The floods of October 27–30, 1950, were confined to southwestern Oregon and northwestern California. These floods were almost entirely the result of rainfall, as little snow had accumulated so early in the season. Generally, the floods resulting from this storm were not the greatest known in the area, though peak discharges on the Smith and Umpqua Rivers may have been as great as the flood of 1861 (Paulsen, 1953).

The floods of January 17–21, 1953, affected all of western Oregon, but the most serious flooding occurred in southwestern Oregon and northwestern California. Peak discharges were generally greater than for the floods of October 1950. Snowmelt was not a factor in flooding in the Coast Range, but contributed to flooding on streams heading in the Cascades (Rantz, 1959).

A series of storms from December 1955 to January 1956 caused widespread flooding in most of California, western Nevada, western Oregon and parts of Idaho. In Oregon, the Willamette River and its tributaries and all coastal rivers were affected. Warm temperatures and rain at high elevation melted much of the accumulated snowpack, resulting in record-breaking streamflows for many streams (Hofmann and Rantz, 1963). Recurrence intervals varied from 10 to 50 years, depending on location (Hubbard, 1991).

Heavy rains falling over southwestern Oregon on December 2, 1962, caused severe flooding in some areas of the Rogue Valley (Taylor and Hatton, 1999). Recurrence intervals for peak discharges for the two forks of Ashland Creek exceeded 30 years, for the Rogue River at Raygold, 25 years, and for the South Fork Little Butte Creek, 100 years.

The storm of December 19–23, 1964, was extreme. The recurrence interval for floods in some areas was in excess of 100 years (Hubbard, 1991). All of Oregon, northern California, and parts of Idaho and southern Washington were affected. Peak discharges were substantially increased due to warm rain falling on accumulated snow. Many areas of the State experienced severe flooding. Flooding in the Willamette Valley, however, was significantly reduced because of the flood control reservoirs built in the previous two decades. The peak discharge on the Umpqua River at Elkton of 265,000 cfs exceeded the 1861 peak discharge of 220,000 cfs (Waananen and others, 1971).

The flood of January 1972 affected a limited area in the lower Willamette Valley, the Sandy River, and rivers of the northern Oregon coast. Peak discharges on some coastal streams exceeded those of the December 1964 flood. Recurrence intervals varied from 10 to 100 years for affected streams (Hubbard, 1991).

During January 13–17, 1974, a series of storms with mild temperatures and intense rain followed a period of heavy snow and freezing rain (Taylor and Hatton, 1999). The resulting snowmelt and rapid runoff caused widespread flooding in western Oregon. Recurrence intervals for peak discharges on several steams in the Umpqua and Rogue River Basins exceeded 50 years, with the West Branch of Elk Creek well in excess of 100 years.

Heavy rain fell over much of Oregon February 22–23, 1986. The rain combined with melting snow to bring flooding to many areas (Taylor and Hatton, 1999). In the Sandy River Basin, many streams had peak discharges with recurrence intervals from 10 to 30 years. The recurrence interval for the Middle Santiam River exceeded 80 years.

The storm of January 9–11, 1990, affected coastal streams of northwest Oregon and parts of southwestern Washington. Flooding was exacerbated by high tides and high winds. Recurrence intervals ranged from 25 to 100 years (Hubbard, 1996).

During the period February 5–9, 1996, warm temperatures and intense rain falling on a deep snowpack combined to create severe flooding throughout the northern part of Oregon (Taylor and Hatton, 1999). In many areas, flood magnitudes were generally comparable to or greater than those of the 1964 flood. The peak on the Nehalem River near Foss was the greatest on record, greatly exceeding the 100 year event. In the Willamette Valley, flood control reservoirs minimized flooding.

From November 18–20, 1996, warm, moist air from the tropical Pacific brought record-breaking precipitation to much of Oregon (Taylor and Hatton, 1999). Melting snow exacerbated flooding in some areas. The recurrence interval for the flood peak for the South Fork Coquille River was nearly 50 years and for the Chetco River nearly 70 years. Recurrence intervals for many streams in the interior valleys and the Cascades were on the order of 10 to 30 years.

From December 30, 1996, to January 5, 1997, warm moist air from the subtropical Pacific passed over the entire northwest (Taylor and Hatton, 1999). Heavy rain, warm temperatures, and rapid snowmelt caused flooding over much of the region. In western Oregon, estimated recurrence intervals in a few areas in the south exceeded 15 years. Hard hit was the town of Ashland, which experienced severe flooding. The flood was extreme, but its recurrence interval at Ashland is unknown.

Magnitude and Frequency Analysis

For a site where peak discharges have been systematically measured, the magnitude of peak discharges can be related to frequency by fitting the observed peaks to a theoretical probability distribution. From the probability distribution, the magnitude of the peak discharge for any return interval can be estimated. In practice, however, it is seldom reasonable to make estimates of flood magnitudes for return intervals greater than about 500 years.

For this study, the logarithms of annual series of peak discharges at 376 streamflow gaging stations in western Oregon, southwestern Washington, and northwestern California (Appendix A) were fitted to the Pearson Type III distribution following guidelines established by the Interagency Advisory Committee on Water Data (1982). These guidelines are commonly known as Bulletin 17B. Where the logarithms of the annual peak discharges are used, the fitted Pearson Type III distribution is referred to as the log-Pearson Type III distribution.

The log-Pearson Type III probability distribution requires three parameters: the mean, standard deviation, and skew[1] of the logarithms of the annual series of peak discharges being fitted. The parameters define a smooth trend line through the observed peak discharges when plotted on a log-probability plot (i.e., the logarithm of the magnitude of each annual peak discharge plotted against its probability of occurrence). However, some peak discharges do not fit the general trend of observed peak discharges. Because the data are ranked, these outliers always occur at the high or low ends of the distribution. The log-Pearson Type III distribution usually cannot fit both the general trend of the observed peak discharges and the outliers. This distorted fit typically does a poor job of representing the high end of the distribution and may significantly over- or under-estimate the largest peak discharges.

