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

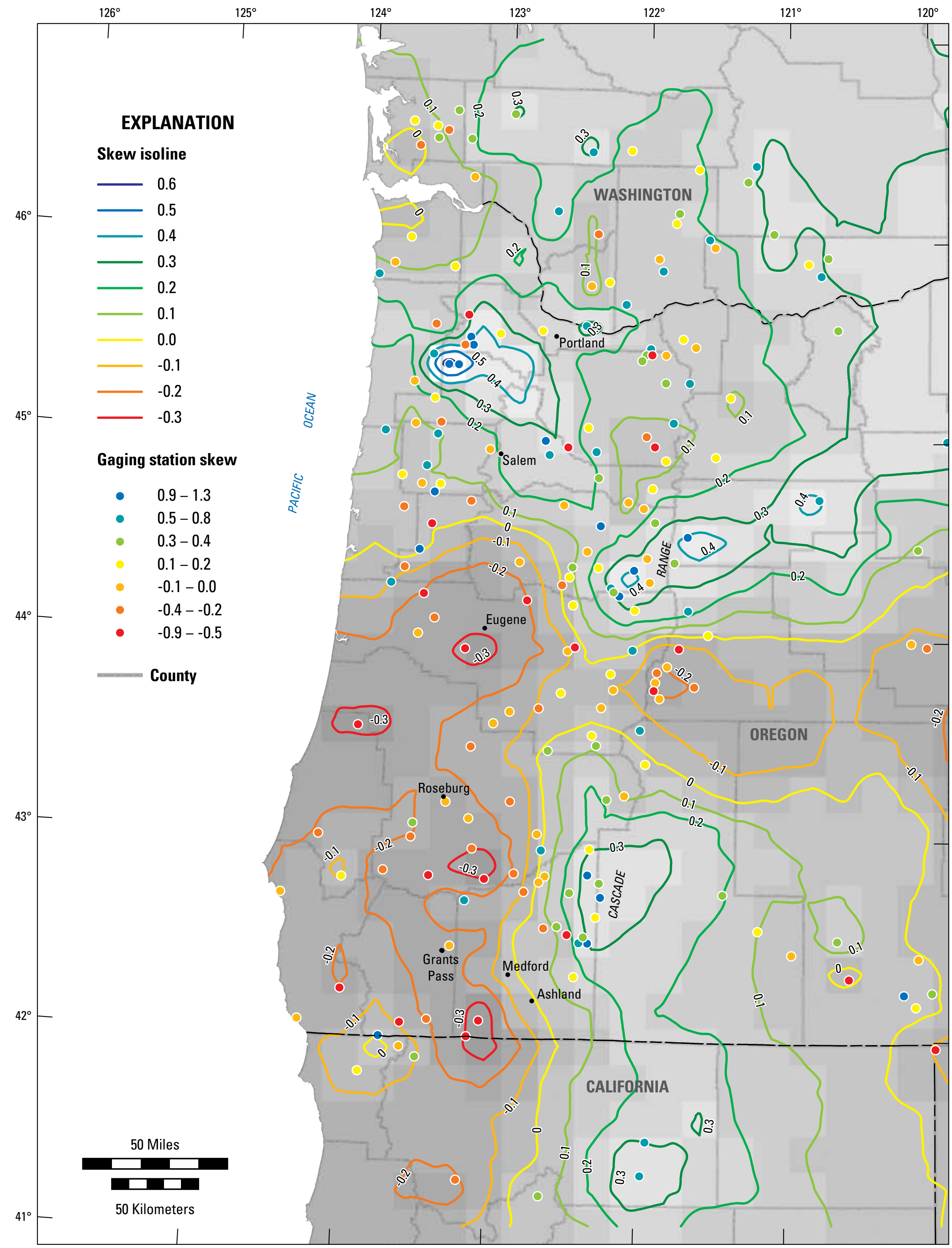

Figure 8. Generalized logarithmic skew coefficients for western Oregon. Isoline interval is 0.1. The colored circles represent skew coefficients of long-term gaging stations and are located at the centroids of their respective watersheds. The shaded background represents the Geologic Information System grid on which the isolines are based. Darker shades represent negative skews, and lighter shades, positive skews. The value of the skew cofficient for each grid cell was calculated as a weighted average of nearby gaging station skews.

