PREDICTING THE 1.5-, 5-, 10-, AND 25-YEAR EVENT FROM THE SNOW ZONES OF COLORADO AND WYOMING

C.A.Troendle[1] and L.S.Porth[2]

ABSTRACT

Scientists have been trying to develop equations to accurately predict streamflow for years. As of yet the most well known equations to do this were developed by the United States Geological Survey (USGS). This paper presents the results of our search to improve equations that predict the 1.5-, 5-, 10-, and 25-year maximum instantaneous flow and maximum mean daily flow for watersheds in the snow zones of Colorado and Wyoming. Using the same set of data utilized by the USGS, but with a more rigorous selection criteria, we have developed prediction equations that not only have a much smaller range of error, but these equations can be applied over a broader geographic area, rather than being restricted to specific Hydrologic Regions.

INTRODUCTION

Engineers, hydrologists, and land managers frequently have the need to predict the magnitude of stream discharge (peaks) for a variety of design purposes. Most often, in wild land situations, streamflow measurements are not readily available thus making prediction of the "design" event difficult. In the absence of a direct measure, deterministic or empirical models must be used to predict the event. This task is also challenging because more often than not, required input data describing watershed condition is usually limited, making implementation of all but the simplest of empirical or nomographic models difficult (e.g. see Troendle and Leaf 1980). In Colorado, as well as much of the western United States, the ability to predict discharges with varying return intervals is critical to developing in-stream flow requirements for aquatics, fishery and channel morphology needs as well as for design purposes such as culvert and storm drain sizing.

The United States Geological Survey (USGS) has developed a series of equations, stratified by hydrologic region within Colorado, that are useful in predicting the 2-, 5-, 10-, 25-, 50-, 100, and 500-year instantaneous peak discharge from high elevation, snow zone, watersheds (Kircher et al. 1985, Jennings et al. 1994). The equations, developed using long-term USGS streamflow data from selected gauging stations, work well and require little data, primarily an estimate of watershed area and basin characteristics such as average annual precipitation.

This paper represents an attempt to revisit the USGS streamflow data used to develop the flood frequency prediction equations. We screened the data for specific application to the development of flood frequency prediction and a refined set of equations are developed based on the screened data set, if warranted and justified by improved model fit.

OBJECTIVES

The objective of this study is to improve our ability to predict the 1.5-, 5-, 10-, and 25-year maximum instantaneous flow and maximum mean daily flow from watersheds in the snow zones of Colorado and Wyoming.

METHODOLOGY

The effort to "improve" the empirical equations was initiated by revisiting the available USGS streamflow record for gauging stations in Colorado and Wyoming. We first identified a set of criteria to use in selecting the streamflow record to be retained for further analysis. First, the minimum of a 10-year streamflow record needed to be available for each site selected. Second, the stream gauge needed to be located at or above 7500 feet in elevation to insure a snowmelt dominated hydrology. Third, the streamflow regime could not be altered by diversion or augmentation, as we would have no basis for estimating or correcting for the impact. These three

---

[1] Consulting Hydrologist, MATCOM Corporation, 2629 Redwing Road, Suite 110, Fort Collins, CO 80526
[2] Statistician, USFS, Rocky Mountain Research Station, 240 West Prospect Road, Fort Collins, CO 80526

Presented at the 68TH Annual Meeting of the Western Snow Conference; Port Angeles, WA, April 2000

selection criteria identified those streamflow records that would be retained for further assessment or evaluation. Though this process was completed for both Colorado and Wyoming, the Wyoming watershed boundaries fell outside the GIS platform available to us. Therefore, the GIS area-based watershed description parameters (such as slope, precipitation, and area) could not be generated for the Wyoming watersheds. However, these sites were later used for testing the developed models where possible.

