Front Street, Columbus, Ohio 43215.

ment (*7*) and basins with a large percentage of strip-mined area (*6*) have been identified as giving biased estimates by observation of residuals between gauging station flood peak estimates and those obtained from regression equations. This situation normally occurs when there are insufficient data available to statistically quantify the effect of that particular characteristic for which flood estimates are biased.

The methods used to develop the regional regression equations and the limitations of the equations are usually thoroughly discussed in the research reports providing the equations. The reports are generally well written, but they are written as research reports for those familiar with the practice of hydrology and not as users manuals for the practicing drainage designer. Because of the disparity between the backgrounds of the report writers and those of the report users, there may often be errors in application of regional regression equations in spite of the report quality. One reason may be that designers are not reading the reports in detail. This is not recommended but is understandable because the equations are generally so easy to apply.

In the following section, the experience of ODOT is used to demonstrate some of the potential misapplications of regional regression equations, and some of the information that has been given to designers to clarify the application of the equations will be presented.

ODOT EXPERIENCE

ODOT has used regional regression equations to estimate rural flood peaks for more than 20 years. ODOT also currently uses regional regression equations to estimate urban flood peak discharges, urban flood hydrographs associated with flood peaks, urban basin lag times, and urban flood volumes associated with flood peaks for small streams. At the time of this report, equations were being developed to estimate rural flood hydrographs associated with flood peaks, rural basin lag times, rural flood volumes associated with flood peaks, and rural and urban flood volume, duration, and frequency relations for small streams. During this time, many misapplications of the equations have occurred and many questions from drainage designers have been received. The following subsections categorize the misapplications and questions.

Determination of Geographic Region

It should be relatively simple for the designer to determine the geographic region in which the ungauged drainage area of interest is located. This is not always the case. The first regional regression equation used by ODOT to estimate flood peak discharges for rural basins (*8*) was

$$Q_{2.33} = 50CA^{0.8}\text{Sl}^{0.298} \tag{1}$$

where

$Q_{2.33}$ = mean annual flood, ft^3/sec;
A = drainage area, mi^2;
Sl = main channel slope, ft/mi, and
C = a regional constant ranging from 0.3 to 1.7.

The 5- through 100-year rural peak discharges were determined by multiplying $Q_{2.33}$ by factors that were a function of recurrence interval and drainage area. This equation was used for the entire state of Ohio.

The factor C was used to eliminate geographic bias in flood estimates from the statewide equation. It was delineated by soil association boundaries of the Soil Conservation Service (SCS). Because these boundaries were not precisely defined on the maps available to drainage designers and did not coincide with drainage basin boundaries, the designers were often left with a difficult subjective decision to make about which value of the coefficient to use. Because of the wide range in C values, a large difference in flood peak estimates could result. The most common difficulty was differentiating between $C = 1.0$ or 1.6. Because practitioners generally use a conservative approach, a 60 percent overestimate in flood peaks often resulted.

Equation 1 has since been replaced twice with updated regional regression equations (*6,9*). In each case, the regional boundaries were defined by drainage basin divides in place of soil association areas. Region boundaries may cross the main channel of a large well-gauged stream, but the location is clearly noted as either just upstream or downstream of a major tributary. Analyses were performed during development of these equations to determine whether any soil parameter had a significant effect on flood peaks. No soil parameters were determined to have a significant effect. Designers no longer had any problem determining which equations to use. They have only to follow the watercourse for the basin of interest down to a major stream if the drainage basin happens to be near a regional boundary.

Determination of Equation Suitability

The designer must also determine whether the regional regression equation is suitable for the drainage basin being analyzed. It may not be as simple as it would seem to confirm that the basin is, for example, rural, unregulated, and within the equation parameter limits.