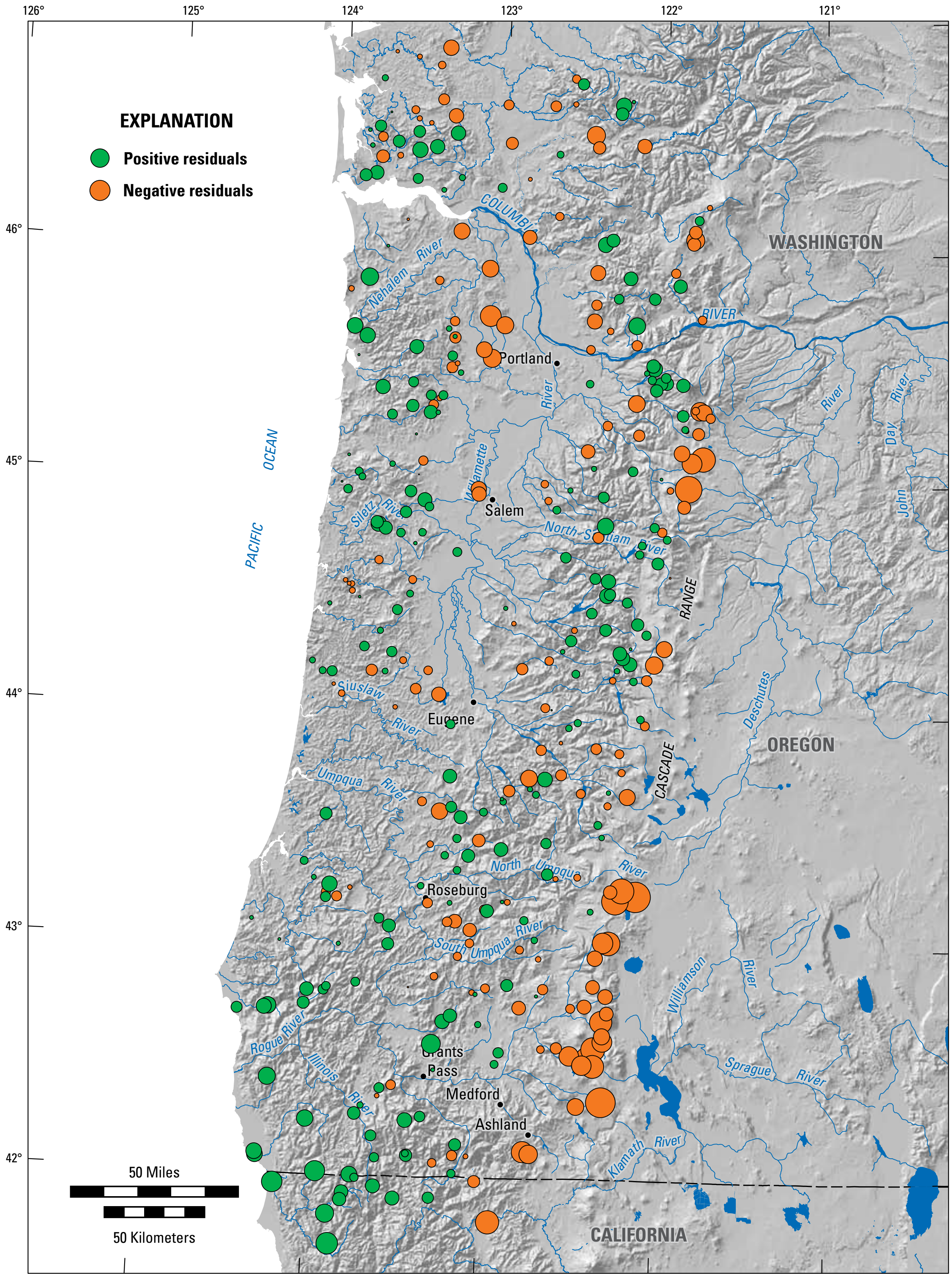

Figure 9. Residuals from a regression of 100-year peak discharges on watershed area for all gaging stations in western Oregon. The size of each circle is proportional to the absolute value of the residual. Negative residuals are shown in orange, and positive residuals in green. Note the line of large negative residuals along the crest of the Cascade Range.