From the more than 400 gauging stations initially available we identified six watersheds, distributed throughout the study area, with long-term stable streamflow records, that could function as reference sites. Three of the watersheds are located in Wyoming: East Fork River near Big Sandy, East Fork Big Goose Creek near Big Horn, and the Encampment River above Hog Park Creek near Encampment. The other three watersheds are in Colorado: Halfmoon Creek near Malta, Dolores River below Rico, and Piney River below Piney Lake near Minturn. Stability of the reference gauges was verified by double-mass plotting the annual instantaneous peak discharge values from each of the watersheds over comparable data from several of the other watersheds (Anderson 1955, Burton 1997, Troendle and Stednick 1999). Stability was assumed for all the double-mass comparisons that reflected a reasonably consistent linear relationship (e.g. see fig. 1).

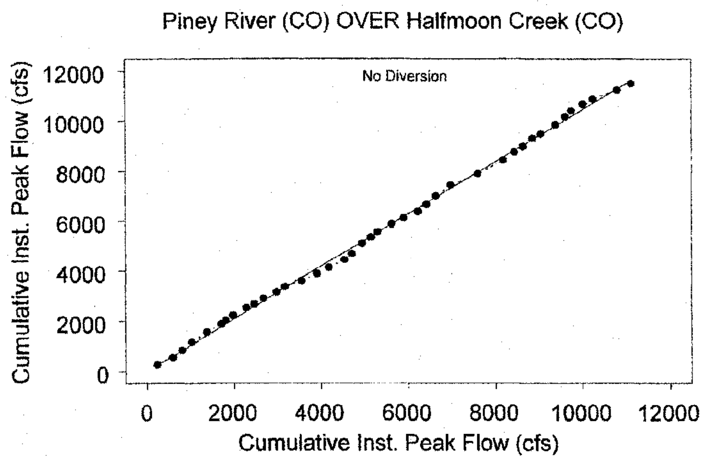

Figure 1. Double-mass plot showing the linear relationship between two of the Colorado "reference" sites.

Once the stability analysis was completed, the six watersheds were considered "reference" sites and used as the base for comparison with the record from all other gauging sites. The overlapping portions of the streamflow record for every other watershed retained for analysis were then "double-mass" plotted against the peak discharge records for the two nearest "reference" watersheds (e.g. see fig. 2). There were a large number of watersheds that

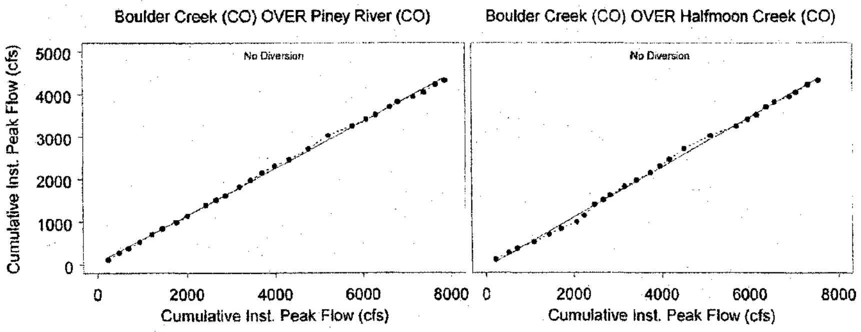

Figure 2. Double-mass plots showing the consistent linear relationship required between a particular site (Boulder Creek) and two "reference" sites (Piney River and Halfmoon Creek) in order to be included in the analysis.

demonstrated substantial deviations in the double-mass plots (e.g. see fig. 3 and fig. 4). The deviations/deflections that we observed could be the result of movement in gage location, significant revisions in rating curves used to estimate discharge, watershed disturbance above the gauging site, or diversions/augmentations in the flow regime (documented or undocumented). We did not attempt to correct or adjust streamflow records or evaluate the cause of departures. Although arbitrary in nature, those watersheds with long-term records that were not stable, relative to the two nearest reference sites (e.g. fig. 3 and fig. 4), were dropped from further analysis. It should be noted that this assessment was made on the maximum instantaneous discharge for each water year and as such, the assessment of data only relates to that particular measure of discharge. Subsequently, inferences cannot be drawn, from this assessment, on other indices of discharge such as mean daily, monthly, or annual flows.