Following procedures recommended in Bulletin 17B, the parameters of the log-Pearson Type III distribution are adjusted for the effects of high and low outliers as well as for historic peaks, for zero-flow peaks[2], and for peaks below the gage threshold. It is beyond the scope of this report to discuss these adjustments, but for those interested, they are treated in detail in Bulletin 17B.

Even after adjustment for outliers, the station skew value may be poorly defined for short record gaging stations. A better estimate of the skew coefficient is obtained by taking a weighted average of the adjusted station skew and a "generalized" skew based on the skew coefficients for long-term stations in the area.

Although generalized logarithmic skew coefficients for the United States are provided with Bulletin 17B, Bulletin 17B recommends that generalized skew coefficients be developed for each area of concern. If the newly developed generalized skew coefficients have a mean squared error less than that of the generalized skew coefficients provided by Bulletin 17B, the newly developed skew coefficients should be used in lieu of those provided in Bulletin 17B.

Generalized skew coefficients for Oregon were developed as part of this study. These generalized skew coefficients were combined with station skew values to obtain weighted skew estimators for each station. The weighted skew values were used in fitting the Pearson Type III distributions. These topics are discussed in detail later in the report.

In general, fitting the theoretical Pearson Type III distribution to the logarithms of the observed peak discharges was straightforward and produced good results. Of the 376 gaging stations, 181 required adjustment for high or low outliers or historic or zero peak discharges. Peak discharge statistics used in fitting the distribution for the gaging stations are listed in Appendix B. The statistics include length of record; number of historical peaks; user-defined high and low-outlier thresholds; number of high and low outliers; number of zero flow peaks and peaks below the gage threshold; the station, Bulletin 17B, generalized, and weighted skews; and the statistics from the trend analysis. The meaning and significance of these statistics can be found in Bulletin 17B.

A visual check of the "goodness of fit" of the theoretical Pearson Type III distribution to the logarithms of the annual peaks was made for each of the 376 gaging stations. Eight gaging stations originally considered for inclusion in this analysis were rejected based on this visual check. The fitted distributions did not reasonably approximate the actual distribution of observed peak discharges.

To make the check, the theoretical distribution and the observed peaks are plotted on a log-probability plot. (Appendix C discusses how the plotting position was determined for the probability axis.) The log-Pearson Type III distribution generally plots as a curved line. The sense and degree of curvature is determined by the skew coefficient. Curvature is concave upward when the skew coefficient is positive and concave downward when it is negative. When the skew coefficient is zero, the distribution plots as a straight line. An example plot for the gaging station on the Nehalem River is shown in figure 4. Note the low outlier.

Peak discharge magnitudes at selected frequencies are obtained from the log-Pearson Type III distribution by this equation:

$$\log Q = \overline{X} + KS \tag{1}$$

where

$\overline{X}$ = the mean of the logarithms of the peak discharges,
K = a factor that is a function of the skew coefficient of the logarithms of the peak discharges and the selected frequency, and
S = the standard deviation of the logarithms of the peak discharges.

Values of K can be obtained from Appendix 3 of Bulletin 17B. The table requires the skew coefficient and the frequency. The 2-, 5-, 10-, 25-, 50-, 100-, and 500-year peak discharges for the 376 gaging stations are listed in Appendix D.

Peak Discharge Data

The data used in this study are the annual series of peak discharges for the 376 gaging stations. An "annual series" of peaks represents the largest instantaneous peak for each water year of record, reported in cubic feet per second. Peaks were measured at both continuous record sites and at crest-stage gage sites that record only the maximum peak discharge for

[1] The mean is a measure of the central tendency of the distribution, the standard deviation is a measure of the dispersion of the distribution about the mean, and the skew is a measure of the asymmetry of the distribution. A distribution with a skew of zero is symmetrical.

[2] For some watersheds, in some years, there is no streamflow. The annual peak discharge in those years is zero.