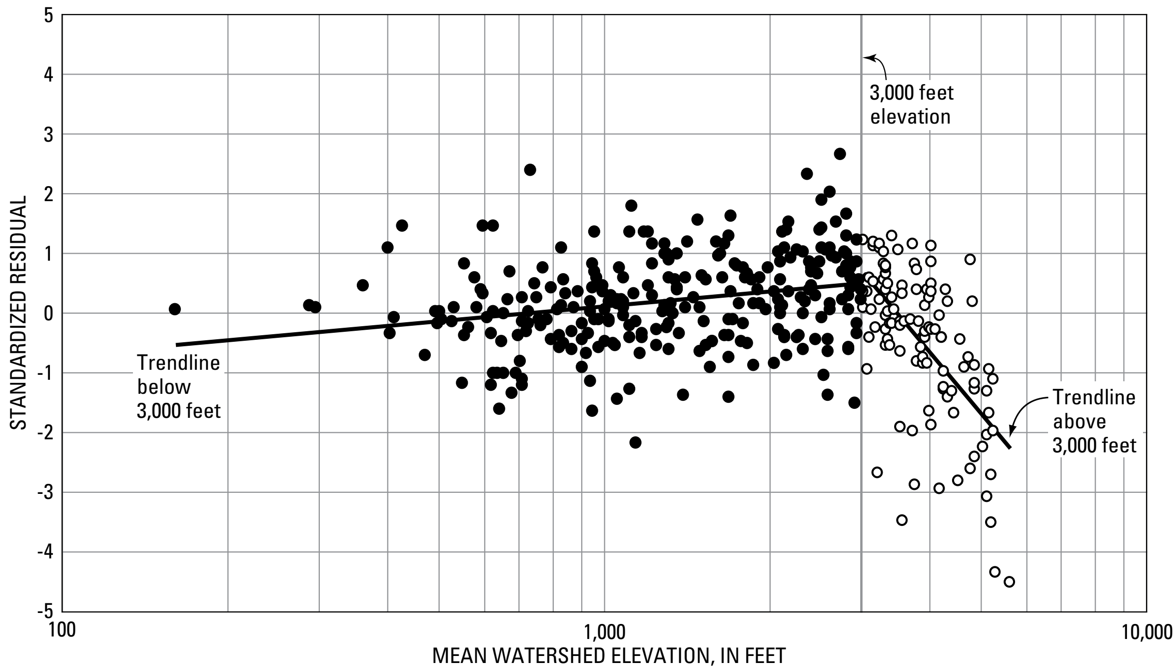

Figure 10. Relation of standardized residuals from a regression of 100-year peak discharges on watershed area to the mean elevation of their respective watersheds.

on mean watershed elevation. Region 2A represents gaging stations with mean watershed elevations above 3,000 feet, and Region 2B, gaging stations with mean watershed elevations below 3,000 feet. Although Regions 2A and 2B cannot be delineated on a map, the locations of the gaging stations associated with each region are shown to give a rough approximation of the areal extent of each region.

Watershed Characteristics

Ninety-two watershed characteristics were available for this study (Appendix F). For each gaging station, the 92 watershed characteristics were estimated using the GIS computer program ARC/INFO 7.2.1 (Environmental Systems Research Institute, Inc., Redlands, California).

In a GIS analysis of watershed characteristics, each characteristic is associated with either a coverage (vector data) or a grid (raster data). For this study, the elevation grid (digital elevation model) came from the National Center for Earth Resources Observation & Science (1999). The precipitation and temperature grids came from the Oregon Climate Service (G.H. Taylor, Oregon State Climatologist, written commun., 2000, 2001). The soils coverage came from the National Cartography and Geospatial Center (1994). The climatologic characteristic grids from the Oregon Climate Service were generated using PRISM (Daly and others, 1997). PRISM stands for Parameter-elevation Regressions on Independent Slopes Model.

To begin, each watershed was delineated from U.S. Geological Survey 1:24,000 scale topographic maps and digitized into a coverage of all watersheds. The locations of the outlet and the centroid, the area, and the perimeter of each watershed were calculated directly from this coverage. For other characteristics, the watershed coverage was over-laid on the respective watershed characteristic coverage or grid. Stream length and percent area of lakes and ponds were determined from an overlay of the hydrography coverage. Relief was calculated simply as the difference of the highest and lowest elevations in the watershed determined from the elevation grid. For all others, the value of the characteristic was calculated as its average over the area of the watershed. The GIS analysis of watershed characteristics was implemented using an Arc Macro Language script. The script is available from the Oregon Water Resources Department on request (*webmaster@wrd.state.or.us*).