As previously stated, ODOT currently uses regional regression equations to estimate flood peak discharges from both rural and urban basins (*10*). The urban peak equations are

$$Q_T = \text{RC} \cdot A^a\text{Sl}^b\text{El}^c(13 - \text{BDF})^d \tag{2}$$

where

Q_T = T year peak discharge, ft^3/sec;
RC, a, b, c, and d = regression constants,
A = drainage area, mi^2;
Sl = main channel slope, ft/mi;
El = basin elevation index, 1,000s of feet; and
BDF = a basin development factor ranging from 0 (little or no development) to 12 (fully developed).

The rural peak equations are of the form

$$Q_T = \text{RCA}^a\text{Sl}^b(\text{St} + 1)^c \tag{3}$$

where St is the percentage of the area in storage and the other terms are as defined for Equation 2.

Many drainage basins in Ohio are partially developed. Hence, these basins fall into a neutral category in which the effects of the development on stream flow may not be significant. For those areas with slight urban development (BDF < 5), it is recommended that both the rural and urban methods be considered (*10*). The designer must then decide which estimate to use. This decision is complicated by the fact that in some cases use of the rural equations provides larger flood peak estimates than use of the urban equations. On the basis of the results of studies demonstrating the effect of location of detention facilities on flood peaks from urbanized areas, it appears reasonable to assume that development in the lower one-third of a basin could reduce flood peaks compared with natural conditions, whereas development in the upper one-third would increase the peaks substantially (*11–13*). Thus designers have been offered the following guidelines for estimating peak discharges using both urban and rural flood peak equations. If development is in only the lower one-third of the basin, use of the lesser peak discharge is recommended; if the development is only in the upper one-third of the basin, use of the greater peak discharge is recommended; if the development is in the middle or spread throughout, use of the average is recommended.

Once it is determined that the basin is rural, the rural equations may still not be applicable to the particular unregulated basin even if its characteristics are within the range of parameter limits. More information about the basin may be required to determine if the regional regression equations can be used to predict flood peaks (or other characteristics) for the basin. In Ohio, for example, basin characteristics of the 275 gaged basins used to develop the final regional regression equations for rural flood peaks were determined from U.S. Geological Survey (USGS) 7.5-min (1 in. = 2,000 ft) topographic quadrangle maps. Furthermore, the main channel from the basin mouth to the divide was defined from the contours on these maps.

In order to confidently use the regional regression equations, the designer should be able to determine the value of each independent parameter by the same means and with the same ease as was done for the gauged sites during equation development. In the particular case of Ohio's flood peak equations, if a basin boundary cannot be defined without precise mapping or a definite channel does not exist (e.g., sheet flow areas), other discharge estimation methods should be considered. A rule of thumb ODOT provides to designers using the Ohio rural flood peak regression equations is the following: "If all the information needed for the equation cannot be confidently obtained from a USGS 1 in. = 2,000 ft map, another method should be considered." ODOT recommends the use of the rational method for determining flood peaks for drainage areas of small sheet flow type.

The suitability of the regression equation may still not be verified. The values of parameters obtained should fall within the range of those parameters at the gauged sites used to develop the equations. This is not always as simple to verify as it may seem. A particular basin may have characteristics each of which has a value within the range of the basin characteristic of the original data set. However, the point formed by these values may be outside the range of points in multiple-dimensional space for the data set used in equation development.

Consider Equation 3 used by ODOT to estimate rural flood peaks. For the data sets used to develop the equations, drainage area and main channel slope are highly correlated. In general, as drainage area decreases, main channel slope increases. Thus, a large basin with a steep channel slope may lie outside the two-dimensional space of the original data set even though each characteristic may have a value within the range of that particular variable. Figure 1 shows an exaggerated illustration of this concept.

If the regional regression equations involve only one or two basin characteristics, it can easily be determined if the basin of interest has characteristics within the multiple-dimensional space either directly or graphically. However, most current regional regression equations contain three or more basin characteristics. The urban peak equations used by ODOT have four variables. Manually solving this problem for three or more variables is virtually impossible. However, with high-powered desktop computers available, simple programs can be developed to determine whether the basin of interest lies with the multiple-dimensional space of the original data set. Currently, the USGS Water Resources Division, Ohio Office, is developing such a microcomputer program for the rural peak equations used by ODOT.