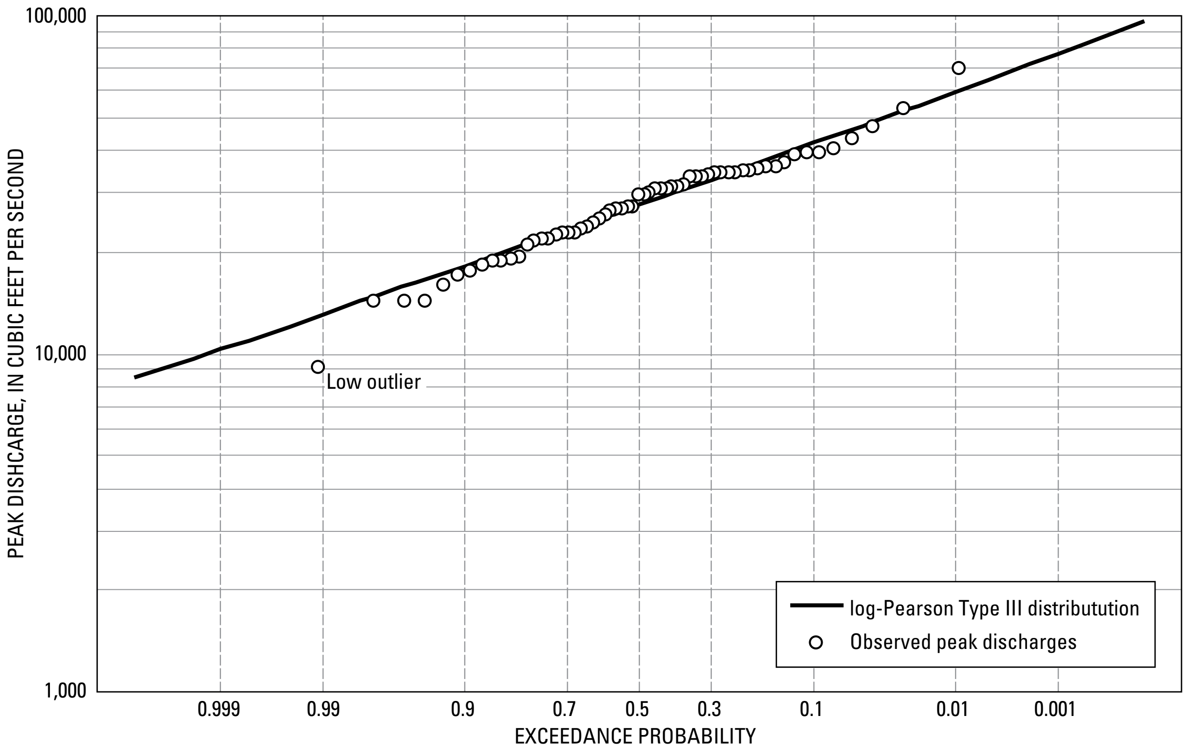


Figure 4. Log-Pearson Type III distribution fitted to the logarithms of peak discharges for the gaging station Nehalem River at Foss, Oregon (14301000).

each year. These measurement sites represent watersheds not significantly affected by reservoir operations, diversions, or urbanization. All sites have 10 or more years of measured peak discharges through water-year 2001. The peak discharges used in this study were measured and reported by the U.S. Geological Survey and the Oregon Water Resources Department. All peak discharge data used in the analysis are available from the Oregon Water Resources Department (*webmaster@wrd.state.or.us*), and all peaks except those originating with the Oregon Water Resources Department are available from the U.S. Geological Survey (*info-or@usgs.gov*).

Quality Assurance

No effort was made to directly check the accuracy of peak discharges reported for the various gaging stations. It was assumed that adequate checks were made by the agency responsible for the peak estimates. However, a few scriveners' errors were discovered during the analysis. Unusual results in fitting the probability distributions or in doing the regression analysis were sometimes the result of erroneous peak values. In the first case, erroneous peaks caused the absolute value of the skew parameter of the distribution to be large. In the second case, erroneous peaks lead to large residuals in forming the prediction equations. In both cases, the observed peaks were examined for errors and corrected as necessary.

Assumptions of the Magnitude and Frequency Analysis

Assumptions of the magnitude and frequency analysis are that the peaks in any systematic series are random, and that they are all derived from the same population. These assumptions mean (1) that the value of one peak does not depend on the value of a preceding peak and (2) that all peaks arise from the same processes, e.g., as the result of rain from a frontal-storm as opposed to rain from a convective storm or as the result of snowmelt. Implicit in the second assumption is that the processes are not changing in time. For example, it is assumed that weather may vary from year to year, but that climate is not steadily getting wetter or drier, or warmer or colder. Other factors are also assumed to remain constant; that land use, for example, does not change substantially over the period the observations are made.

Test for Random Peaks

A usual test for randomness is to check each series of annual peaks for a statistically significant linear serial correlation, i.e., a trend (Thomas and others, 1993; Wiley and others, 2000). A significant trend suggests that systematic, nonrandom changes in peak discharge characteristics are occurring in time. A trend test is not definitive; it is cause for investigation, not necessarily for the elimination of a gaging station from the analysis.

Peak discharges from the 376 gaging stations were tested for linear correlation. The resulting information was analyzed in two ways: (1) to check for regional, climate dependent trends and (2) to check for local trends resulting from significant physical changes to a watershed. Local changes include human caused changes due to land use or water management as well as natural changes such as a volcanic eruption. Local trends that can be attributed to physical changes in the watershed may require all or part of a gaging station's period of record to be removed from consideration.

In the regional analysis, no consistent long-term trend was found, although there is evidence of a regional fluctuation of peak discharges between wet and dry periods. This fluctuation led to a higher than expected number of significant trends in long-term gaging station records. The evidence is too weak, however, to support a strong conclusion as to whether the fluctuation is truly periodic or what its period might be. Locally, no significant trend could be linked to physical changes in the associated watershed.

No gaging stations were eliminated from consideration based on the trend analysis. The details of the trend analysis are found in the Appendix E. It should be noted that watersheds known to be affected by regulation, significant diversion, urbanization, or the eruption of Mount St. Helens were not considered for the analysis.

Test for Mixed Populations

For some watersheds in western Oregon, more than one hydrologic process may generate peak discharges. While it is convenient to think of these processes as giving rise to distinct populations of peak discharges, the processes occur in unpredictable combinations and the populations overlap considerably. For example, rain-on-snow events probably form a continuum from pure rain to pure snowmelt.

For watersheds where more than one hydrologic process generates peak discharges, the log-Pearson Type III distribution may poorly fit the distribution of annual peak discharges. When plotted on a log-probability plot, a mixed population of peak discharges may show a sharp break in slope or a curve that reverses direction. The fitted distribution usually has a large skew coefficient. If the peak discharges are separated into homogeneous populations, log-Pearson Type III distributions fitted to the separate populations may be significantly different from one another. In these cases, the distributions may be combined by the method described by Crippen (1978).

Often, however, the distribution of a mixed population of peak discharges does not exhibit a break in slope or a curve that reverses direction. If the distribution is well approximated by a log-Pearson Type III distribution, and if each of the separate populations is well represented in the mixed population, then there is no benefit to dividing the peak discharges into separate populations. The log-Pearson Type III distribution fitted to the mixed population will be close to the composite distribution calculated from the separate populations (Advisory Committee on Water Information, 2002).

Log-probability plots of observed peaks for the 376 gaging stations used in this study were examined for sharp breaks in slope or a curve that reverses direction. Particular attention was given to high elevation gaging stations, where a mixed population of peak discharges is most likely to occur, and to distributions with large absolute values of skew. Distributions for only four gaging stations showed breaks in slope and none showed a curve that reverses direction. Other distributions had large absolute values of skew, but all had high or low outliers and were corrected to the extent possible by the procedures outlined in Bulletin 17B.