Most of the 92 characteristics were not used in the regression analysis. Some of the characteristics, such as the location of the centroid of a watershed, perimeter length or minimum watershed elevation, are poorly (or not at all) related to peak discharges. Others, such as percent of a watershed above 3,000 feet, tend to cluster at one or two values. For example, most coastal watersheds have zero percent of their area above 3,000 feet. Many of the characteristics, including the various monthly precipitation or temperature characteristics, are highly correlated with each other. Using combinations of these characteristics in a regression analysis does not add information and may lead to unstable and unreliable regression coefficients. Based on these considerations and some trial regressions using ordinary least-squares regression analysis, 15 characteristics were selected for the generalized least-squares regression analysis

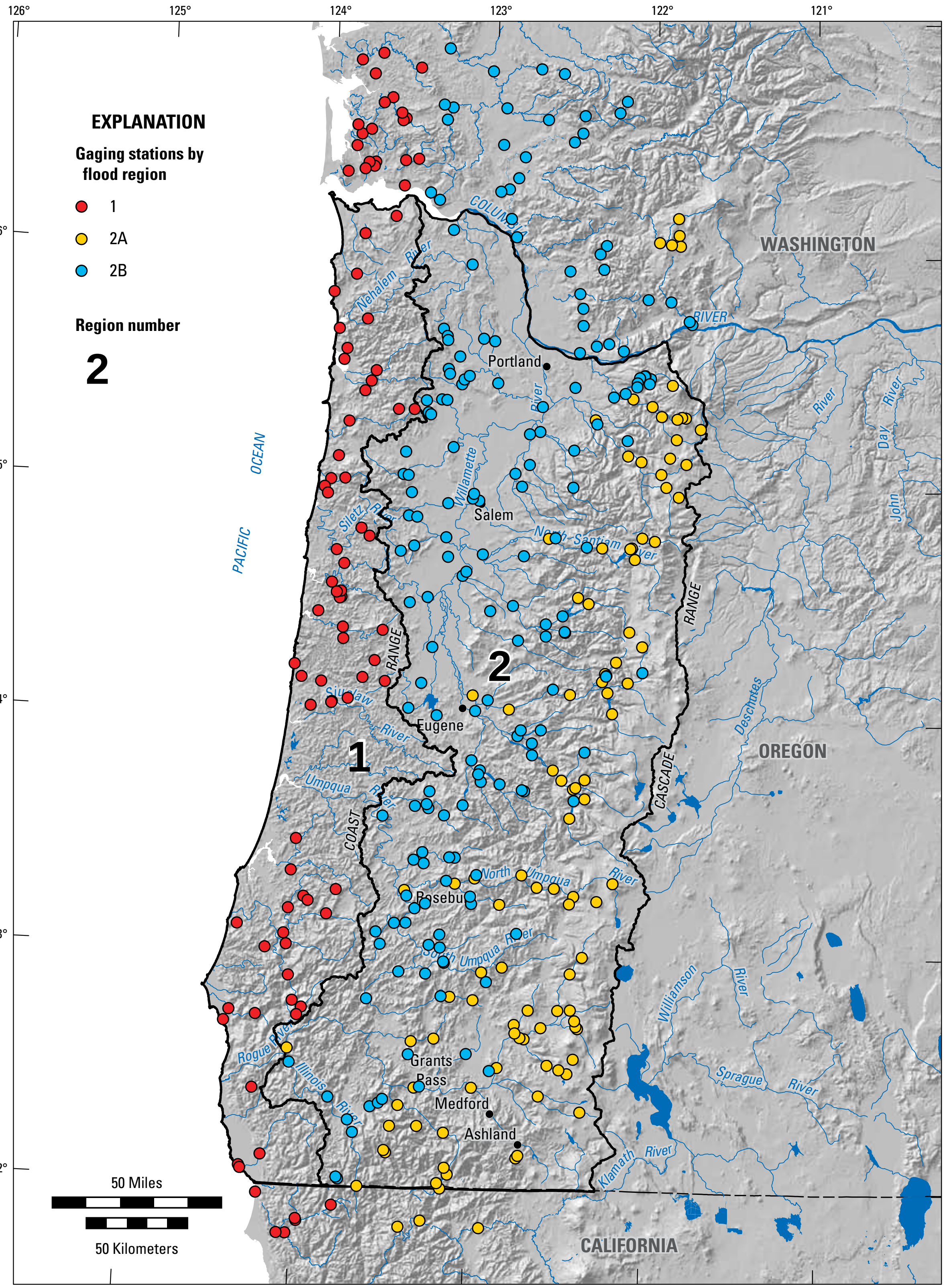

Figure 11. Flood regions of western Oregon. Regions 2A and 2B cannot be separated into discrete areas and are shown together as Region 2; however, the gaging stations associated with Regions 2A and 2B give a rough approximation of the areal extent of each region.

(table 2). These 15 characteristics for each of the 376 gaged watersheds used in the regional regression analysis are given in Appendix G.

The 15 selected characteristics were checked for collinearity. Matrices of the correlation coefficients for the characteristics of the watersheds for each of the three flood regions are shown in tables 3, 4, and 5. High correlation coefficients (absolute values greater than about 0.80) were detected. These pairs of characteristics were not allowed to appear together in a prediction equation.

The area determined for each gaged watershed from the spatial analysis was compared to its published value. Where significant differences occurred, the delineation of the watershed was checked. Errors in both the delineations and in the published areas were discovered in this way. The distribution of gaged watersheds by area and region is shown in table 6.

Description of the Watershed Characteristics

The computed characteristics represent the contributing watershed upstream of the gaging station, or other point of interest. The watershed is delineated based on topography as shown on U.S. Geological Survey 1:24,000-scale topographic maps.

Drainage area is the size of the watershed in square miles.

Maximum watershed relief is the maximum difference in elevation, in feet, between the lowest and the highest points in a watershed. The lowest point in the watershed is the outlet (or pour point) of the watershed. Relief is often highly correlated with area.