Application of the Equations

After it is determined that the regional regression equations are applicable to the ungauged site being studied and which particular regional regression equation should be used, the appropriate equation must be applied properly.

As has been previously discussed, the designer should be able to obtain values of basin characteristic parameters by the same methods used for the data in development of the regional regression equations. For some parameters, it may be imperative to obtain the parameter value in the same way, whereas for others it may not. For example, Equation 3 is examined in more detail.

Although the designer should be able to delineate the drainage area boundary on a 1 in. = 2,000 ft topographic map, the drainage area may be determined from a topographic map of a different scale. The scale of a map should not significantly affect the accuracy of the determination, nor should a different scale than 1 in. = 2,000 ft cause constant over- or underestimation of the value. However, use of a less accurate map may increase the variability of the drainage area estimate and is not recommended.

The same cannot be said for determination of main channel slope, which is defined as "the difference in elevation at points

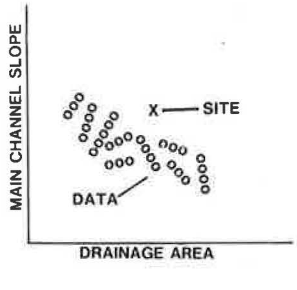

FIGURE 1 Illustrative plot of basin characteristics outside the multidimensional space of the data.

10 and 85 percent of the distance along the main channel from the specified location on the channel to the topographic divide, divided by the distance between the points" (*6*). Figure 2 shows how a sinuous drainage basin channel might be plotted on different scale maps. Although the 10 and 85 percent elevations would remain relatively the same regardless of the map scale, the length of the channel would probably increase as the scale of the map becomes larger. Thus, the main channel slope would decrease. If the regression equations were developed using values of main channel slope determined from 1 in. = 2,000 ft scale map, use of larger scale maps by the designer would consistently underestimate the main channel slope, which would result in consistently underestimated flood peaks. The converse would be true if a smaller scale were used. A not uncommon error in main channel slope estimate of 20 percent results in a 4 to 13 percent error in flood estimates for the usual 0.2 to 0.6 range in main channel slope exponent. ODOT has recommended that designers use 1 in. = 2,000 ft topographic maps to estimate main channel slope.

Storage is defined for Equation 3 as "the percentage of the contributing drainage area occupied by lakes, ponds, and swamps as explicitly shown on USGS 7.5-min (1 in. = 2,000 ft scale) topographic quadrangle maps" (*6*). The equation does not apply to drainage basins where reservoirs provide enough storage to cause the basin to be considered regulated because such basins were not included in the regression analyses (*6*). The parameter involves only surface area and no consideration is given to whether the pond or swamp is full or empty, if any freeboard exists, or what total number of acre-feet of storage is provided. The definition of storage provided by Webber and Bartlett (*9*) was not as clear regarding how storage was determined. Several designers had underestimated this parameter using the equations from that report because field checks indicated that the ponds in the area were "generally full at the time of the field check and appeared to provide little storage capacity." Failure to take full account of storage will consistently overestimate flood peaks. Ignoring a 2 percent storage value in the current equations (*6*) for rural peaks results in a 38 to 49 percent overestimate. To avoid this problem, the current report clearly defined storage exactly as it was determined for the gauged sites used in the multiple regression analyses.

The fact that all information required for using Equation 3 can or should be obtained from 1 in. = 2,000 ft scale top maps does not excuse the designer from verifying that the information on the topographic map is accurate and current. Verification may include a combination of field checks, review of recent aerial photographs, consultation with local officials, etc. Thorough field inspection is recommended for urban basins. To apply Equation 2 for urban areas, the designer must do a field check to determine the basin development factor (BDF), and must use local sewer and street maps to verify the actual drainage area.