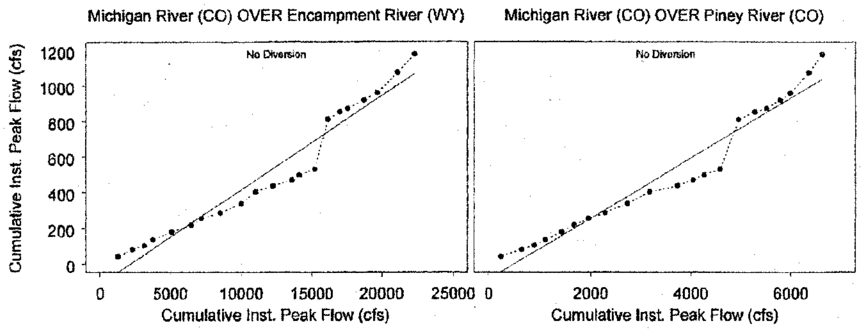

Figure 3. Double-mass plots showing the inconsistent relationships one particular site (Michigan River) had with two "reference" sites (Encampment River and Piney River) that resulted in removal from the analysis. The departure clearly resides in the Michigan River streamflow data.

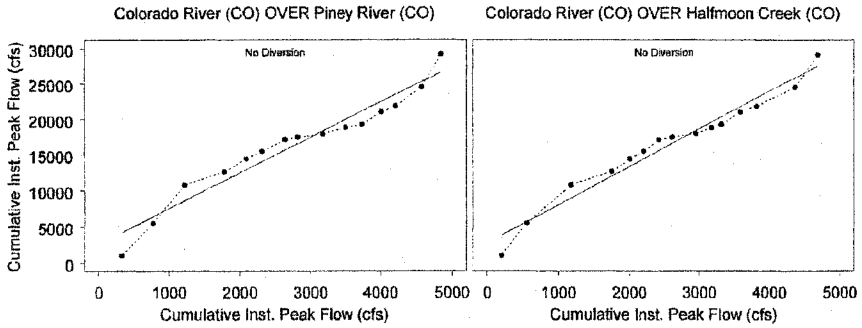

Figure 4. Another set of double-mass plots showing the inconsistent relationship one particular site (Colorado River) had with two "reference" sites (Piney River and Halfmoon Creek) that resulted in removal from the analysis.

Initially, 411 Colorado stream gauges were identified as occurring at or above 7500 feet in elevation and having data for annual maximum instantaneous discharge. Maximum mean daily discharge was also available for the majority of these stations. Of those, 262 gauges had records at least 10 years in length, although 30 of those 262 stations lacked published "station" remarks. However, of the 262 gauges having 10 years or more of record, only 56 were non-diverted. The streamflow record from all 262 streams was analyzed (double-mass analysis with two reference sites) and the data set was further reduced to 30 gauges with a stable flow regime from which flood-forecasting equations could then be developed. This number was significantly less than the 200+ sites used by the USGS (Kircher et al 1985, Jennings et al. 1994) in developing the published regional equations for Colorado. Similarly, we initially identified 80 stream gauging stations above 7500 feet in elevation in Wyoming, 61 of which had 10 or more years of records, but only 14 of which were not diverted. This number was ultimately reduced to 11 usable sites. As mentioned earlier, the Wyoming data were not used in model development phase but were later used to validate the Colorado based models that were developed.

Once the 30 sites were identified in Colorado, the U.S. Forest Service (personal communication, Eric Butler, R2-GIS specialist) assisted in the database development. The Oregon State Climate Center digital precipitation map was obtained for Colorado and used as a layer on the Digital Elevation Map (DEM). The 30 stream gauging sites were located on the DEM (fig. 5), watershed boundaries were identified, and average slope, elevation, and precipitation for each watershed were determined. Watershed areas ranged from 0.9 to 131.0 mi^2, elevations ranged from 7560 to 10430 feet, average slope ranged from 12 to 56%, and average annual precipitation ranged from 20 to 43 inches. In addition, the percentage of each watershed area occupied by forest vegetation was calculated, but this did not prove to be a significant parameter in predicting peak discharges.