Mean watershed slope is calculated as the average of the slope of all the cells of the digital elevation model found within the watershed boundaries. Slope is given in degrees. For example, a 0 degree slope is horizontal, and a 90 degree slope is vertical.

Mean watershed elevation is calculated as the average of the elevations of all the cells of the digital elevation model found within the watershed boundaries. It is reported in feet.

Mean January precipitation, mean July precipitation, 24-hour 2-year precipitation intensity, and annual snowfall are calculated as the average of the values of all the cells of their respective grids found within the watershed boundary. All are reported in inches. Each of the grids represents averages for water years 1961 to 1990.

Mean minimum January temperature, mean minimum July temperature, mean maximum January temperature, and mean maximum July temperature are calculated as the average of the values of all the cells of their respective grids found within the watershed boundary. All are reported in degrees Fahrenheit. Each of the grids represents averages for water years 1961 to 1990.

Table 2. Watershed characteristics considered for the regression analysis.

[Units: mi², square miles; ft, feet; in, inches; in/hr, inches per hour; °, degrees; °F, degrees Fahrenheit]

Characteristic	Units	Data type	Scale or resolution	Source
Drainage area	mi^2	vector	1:24,000	Water Resources Department
Maximum watershed relief	ft	grid	30 m	U.S. Geological Survey
Mean watershed slope	°	grid	30 m	U.S. Geological Survey
Mean watershed elevation	ft	grid	30 m	U.S. Geological Survey
Mean January precipitation	in	grid	4,000 m	Oregon Climate Service
Mean July precipitation	in	grid	4,000 m	Oregon Climate Service
2-year 24-hour precipitation intensity	in	grid	3,000 m	Oregon Climate Service
Annual snowfall	in	grid	4,000 m	Oregon Climate Service
Mean minimum January temperature	°F	grid	4,000 m	Oregon Climate Service
Mean minimum July temperature	°F	grid	4,000 m	Oregon Climate Service
Mean maximum January temperature	°F	grid	4,000 m	Oregon Climate Service
Mean maximum July temperature	°F	grid	4,000 m	Oregon Climate Service
Soil storage capacity	in	vector	1:250,000	Natural Resources Conservation Service
Soil permeability	in/hr	vector	1:250,000	Natural Resources Conservation Service
Soil depth	in	vector	1:250,000	Natural Resources Conservation Service

Table 3. Correlation matrix of predictor variables for the 91 gaging stations of Region 1, coastal watersheds.