This result suggests that peak discharges for each gaging station may be treated as coming from the same population whether they do or not. A few examples will illustrate the point. Figure 5 shows the monthly distributions for four high elevation watersheds. The watersheds were selected because they have a mix of winter and spring peak discharges: Salmon River near Government Camp, Oregon (14134000), Oak Grove Fork above power plant intake, Oregon (14209000), Clearwater River above Trap Creek near Toketee Falls, Oregon (14314500), and Imnaha Creek near Prospect, Oregon (14331000). Note that the distributions are all bimodal.

Log-probability plots of the peak discharges for the four gaged watersheds are shown in figure 6. The peak discharges are identified as to their season of occurrence. Also shown is the log-Pearson Type III distribution fitted to the peak discharges. The distributions of the peak discharges for the four watersheds do not show breaks in slope or curves that reverse direction. The fitted log-Pearson Type III distributions are all reasonable.

The four distributions with breaks in slope, mentioned earlier, may represent mixed populations of peak discharges, but they were not treated as such. Instead, all peak discharges below the break were treated as low outliers and a conditional probability adjustment was made to the fit of the log-Pearson Type III distribution. This part of the analysis is discussed in detail in the next section.

Low Outliers and Mixed Populations

Bulletin 17B describes a statistical test to identify low outliers. This test usually does a good job of identifying low outliers, and the subsequent conditional probability adjustment satisfactorily improves the fit of the log-Pearson Type III distribution.

Sometimes more than one low peak discharge will fall outside the general trend of all peak discharges. Often in these cases, the statistical test will not identify all the low peak discharges as outliers even though they adversely affect the fit of the log-Pearson Type III distribution. These cases can be identified by a visual inspection of a log-probability plot of the fitted distribution and the observed peak discharges. Because these distributions have large negative skew coefficients, the fitted distributions have strong downward curves. Unless these low outliers are censored, the fit of the log-Pearson Type III

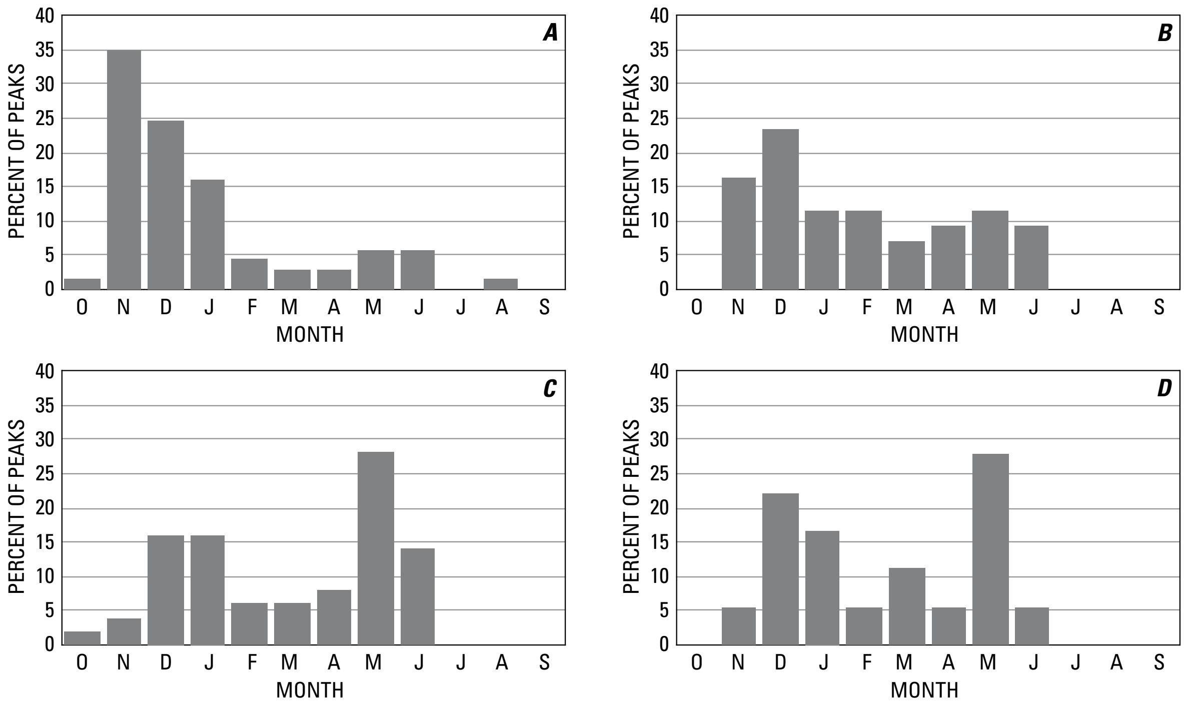


Figure 5. Distributions of the monthly occurrences of annual peak discharges for four gaging stations: (A) Salmon River near Government Camp, Oregon (14134000), (B) Oak Grove Fork above power plant intake, Oregon (14209000), (C) Clearwater River above Trap Creek near Toketee Falls, Oregon (14314500), and (D) Imnaha Creek near Prospect, Oregon (14331000).

distribution is compromised, with the upper end of the distribution poorly defined and often overestimated.

The Advisory Committee on Water Information (2002) offers suggestions on how to determine how many low peak discharges to censor. In general, low peaks are censored one at a time until the conditional probability distribution based on the remaining peaks stops changing significantly.