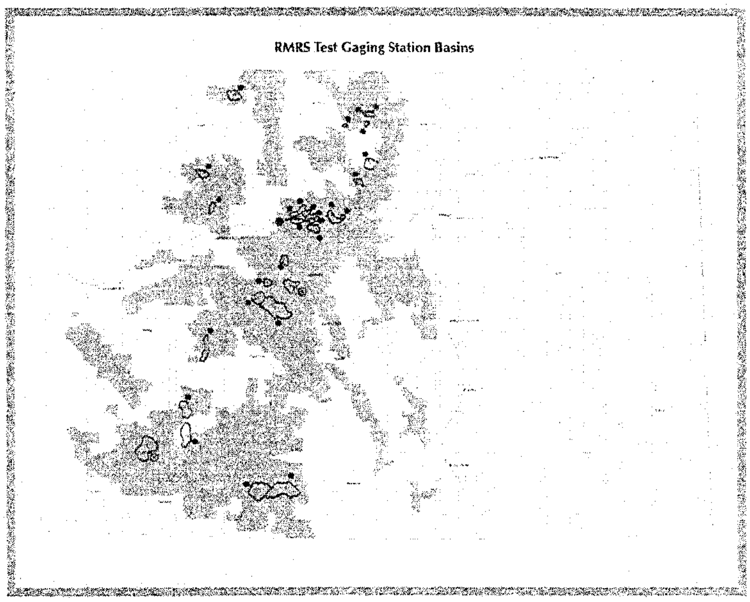

Figure 5. Digital Elevation Map denoting location of 30 Colorado stations (•) used to develop new flow prediction equations. Three of the 6 reference watersheds (⊙) are shown, but the other three (East Fork River near Big Sandy, East Fork Big Goose Creek near Big Horn, and the Encampment River above Hog Park Creek near Encampment) are just above the CO/WY border.

The analysis focuses on the prediction of the 1.5-, 2-, 5-, 10-, and 25-year instantaneous peak and maximum mean daily discharge, expressed as either the Weibull or the Log-Pearson Type III probability of occurrence. Because the return interval for the Weibull distribution is plotted from observed data, only 18 of the 30 watersheds had station record of sufficient length to estimate the 25-year discharge. The estimates of the 1.5-, 2-, 5-, 10-, and 25-year events were generally similar for both the Weibull and Log-Pearson III distributions (fig. 6). Stepwise, forward, and backward multiple regression techniques (SAS) were all used to help develop the "best" models to predict peak discharge (Y) for each of the return intervals using a variety of watershed descriptive variables (X_i). The independent GIS-area based variables (X_i) consisted of: 1) mean annual watershed precipitation, 2) seasonal flow, 3) watershed area, 4) average watershed slope, 5) average watershed elevation, and 6) the percent of watershed area in forest. In a parallel analysis, precipitation (P_2) estimated as the mean annual value for the mapping point closest to the gauging station was substituted for the GIS area-based estimate (P). We also substituted the USGS published estimate of watershed area (A_2) for the GIS derived estimate of area (A). The

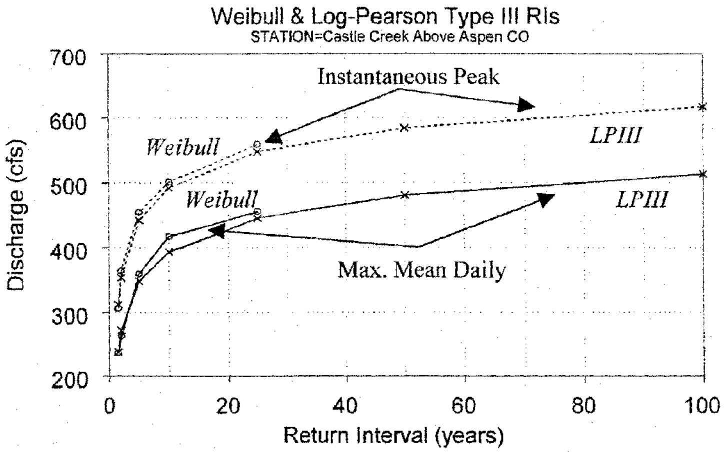

Figure 6. Log-Pearson Type III and Weibull plotting position return intervals calculated using instantaneous maximum flows and maximum mean daily flows.