[Variables: All variables are log-transformed. Area, drainage area, in square miles; Relief, maximum difference in elevation, in feet; Slope, mean watershed slope, in degrees; Elev, mean watershed elevation, in feet; Jan P, mean January precipitation, in inches; Jul P, mean July precipitation, in inches; I24-2, 2-year 24-hour precipitation intensity, in inches; Snow, annual snowfall, in inches; Mn Jan T, mean minimum January temperature, in degrees Fahrenheit; Mn Jul T, mean minimum July temperature, in degrees Fahrenheit; Mx Jan T, mean maximum January temperature, in degrees Fahrenheit; Mx Jul T, mean maximum July temperature, in degrees Fahrenheit; Soil C, soil storage capacity, in inches; Soil P, soil permeability, in inches per hour; Soil D, soil depth, in inches. Correlations greater than 0.80 are in bold face.

	Area	Relief	Slope	Elev	Jan P	Jul P	I24-2	Snow	Mn Jan T	Mn Jul T	Mx Jan T	Mx Jul T	Soil C	Soil P	Soil D
Area	1.00														
Relief	**0.82**	1.00													
Slope	0.26	0.64	1.00												
Elev	0.28	0.64	0.77	1.00											
Jan P	0.11	0.29	0.26	0.27	1.00										
Jul P	-0.05	0.07	-0.02	0.13	0.49	1.00									
I24-2	0.18	0.40	0.42	0.42	**0.91**	0.38	1.00								
Snow	0.35	0.63	0.57	0.71	0.61	0.27	0.66	1.00							
Mn Jan T	0.01	-0.15	-0.14	-0.30	-0.48	-0.50	-0.40	-0.54	1.00						
Mn Jul T	-0.09	-0.20	-0.14	-0.16	-0.60	-0.47	-0.53	-0.38	0.59	1.00					
Mx Jan T	0.01	-0.15	-0.12	-0.26	-0.64	-0.62	-0.54	-0.62	**0.86**	0.48	1.00				
Mx Jul T	-0.11	-0.19	-0.07	-0.08	-0.66	-0.66	-0.51	-0.48	0.39	0.53	0.62	1.00			
Soil C	-0.14	-0.31	-0.41	-0.53	0.24	0.37	0.03	-0.20	-0.14	-0.37	-0.21	-0.49	1.00		
Soil P	0.30	0.41	0.35	0.38	0.28	0.44	0.34	0.39	-0.32	-0.15	-0.39	-0.38	-0.09	1.00	
Soil D	-0.09	-0.15	-0.30	-0.30	0.32	0.57	0.05	0.01	-0.21	-0.25	-0.35	-0.54	0.59	0.15	1.00

Table 4. Correlation matrix of predictor variables for the 107 gaging stations of Region 2A, western interior watersheds with mean elevations above 3,000 feet.

[Variables**:** All variables are log-transformed. Area, drainage area, in square miles; Relief, maximum difference in elevation, in feet; Slope, mean watershed slope, in degrees; Elev, mean watershed elevation, in feet; Jan P, mean January precipitation, in inches; Jul P, mean July precipitation, in inches; I24-2, 2-year 24-hour precipitation intensity, in inches; Snow, annual snowfall, in inches; Mn Jan T, mean minimum January temperature, in degrees Fahrenheit; Mn Jul T, mean minimum July temperature, in degrees Fahrenheit; Mx Jan T, mean maximum January temperature, in degrees Fahrenheit; Mx Jul T, mean maximum July temperature, in degrees Fahrenheit; Soil C**,** soil storage capacity, in inches; Soil P, soil permeability, in inches per hour; Soil D, soil depth, in inches. Correlations greater than 0.80 are in bold face]