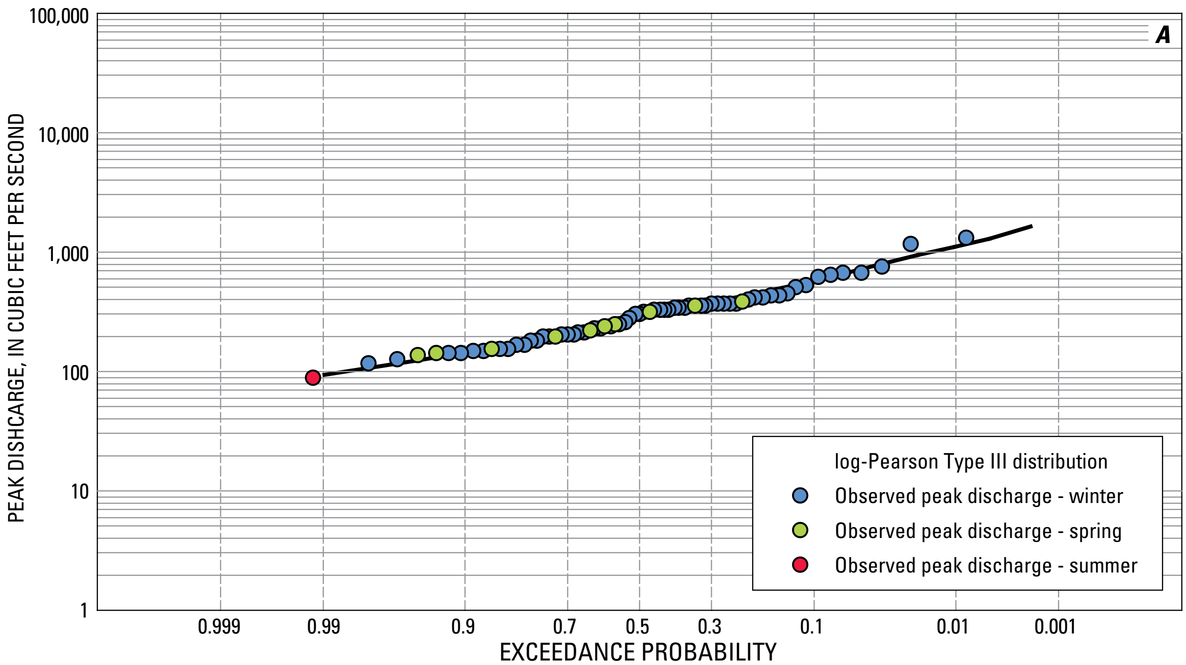
Not all low outliers are statistical outliers, however. In some cases, especially in drier areas, the low peak discharges may represent a separate population (Thomas and others, 1993). If there are a sufficient number of low peak discharges from this other population, the statistical test may not identify any of them as outliers, but the skew coefficient will have a large negative value, and a log-probability plot of the observed peaks will show a break in slope.

In many cases, it cannot be determined whether the low peaks come from a separate population or are statistical outliers. Even if these low peak discharges represent a separate population, it is not necessary to treat them as such. It is sufficient to treat them as low outliers. They do not provide any information about the *magnitude* of peak discharges at the upper end of the distribution; however, the low peaks do provide information about the *frequency*. The peaks below the low threshold are used with any zero peaks in a conditional probability adjustment as described in Bulletin 17B, Appendix 5.

Figure 7 shows examples of how the fit of the log-Pearson Type III distribution was improved for two gaging stations: (1) Blue River above Quentin Creek near Blue River, Oregon (14161000) and (2) Big Butte Creek near McLeod, Oregon (14337500).

For the Blue River gaging station, the outlier test identified one low outlier. The fitted distribution (dotted line) did not fit the observed peaks well, overestimating peak discharge at the high end. A visual inspection of the log-probability plot suggested that there were two additional low outliers. Increasing the low outlier threshold to 750 cfs censored these two peak discharges. The resulting fitted distribution (solid line) follows the observed peak discharges at the high end of the distribution much better than the distribution based on one outlier.

For the Big Butte Creek gaging station, the outlier test identified no low outliers. A visual inspection of the log-probability plot for the fitted and observed distributions shows a break in slope in the observed distribution, and the station skew has a large negative value (-0.632). The fitted distribution (dotted line) does not fit the observed distribution well, and it overestimates peak discharge at the high end. In this case, the low outlier threshold was increased to 2,000 cfs, sufficient to censor all low peak discharges below the break in slope (Thomas and others, 1993). The resulting fitted distribution (solid line) follows the observed peaks better above the break in slope and does not overestimate at the high end.