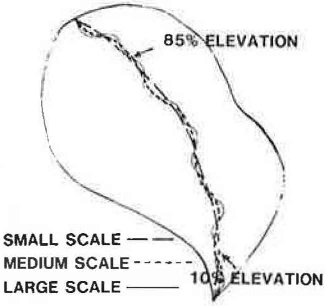

FIGURE 2 Main channel plotted on different scale maps.

Extrapolation of the regional regression equations beyond the limits of the original data used to develop the equations is not recommended. However, there are times when the drainage basin being studied has one parameter with a value just outside the recommended range and no other method of flood estimation appears more reasonable. The designer must evaluate whether or not this extrapolation may produce an over- or underestimate of the flood characteristic. Normally, if the values of the other parameters are well within the ranges of the original data set, slight extrapolation may be reasonable unless computation of the flood characteristic is unusually oversensitive to small changes in the value of the extrapolated parameter.

Before development of regression Equation 3, the equations used for one geographic region in Ohio were as follows (*9*):

$$Q_2 = 42.6A^{0.802}\mathrm{Sl}^{0.225} \tag{4}$$

$$Q_{100} = 52.6A^{0.857}\mathrm{Sl}^{0.619} \tag{5}$$

with all symbols as defined for previous equations. Q_5 through Q_{50} were similar with the power of slope increasing proportionally and the regression constant and the power of area increasing only slightly. The increase in peak discharge estimates from Q_2 through Q_{100} is more commonly the result of a comparable increase in the regression constants with the exponents of the parameters changing much less. However, in this particular regression analysis the increase from Q_2 through Q_{100} is reflected more in the unusually large increase in the power of slope, which apparently compensates for the lack of change in the regression constant. These equations were carefully checked for bias with regard to slope, but none was observed. Although the increase in exponent for main channel slope was unusual, the sensitivity of flood peak estimates to relative changes in main channel slope was not considered excessive. Therefore, no effort was made to change the form of the variable in the equation to reduce sensitivity. Both Equation 3 and all other regional regression equations from the earlier work (*8,9*) have increasing regression constants as recurrence interval increases, whereas changes in the slope exponents are much smaller and the exponent may increase slightly or decrease as the recurrence interval increases.

The particular region for which Equations 4 and 5 were developed has very steep watersheds. It is not usual to find ungauged drainage basins that have main-channel slopes outside the range of those in the original data set. Extrapolating Equation 4 to steeper slopes would appear reasonable because the value of Q_2 is somewhat insensitive to an increase in slope. However, extrapolation of Equation 5 for steeper slopes may overestimate Q_{100} because the value of that flood peak appears unusually sensitive to any increase in slope. Any flood characteristics computed by equations such as Equation 5 through extrapolation of slopes should be checked by an independent method.

Where the ungauged basin has a particular basin characteristic for which the regional regression equations are known to give biased estimates, subjective adjustment of the estimated flood characteristics may be required. For example, flood peaks estimated by use of the current regional regression equations may need to be increased for steep drainage basins oriented within 20 degrees of the prevailing storm movement (*7*), whereas peak estimates for basins with a larger percentage of strip mine areas may need to be decreased (*6*). In each case, the magnitude of the adjustment may be estimated comparing the presumed biased estimates with estimates based on methods that are either less sensitive to the particular basin characteristic causing the bias or which take into account the effect of that characteristic.

Equations relating flood peaks to channel shape (geometry) characteristics have been developed for Ohio (*14,15*). They are not as accurate as the basin characteristic equations with regard to the regression statistics. However, they provide a means to check flood peak estimates computed by questionable application of the basin characteristic equations. They were developed at ODOT's request to provide an independent method of flood peak estimation for watersheds where the more conventional basin characteristic equations are known to produce biased estimates. The channel-shape equations should be used to help quantify an adjustment to the flood estimates computed by use of the basin characteristic equation.

RECOMMENDATIONS

On the basis of ODOT experience with the use of the regional regression equations to estimate flood characteristics for ungauged watersheds, the following recommendations are presented.