	Area	Relief	Slope	Elev	Jan P	Jul P	I24-2	Snow	Mn Jan T	Mn Jul T	Mx Jan T	Mx Jul T	Soil C	Soil P	Soil D
Area	1.00														
Relief	**0.81**	1.00													
Slope	-0.13	-0.02	1.00												
Elev	-0.06	0.05	-0.52	1.00											
Jan P	-0.16	-0.07	0.18	-0.29	1.00										
Jul P	-0.13	-0.05	-0.31	0.15	0.70	1.00									
I24-2	-0.15	-0.07	0.20	-0.04	**0.85**	0.61	1.00								
Snow	0.08	0.13	-0.56	0.51	0.44	**0.80**	0.41	1.00							
Mn Jan T	-0.02	-0.08	0.71	-0.76	0.20	-0.32	0.18	-0.60	1.00						
Mn Jul T	-0.11	-0.13	0.59	-0.68	0.11	-0.36	0.10	-0.65	**0.88**	1.00					
Mx Jan T	0.21	0.09	0.47	-0.46	-0.31	-0.63	-0.22	-0.68	0.69	0.46	1.00				
Mx Jul T	0.18	0.04	0.31	-0.48	-0.39	-0.62	-0.32	-0.69	0.54	0.43	**0.88**	1.00			
Soil C	0.11	-0.01	-0.16	-0.18	0.39	0.53	0.20	0.41	-0.11	-0.31	-0.08	-0.04	1.00		
Soil P	0.10	0.13	-0.29	0.33	0.35	0.53	0.27	0.66	-0.48	-0.61	-0.43	-0.42	0.39	1.00	
Soil D	0.16	0.03	-0.52	0.04	0.33	0.54	0.11	0.56	-0.26	-0.29	-0.30	-0.20	0.69	0.44	1.00

Table 5. Correlation matrix of predictor variables for the 178 gaging stations of Region 2B, western interior watersheds with mean elevations less than 3,000 feet.

[Variables: All variables are log-transformed. Area, drainage area, in square miles; Relief, maximum difference in elevation, in feet; Slope, mean watershed slope, in degrees; Elev, mean watershed elevation, in feet; Jan P, mean January precipitation, in inches; Jul P, mean July precipitation, in inches; I24-2, 2-year 24-hour precipitation intensity, in inches; Snow, annual snowfall, in inches; Mn Jan T, mean minimum January temperature, in degrees Fahrenheit; Mn Jul T, mean minimum July temperature, in degrees Fahrenheit; Mx Jan T, mean maximum January temperature, in degrees Fahrenheit; Mx Jul T, mean maximum July temperature, in degrees Fahrenheit; Soil C, soil storage capacity, in inches; Soil P, soil permeability, in inches per hour; Soil D, soil depth, in inches. Correlations greater than 0.80 are in bold face]

	Area	Relief	Slope	Elev	Jan P	Jul P	I24-2	Snow	Mn Jan T	Mn Jul T	Mx Jan T	Mx Jul T	Soil C	Soil P	Soil D
Area	1.00														
Relief	**0.84**	1.00													
Slope	0.24	0.54	1.00												
Elev	0.28	0.64	0.65	1.00											
Jan P	0.11	0.29	0.19	0.28	1.00										
Jul P	-0.06	0.07	-0.08	0.14	0.50	1.00									
I24-2	0.18	0.39	0.37	0.43	**0.92**	0.39	1.00								
Snow	0.35	0.62	0.45	0.71	0.63	0.29	0.67	1.00							
Mn Jan T	0.00	-0.16	-0.03	-0.30	-0.48	-0.51	-0.41	-0.55	1.00						
Mn Jul T	-0.10	-0.21	-0.10	-0.18	-0.60	-0.48	-0.53	-0.40	0.60	1.00					
Mx Jan T	0.01	-0.15	0.00	-0.26	-0.65	-0.64	-0.55	-0.63	**0.87**	0.49	1.00				
Mx Jul T	-0.10	-0.18	-0.03	-0.09	-0.66	-0.67	-0.52	-0.50	0.40	0.53	0.64	1.00			
Soil C	-0.13	-0.30	-0.41	-0.53	0.23	0.37	0.02	-0.19	-0.12	-0.35	-0.21	-0.49	1.00		
Soil P	0.31	0.42	0.34	0.36	0.30	0.46	0.36	0.39	-0.33	-0.17	-0.39	-0.41	-0.08	1.00	
Soil D	-0.09	-0.14	-0.33	-0.30	0.31	0.56	0.05	0.01	-0.21	-0.24	-0.35	-0.53	0.60	0.17	1.00

Table 6. Numbers of gaging stations and their average record length by area and region.