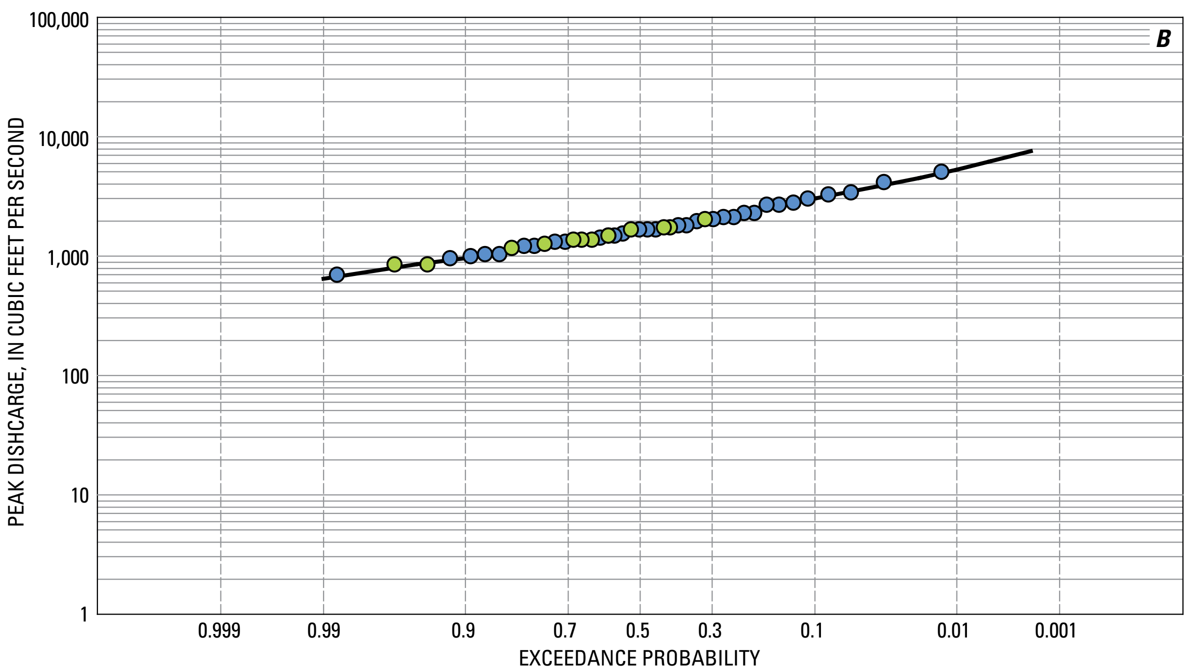


Figure 6. Log-probability plots for annual peak discharges for four gaging stations: (A) Salmon River near Government Camp, Oregon (14134000), (B) Oak Grove Fork above power plant intake, Oregon (14209000), (C) Clearwater River above Trap Creek near Toketee Falls, Oregon (14314500), and (D) Imnaha Creek near Prospect, Oregon (14331000). Peak discharges are identified as to their season of occurrence: winter is November to March, spring is April to June, and summer is July to September. Also shown are the log-Pearson Type III distributions that were fitted to the peaks.

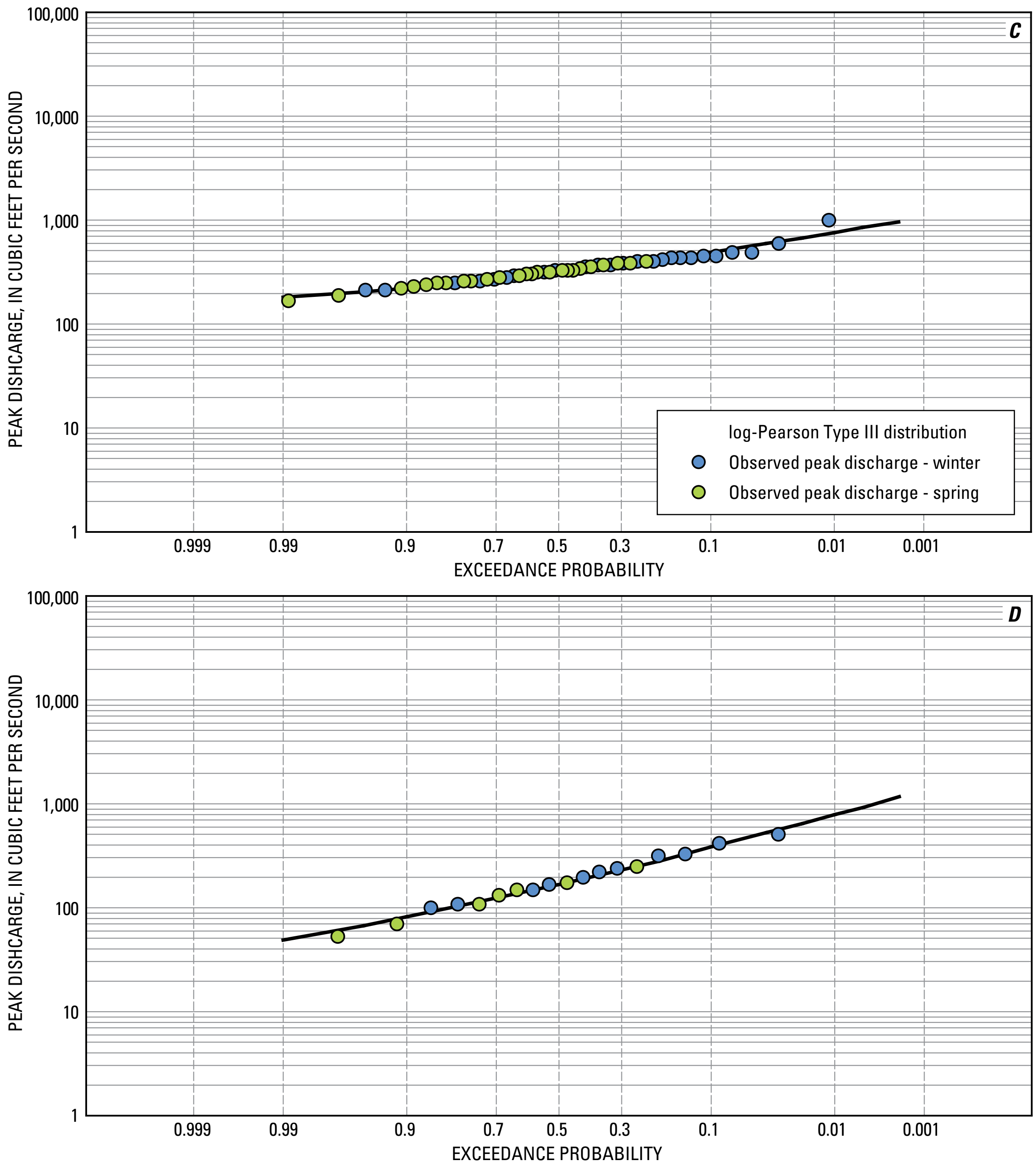


Figure 6. —Continued.