purpose for these substitutions in the analysis was two-fold. First we wished to document the effect, if any, of refining the estimates of area and precipitation using the GIS database. Second, equations that can be driven with more readily available data such as precipitation estimated from the nearest gauge (point) and area estimated from USGS quadrangle sheets would be easier and more cost effective to implement, especially if there is no loss in the resolution of prediction.

RESULTS

Models were fit to estimate either the Log-Pearson Type III or the Weibull plotting position peak flows for the 1.5-, 5-, 10-, and 25-year events for both the maximum instantaneous discharge (table 1) and the maximum mean daily discharge (table 2) for the annual series. The Log-Pearson Type III distribution is commonly used but oftentimes extends well beyond the data, or what we "know". The Weibull plotting position on the other hand is empirically based, and represents a direct reflection of the available data. We chose to work with both recurrence intervals. Although we attempted to use a suite of independent variables to find both the Log-Pearson and Weibull recurrence intervals, generally, only precipitation (P) and area (A) proved to be significant ($p = 0.05$) and were retained in the final prediction models. As noted earlier, point estimates for precipitation (P_2) and map estimates of area (A_2) were substituted for the GIS derived estimates of those parameters and the models refit (table 3). A comparison of the model fit using either estimate of precipitation and area for the Log-Pearson Type III Peak discharge showed that the coefficients of regression appear to differ for the intercept (b_0), and the exponents on the precipitation (b_1) and area (b_2), but the error terms for both sets of equations are similar. There is little difference in relative standard error (SE) of the models, which is the square root of the model mean squared error divided by the mean discharge, and the relative SE of the individual and mean predictions, which are calculated as the square root of the average SE of prediction divided by the mean discharge. To utilize the equations in table 3, precipitation and area are simply applied to the appropriate equation. As an example, if average annual precipitation is 30 inches and area is 50 square miles, then the 1.5-year Log-Pearson Type III return interval flow would be calculated as: $Q_{1.5}=1.179(\mathbf{30})^{1.019}(\mathbf{50})^{0.720}=630.894$ cubic feet per second per square mile (cfs).

Table 1. Models developed to predict both Log-Pearson Type III and Weibull plotting position instantaneous peak discharges using GIS area-based input variables.

RI (yrs)	N	Discharge Model	Mean Q (cfs)	Model SE (%)	Avg. SE of Mean Pred (%)	Avg. SE of Indiv Pred (%)
Log-Pearson Type III:						
1.5	30	$Q=1.101(P)^{1.098}(A)^{0.591}$	305	34	10	36
5	30	$Q=1.273(P)^{1.148}(A)^{0.638}$	497	31	9	33
10	30	$Q=1.052(P)^{1.227}(A)^{0.658}$	584	29	8	31
25	30	$Q=0.726(P)^{1.356}(A)^{0.683}$	692	27	8	29
Weibull Plotting Position:						
1.5	30	$Q=0.886(P)^{1.172}(A)^{0.577}$	302	32	9	33
5	30	$Q=1.069(P)^{1.189}(A)^{0.648}$	500	32	9	34
10	30	$Q=1.523(P)^{1.133}(A)^{0.649}$	589	31	9	32
25	18	$Q=0.941(P)^{1.339}(A)^{0.629}$	790	22	9	24

Table 2. Models developed to predict both Log-Pearson Type III and Weibull plotting position maximum mean daily discharges using GIS area-based input variables.