1. Where possible, watershed divides should be used for regional regression equation boundaries.
2. The methods used to determine the value of each parameter during development of the regional regression equations should be clearly defined in research reports.
3. Users of regional regression equations should be advised if the values of parameters must be determined by the same means as those used for the original data set.
4. Microcomputer programs should be developed to determine whether selected ungauged sites have parameter values within the multidimensional space of the original data set used to develop the regional regression equations.
5. Short courses should be held for users of the equations as part of the implementation process for any new reports.
6. Independent methods of estimating flood characteristics should be developed to be used as check methods in which use of the regional regression equations produces questionable results.
7. Research should be conducted to quantify the effects on flood characteristics of main-channel orientation, strip-mined areas, and other basin characteristics that affect the accuracy of flood characteristic estimates.

ACKNOWLEDGMENTS

This paper was funded by the Ohio Department of Transportation.

Sincere appreciation is given to James M. Sherwood and other personnel of the Water Resources Division, Ohio Office, U.S. Geological Survey, for their critical review of this paper.

REFERENCES

1. Design and Construction of Sanitary and Storm Sewers. *Manual and Report on Engineering Practice 37*, ASCE, New York, 1970.
2. *Urban Hydrology for Small Watersheds*. Technical Release 55; Soil Conservation Service, U.S. Department of Agriculture, Washington, D.C., 1975.
3. T. Dalrymple. *Flood Frequency Analyses*, Water Supply Paper 1543-A. U.S. Geological Survey, 1960.
4. F. D. Masch. *Hydrology*. FHWA-IP-84-15 (HEC 19). FHWA, U.S. Department of Transportation, 1984.
5. V. P. Singh. Hydrologic Frequency Modeling. *Proc., International Symposium of Flood Frequency and Risk Analysis*, D. Reidel Publishing; 1987.
6. G. F. Koltun and J. W. Roberts. *Techniques for Estimating Flood-Peak Discharges of Rural Unregulated Streams in Ohio*. Water Resources Investigations Report 89-4126, U.S. Geological Survey, 1989.
7. J. O. Hurd. Effect of Main Channel Orientation on Flood Peaks for Streams in Ohio. In *Transportation Research Record 1073*; TRB, National Security Council, Washington, D.C., 1986.
8. W. P. Cross and R. I. Mayo. *Floods in Ohio Magnitude and Frequency*. Bulletin 43. Ohio Department of Natural Resources, Columbus, 1969.
9. E. E. Webber and W. P. Bartlett, Jr. *Floods in Ohio Magnitude and Frequency*. Bulletin 45. Ohio Department of Natural Resources, Columbus, 1977.
10. J. M. Sherwood. *Estimating Peak Discharges, Flood Volumes, and Hydrograph Shapes of Small Ungaged Urban Streams in Ohio*. Water Resources Investigations Report 86-4197. U.S. Geological Survey, 1987.
11. R. H. McCuen. A Regional Approach to Urban Stormwater Detention. *Geophysical Research Letters*, Vol. 1, No. 7, 1974.
12. D. C. Curtis and R. H. McCuen. Design Efficiency of Stormwater Detention Basins. *Journal of the Water Resources Planning and Management Division, ASCE*, May 1977.
13. A. C. Flores, A. M. Bedient, and L. W. Mays. Method for Optimizing Size and Location of Urban Detention Storage. *Proc., International Symposium on Urban Hydrology, Hydraulics, and Sediment Control*, 1982.
14. E. E. Webber and J. W. Roberts. Floodflow Characteristics Related to Channel Geometry in Ohio. Open File Report 81-1105. U.S. Geological Survey, 1981.
15. D. K. Roth. Estimation of Flood Peaks from Channel Characteristics in Ohio. Water Resources Investigations Report 85-4175. U.S. Geological Survey, 1985.

Publication of this paper sponsored by Committee on Hydrology, Hydraulics, and Water Quality.