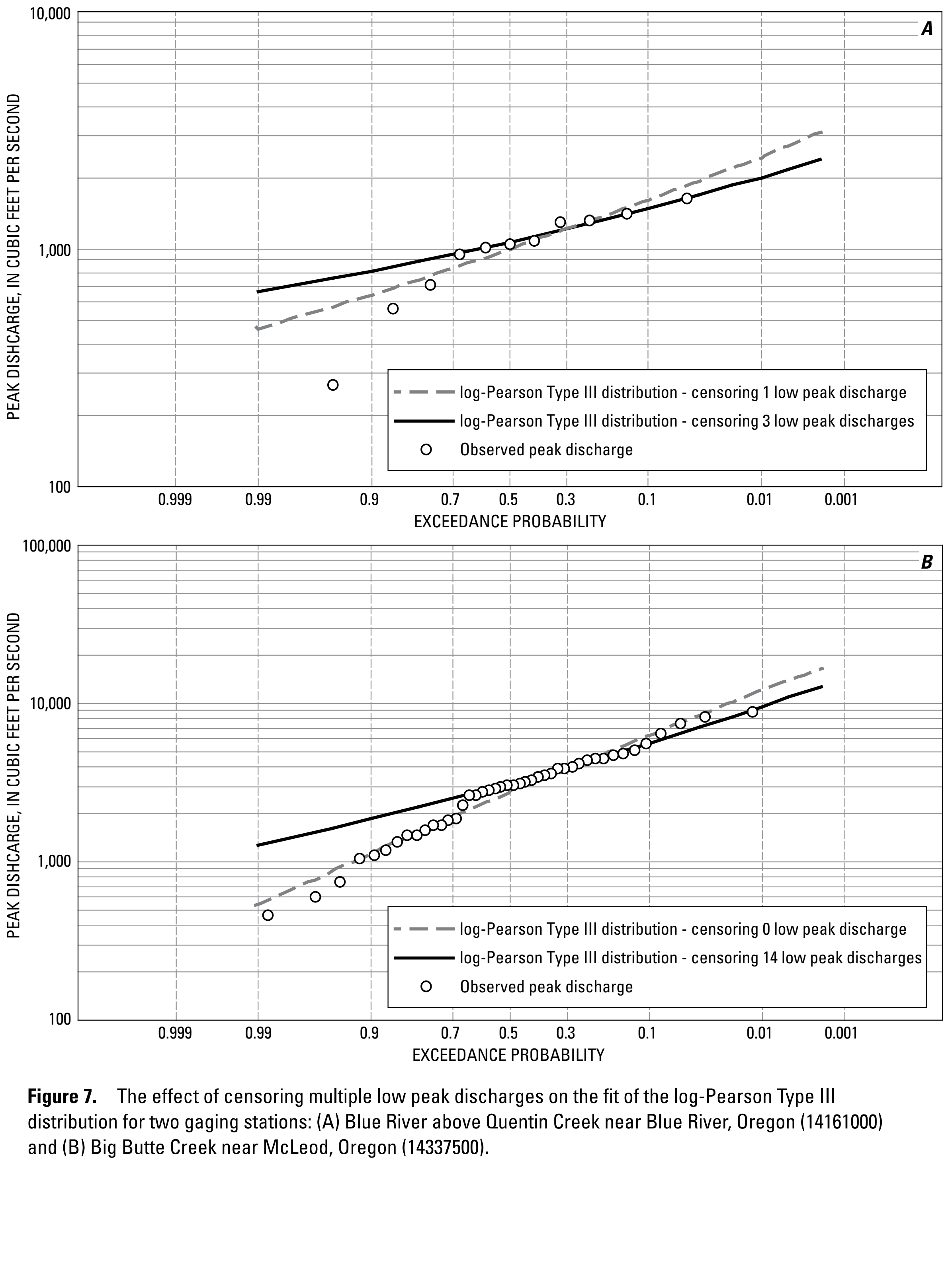


Figure 7. The effect of censoring multiple low peak discharges on the fit of the log-Pearson Type III distribution for two gaging stations: (A) Blue River above Quentin Creek near Blue River, Oregon (14161000) and (B) Big Butte Creek near McLeod, Oregon (14337500).

Generalized Skew

The skew coefficient of an annual series of peaks is sensitive to extreme values, especially when records are short. A more accurate estimate of the skew coefficient is obtained by weighting the station skew with a generalized skew value based on the skew coefficients of nearby long-term gaging stations. The weighting is based on the relative mean-square errors of the station and generalized skew and is given by this equation:

$$G_W = \frac{MSE_{\overline{G}}(G) + MSE_G(\overline{G})}{MSE_{\overline{G}} + MSE_G} \qquad (2)$$

where

G_W = weighted skew coefficient,
G = adjusted station skew,
$\overline{G}$ = generalized skew,
$MSE_{\overline{G}}$ = mean-square error of the generalized skew, and
MSE_G = mean-square error of the station skew.

Included with Bulletin 17B is a map of the generalized logarithmic skew coefficients of annual maximum streamflows for the entire United States. Although many peak discharge frequency studies use this map to obtain generalized skew values, Bulletin 17B recommends that users of the guide develop their own generalized skew coefficients for their area of interest using the procedures outlined in the bulletin.

Bulletin 17B outlines three methods for developing generalized skew coefficients: (1) drawing skew isolines on a map, (2) developing skew prediction equations, and (3) using the mean of station skew values. These generalized skews are to be developed using at least 40 stations with 25 or more years of record. The isoline map is drawn by hand from station skews plotted at the centroid of their watersheds. The prediction equations are developed to relate station skews to predictor variables that include the physical or climatological characteristics of the watersheds.

For this analysis, all three methods were tried. For the isoline method, rather than drawing the map by hand as suggested by Bulletin 17B, the map was drawn using GIS techniques, by the method described by Lumia and Baevsky (2000). How this method was adapted for this analysis is described in the next section. For the skew prediction equation method, useful equations could not be developed. There is not a good linear correlation between station skew and any of the available watershed characteristics.

The analyses were done statewide and were based on 267 gaging stations with more than 25 years of record in Oregon, southern Washington, western Idaho, northwestern Nevada, and northern California. The skews used in each analysis were the station skews adjusted for the effects of high and low outliers, zero peak discharges, and peak discharges below the gage threshold (see Bulletin 17B).

The isoline and average skew methods were evaluated based on a comparison of their mean-square errors to that of the generalized skew map provided with Bulletin 17B, the method with the smallest mean-square error being preferred. Mean-square errors for the isoline method and for the generalized skew map of Bulletin 17B were calculated by estimating the skew at each of the long-term stations by each method, squaring the difference between the station skew and the generalized skew, and taking the mean of the squared differences:

$$MSE = \frac{\sum_{i=1}^{n}(G_i - \overline{G_i})^2}{n} \qquad (3)$$

where

MSE = mean-square error,
G_i = station skew for gaging station i,
$\overline{G_i}$ = generalized skew for gaging station i,
n = number of stations.

For the method where the generalized skew is estimated as the mean of all station skews, the mean-square error is simply the variance of the station skews.