RI (yrs)	N	Discharge Model	Mean Q (cfs)	Model SE (%)	Avg. SE of Mean Pred (%)	Avg. SE of Indiv Pred (%)
Log-Pearson Type III:						
1.5	30	$Q=0.776(P)^{1.090}(A)^{0.629}$	240	29	8	30
5	30	$Q=1.116(P)^{1.060}(A)^{0.683}$	380	28	8	29
10	30	$Q=1.162(P)^{1.076}(A)^{0.694}$	436	27	7	28
25	30	$Q=1.149(P)^{1.111}(A)^{0.702}$	500	26	7	27
Weibull Plotting Position:						
1.5	30	$Q=1.206(P)^{0.963}(A)^{0.626}$	238	28	8	29
5	30	$Q=0.743(P)^{1.164}(A)^{0.698}$	385	27	7	28
10	30	$Q=1.060(P)^{1.098}(A)^{0.702}$	442	29	8	30
25	18	$Q=36.191(A)^{0.822}$	598	28	8	29

Table 3. Final models developed to predict both the Log-Pearson Type III and Weibull plotting position instantaneous peak and maximum mean daily flows using more readily available input variables.

RI (yrs)	**N**	**Discharge Model**	**Mean Q (cfs)**	**Model SE (%)**	**Avg. SE of Mean Pred (%)**	**Avg. SE of Indiv Pred (%)**
Log-Pearson Type III instantaneous peak discharges:						
1.5	30	$Q=1.179(P_2)^{1.019}(A_2)^{0.720}$	305	32	9	34
5	30	$Q=3.034(P_2)^{0.848}(A_2)^{0.745}$	497	34	10	35
10	30	$Q=5.836(P_2)^{0.692}(A_2)^{0.747}$	584	36	10	37
25	30	$Q=14.482(P_2)^{0.460}(A_2)^{0.746}$	692	39	11	41
Weibull Plotting Position instantaneous peak discharges:						
1.5	30	$Q=2.076(P_2)^{0.878}(A_2)^{0.686}$	302	35	10	36
5	30	$Q=2.726(P_2)^{0.871}(A_2)^{0.756}$	500	35	10	37
10	30	$Q=3.930(P_2)^{0.810}(A_2)^{0.754}$	589	34	10	35
25	18	$Q=51.389(A_2)^{0.803}$	790	27	8	28
Log-Pearson Type III maximum mean daily discharges:						
1.5	30	$Q=1.563(P_2)^{0.834}(A_2)^{0.739}$	240	32	9	33
5	30	$Q=2.436(P_2)^{0.786}(A_2)^{0.785}$	380	31	9	33
10	30	$Q=3.059(P_2)^{0.752}(A_2)^{0.790}$	436	31	9	33
25	30	$Q=3.946(P_2)^{0.713}(A_2)^{0.792}$	500	32	9	33
Weibull Plotting Position maximum mean daily discharges:						
1.5	30	$Q=2.647(P_2)^{0.691}(A_2)^{0.717}$	238	31	9	32
5	30	$Q=2.167(P_2)^{0.807}(A_2)^{0.801}$	385	33	9	34
10	30	$Q=2.541(P_2)^{0.798}(A_2)^{0.804}$	442	33	9	35
25	18	$Q=2.530(P_2)^{0.804}(A_2)^{0.856}$	598	22	8	24

The suites of equations developed in this analysis apply to all watersheds above 7500 feet in elevation in Colorado. The USGS equations (Kircher et al. 1985, Jennings et al. 1994) were regional. A comparison of models, for similar return intervals, predicting the Log-Pearson Type III peak discharge developed as part of this effort and those of Kircher et al. (1985) for the Rio Grande hydrologic region is presented in table 4. The regression coefficients are somewhat similar but the errors terms are much smaller for the Colorado models. The implication is that although screening the data set reduced the sample size, model fit improved because extraneous variability was reduced. We made no attempt to compare predictions using both sets of equations, we are simply reporting the reduction in prediction error that can be achieved by screening the data prior to model fitting. The other benefit of the new models is applicability. Rather than being restricted to a specific hydrologic region, these models can be applied over a broader geographic area.