The mean-square error for the isoline method (MSE = 0.112) was significantly smaller than for either the mean of all stations skews (MSE = 0.222) or the generalized skew from Bulletin 17B (MSE = 0.302 for all of the United States or MSE = 0.227 for the area of the generalized skew analysis).

Developing Generalized Skew Isolines

Lumia and Baevsky's (2000) method assigns skew values to cells of a grid overlaid on the area of interest. The isolines are drawn from this grid. The grid values are estimated by a weighted average of the skews of nearby long-term gaging stations. The station skews, plotted at the centroids of their watersheds, are weighted by their distance from the grid cell and by their length of record. The closer the centroid of the watershed and the longer the station record, the more weight the station skew is given in the calculation. Lumia and Baevsky used the ARC/Info (Environmental Systems Research Institute, Inc., Redlands, California) routine GRID IDW to determine the skew value at each cell (Y.H. Baevsky, U.S. Geological Survey, written commun., 2001), and used that routine's default values for grid spacing, 10,000 meters, and number of stations, 12 (R. Lumia, U.S. Geological Survey, written commun., 2001). LATTICECONTOUR was used to determine the isolines.

This study also used these routines, however, the grid spacing and number of stations were varied to see the effect on the resulting skew isoline map. As the grid spacing decreases, the isolines become increasingly angular and blocky. As the number of stations decreases, the number of isolines increases and peaks and valleys appear around some stations. The gradients near these stations become increasingly steep.

The generalized skew map selected for this study was based on a grid spacing of 20,000 meters and 12 stations. The part of the map for western Oregon is shown in figure 8. This map was selected because it had the smallest mean square error while having skew isolines that are smooth and with no peaks or valleys. This map offers considerable improvement in mean-square error over either the generalized skew map provided by Bulletin 17B or the average of the skews of the 267 stations.

Figure 8 is provided for illustration only. A GIS (ARC/INFO) grid of the generalized skew coefficients may be obtained from the Oregon Water Resources Department (*webmaster@wrd.state.or.us*). It is recommended that generalized skew for a watershed be determined from this grid (using a GIS overlay analysis) rather than from a plotted map of generalized skew isolines.

Estimation of Magnitude and Frequency of Peak Discharges at Ungaged Sites

Peak discharges for an ungaged watershed may be estimated from prediction equations that relate peak discharge to climatologic and physical characteristics of the watershed (Thomas and Benson, 1969; Riggs, 1973). The predition equations are derived using multiple linear-regression techniques. This generalization or regionalization of peak discharges from gaged to ungaged watersheds is known as a "regional regression analysis."

For this study, a combination of regression techniques was used to derive the prediction equations. A preliminary analysis using ordinary least-squares regression was done to define flood regions of homogeneous hydrology and to determine which climatological and physical characteristics of the watersheds would be most useful in the prediction equations. The final prediction equations were derived using generalized least-squares regression (Tasker and others, 1986; Tasker and Stedinger, 1989). The computer model, GLSNET (version 2.5), developed by the U.S. Geological Survey (2000) was used for the generalized least-squares analysis.

Flood Regions

When using regression techniques to derive prediction equations, the accuracy of the equations may be improved by doing the derivations for regions of relatively uniform hydrology called, herein, flood regions. Three flood regions were defined for this study. In order to define these regions, a simple cluster analysis was used (Wiley and others, 2000). First, an ordinary least-squares regression was done using 100-year peak discharges as the response variable and drainage area as the only predictor variable. Then, the residuals from the regression were plotted at the centroids of their respective watersheds on a map of the study area. Clusters of residuals of similar sign and magnitude were presumed to indicate areas of similar hydrology and were defined as flood regions. This procedure was repeated for each flood region as it was defined until no clusters of residuals were apparent,

Immediately apparent from the plot of residuals was a line of large negative residuals along the crest of the Cascade Range (fig. 9). Assuming these large negative residuals to be related to elevation, all the residuals were plotted against the mean elevation of their corresponding watersheds.

Figure 10 shows that the relationships between residuals and mean watershed elevation above and below 3,000 feet are remarkably different. Below 3,000 feet, the residuals increase slightly with elevation. Above 3,000 feet, the trend reverses, and the residuals rapidly decrease with elevation. The model greatly over predicts at the highest elevations. The behavior of the residuals relative to elevation demonstrates the earlier observation that the hydrologic processes generating peak discharges above and below 3,000 feet are different.

The gaging stations for western Oregon were divided into two groups based on elevation, those above 3,000 feet and those below. In each group, the 100-year peak discharges were regressed on area and the residuals plotted. For the gaging stations above 3,000 feet, no clear groupings of residuals occurred. The gaging stations above 3,000 feet, then, represent one flood region.

The plot of residuals for gaging stations with mean watershed elevations below 3,000 feet showed large positive to slightly negative residuals west of the crest of the coastal mountains and large negative to slightly positive residuals in the remaining area. Based on this distribution of residuals, the gaging stations were divided into two groups, east and west of the crest of coastal mountains. For the gaging stations in each group, the 100-year peak discharges were regressed on drainage area and the residuals were plotted. As no clear grouping of residuals occurred in either group, the area associated with each group of stations was defined as a flood region and no further divisions were made.

The three flood regions in western Oregon are shown on figure 11. It is not possible, however, to show a boundary between watersheds with mean elevations above and below 3,000 feet. The 3,000-foot elevation contour is *not* the boundary. Consider a large watershed with mean elevation above 3,000 feet. It may contain subwatersheds with mean elevations less than 3,000 feet. An areally delineated region containing the large, high elevation watershed cannot also contain the smaller, lower elevation watersheds. This dilemma cannot be resolved on a map.

To facilitate identification and labeling of the regions, western Oregon first is divided into two regions: Region 1, west of the crest the coastal mountains, and Region 2, east of the crest of the coastal mountains. All of the gaged watersheds with elevations above 3,000 feet occur in Region 2. Region 2, then, is divided into two subregions, 2A and 2B, based