Table 4. Model Comparisons using Log-Pearson Type III instantaneous peak discharges.

	*USGS Rio Grande		Colorado	
RI (yrs)	Model	**SE (%)	Model	SE (%)
5	$0.229\ (P)^{1.55}\ (A)^{0.777}$	61	$1.273\ (P)^{1.148}\ (A)^{0.638}$	33
10	$0.487\ (P)^{1.40}\ (A)^{0.760}$	58	$1.052\ (P)^{1.227}\ (A)^{0.658}$	31
25	$1.060\ (P)^{1.25}\ (A)^{0.742}$	57	$0.726\ (P)^{1.356}\ (A)^{0.683}$	29

*From Kircher et al., 1985.
** Percent standard error varied from 42-63 for various regions and return intervals.

The hypothesis of a greater range of applicability was tested on an independent set of data consisting of streamflow record from both Colorado and Wyoming. As noted earlier, there were 11 stream gauges in Wyoming that expressed stable, long-term stream records. In addition there were 28 other sites in Colorado for which there appeared to be a stable record but for which we could not obtain the GIS related information necessary to be included in the initial part of this study. A gage estimate of annual precipitation and the mapped area of each of the watersheds are available for these 39 additional watersheds. The point-based equations presented in table 3 were used to predict for the Log-Pearson Type III 1.5- and 25-year instantaneous maximum discharges for each of the 39 test watersheds. The estimated discharge was then plotted against the actual Log-Pearson value for each of the watersheds (fig. 7 and fig. 8). In general the predictions were quite good and within the range in error we expected when compared to the 1:1 line, implying equality of prediction and observation. A similar relationship exists across all return intervals and with the maximum mean daily discharge.

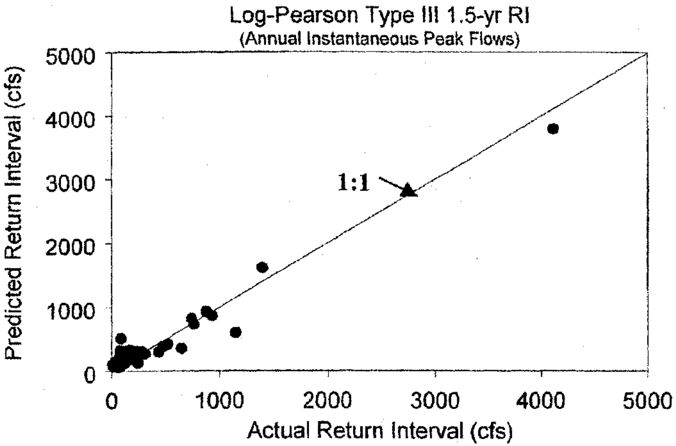

Figure 7. Actual vs. predicted 1.5-year Log-Pearson Type III return interval calculated using the annual instantaneous peak flows for watersheds in Wyoming and Colorado that were not used to develop the models.

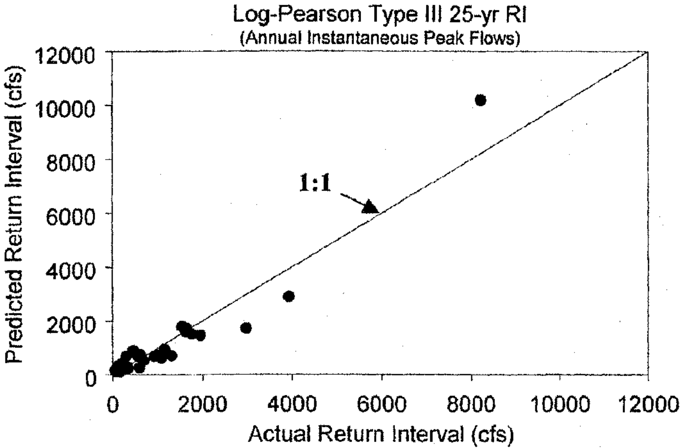

Figure 8. Actual vs. predicted 25-year Log-Pearson Type III return interval calculated using the annual instantaneous peak flows for watersheds in Wyoming and Colorado that were **not** used to develop the models.

Work in progress shows that a comparison of the 1.5-, 10-, and 25-year USGS equations for the southwest region and the corresponding statewide equations (table 3), show the predicted flows from the two sets of equations are not substantially different from one another. This means that these statewide models, developed using only 30 watersheds from all across Colorado, can predict flow in the southwest region just as well as the USGS southwest region equations which were developed using 40 watersheds from within the southwest region. Comparisons have not yet been made between the other regional USGS equations and the statewide equations.

SUMMARY

Frequently, data collected for one purpose is used for many other purposes. Sometimes the application is appropriate, sometimes the application is less appropriate. Often times the problem of appropriateness lies not as much with the quality of the data as it does with the application of the data. The USGS streamflow record used in this analysis was collected for a number of reasons from a variety of watersheds, not all of which lend well to the application we chose to make of the resulting information. The screening of the data set and the subsequent analysis reported in this study focused on the use of the data for a specific purpose; to develop prediction models for peak discharges. A different purpose, or application, might have required a different set of selection criteria and probably would have resulted in a different acceptance/rejection outcome. We were primarily concerned with peak discharges. As such, we conclude that screening the data set to eliminate those stations that demonstrate an inconsistent, erratic, and unexplainable deflection in peak discharge response results in a data set better suited to the intended application. For example, streams that are diverted or augmented may be well monitored, the discharge may be properly estimated, but the discharge metric may not represent a natural response to the existing watershed conditions. Inclusion of that data in the model fitting process will bias the resulting model and therefore should not be included. The same logic applies to other disturbance factors. Screening the data set appears to have reduced variation in the dependent variable resulting in an improved fit in the final models even though the data set is greatly reduced.

As part of this analysis we also attempted to look at the degree of improvement in prediction that can be achieved by improving the technology used in quantifying the independent variables. Using a GIS platform did not appear to result in improved parameter estimates, at least to the degree that the resulting models were improved. The use of GIS derived databases increased the effort expended in the data generation process and decreased the general utility of the resulting models because of the resulting input requirements. However, the point-based models developed as a result of this effort appear to be not only reliable, but more broadly applicable than their predecessors.

REFERENCES

Anderson, H.W., 1955, Detecting hydrologic effects of changes in watershed conditions by double-mass analysis, Trans. Amer. Geophys. Union, v. 36, pp. 119-125.

Burton, Timothy A., 1997, Effects of basin scale timber harvest on water yield and peak streamflow, Journal of the Am. Water Resour. Assoc. 33(6):1187-1194.

Jennings, M.E., Thomas, W.O. Jr., and Riggs, H.C., 1994, Nationwide summary of U.S. Geological Survey regional regression equations for estimating magnitude and frequency of floods for ungaged sites, 1993: U.S. Geological Survey Water-Resources Investigations Report 94-4002.

Kircher, J.E., Choquette, A.F., and Richter, B.D., 1985, Estimation of natural streamflow characteristics in western Colorado: U.S. Geological Survey Water Resources Investigation Report 85-4086.

SAS Institute Inc., SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2, Cary, NC: SAS Institute Inc., 1989. 846 pp.

Troendle, C.A. and Leaf, 1980, Hydrology, *In*: An Approach to Water Resource Evaluation of Non-Point Silvicultural Sources, EPA 60018-80-012, Envir. Res. Lab., Athens, Georgia, pp. III-1 to III-173.

Troendle, C.A. and Stednick, J.D., 1999, Discussion of "Effects of basin scale timber harvest on water yield and peak streamflow" by Timothy A. Burton, Journal of the Am. Water Resour. Assoc. 35(1):177-181.