Area, in square miles	Region 1		Region 2A		Region 2B	
	Number of gages	Average record length, in years	Number of gages	Average record length, in years	Number of gages	Average record length, in years
<1	13	15. 8	1	14.0	12	19.9
1–3	16	18.1	7	15.0	23	18.4
3–10	14	16.9	10	21.9	22	18.0
10–30	14	17.0	15	23.1	21	23.4
30–100	15	25.9	24	27.7	48	28.0
100–300	14	39.1	25	42.8	30	40.6
300–1,000	5	46.8	17	37.9	15	34.9
1,000–3,000	0	N/A	7	36.1	4	35.3
>3,000	0	N/A	1	16.0	3	67.3
Total	91	23.5	107	31.2	178	28.0

Soil capacity is the maximum volume of water the soil is expected to hold. It is calculated as the area-weighted average of the soil capacity for all the soils found within the watershed boundary. Soil capacity for a given soil is its porosity times its depth. Soil capacity is reported in inches.

Soil permeability is the rate at which water is expected to infiltrate the soil. It is calculated as the area-weighted average of the infiltration rate for all the soils found within the watershed boundary. It is reported in inches per hour.

Soil depth is the depth of soil to bedrock averaged over the watershed. It is reported in inches.

Selection of Gaging Stations

Within the study area and adjacent parts of Washington and California there are between 450 and 500 gaging stations where peak discharges have been systematically recorded. Of these, 399 stations had more than 10 years of record and were in rural watersheds unaffected by significant diversion, regulation or urbanization. Twenty-four of these stations were eliminated for a variety of reasons.

The locations of four gaging stations could not be determined: Darlingtonia Creek at Darlingtonia, California (11530950), Lookout Creek tributary no. 3 near Blue River, Oregon (14161200), South Fork Weiss Creek near Waldport, Oregon (14306850), and Buck Creek tributary near Scottsburg, Oregon (14323020). Published information about the station location (latitude and longitude, physical description, public land survey, and drainage area) could not be reconciled with any actual watershed on 1:24,000 scale topographic maps.

Peak discharges at four gaging stations are located at the outlets of large natural lakes: Tenmile Creek near Lakeside, Oregon (14323200), Eel Creek at Lakeside, Oregon (14323300), Waldo Lake outlet near Oakridge, Oregon (14147000), and McKenzie River at outlet of Clear Lake, Oregon (14158500). The lakes all occupy more than 5 percent of the drainage area above their respective stations. Peak discharges are presumed to be significantly attenuated.

Peak discharges at eight gaging station poorly fitted the log-Pearson Type III distribution: Beaver Creek near Klamath River, California (11517800), Soap Creek tributary near Fort Jones, California (11518610), Middle Fork Willamette River at Jasper, Oregon (14152000), Grant Creek near Falls City, Oregon (14190350), Collawash River tributary near Breitenbush Hot Springs, Oregon (14208200), Kink Creek near Government Camp, Oregon (14209100), South Fork Deer Creek near Dixonville, Oregon (14312170), and Star Gulch near Ruch, Oregon (14362250). For each station, the upper end of the distribution is poorly defined, and peak discharges estimated from it are uncertain. As an example, the fit for gaging station Collawash River tributary near Breitenbush Hot Springs, Oregon (14208200) is shown on figure 12.

Thielsen Creek near Diamond Lake, Oregon (14312700) is underlain by young, highly porous volcanic rock. The watershed boundary is uncertain and significant stream losses occur.

In several cases, gaging stations occur near each other on the same stream reach. In six of these cases, one or other of each pair was eliminated: Middle Santiam River near Upper Soda, Oregon (14185700), Wiley Creek at Foster, Oregon (14187100), North Umpqua River below Steamboat Creek near Glide, Oregon (14316800), Rogue River below Prospect, Oregon (14330000), Applegate River near Ruch, Oregon (14363000), and Illinois River at Kerby, Oregon (14377000). For each pair, estimated peak discharges at the upstream station are greater than at the downstream station. The apparent decrease in discharge occurs not because of stream losses, but because of uncertainty in estimating the peak discharges. For each pair, only the station considered the most reliable was retained. The stations were judged on their length and quality of record and their fit to the probability distribution.

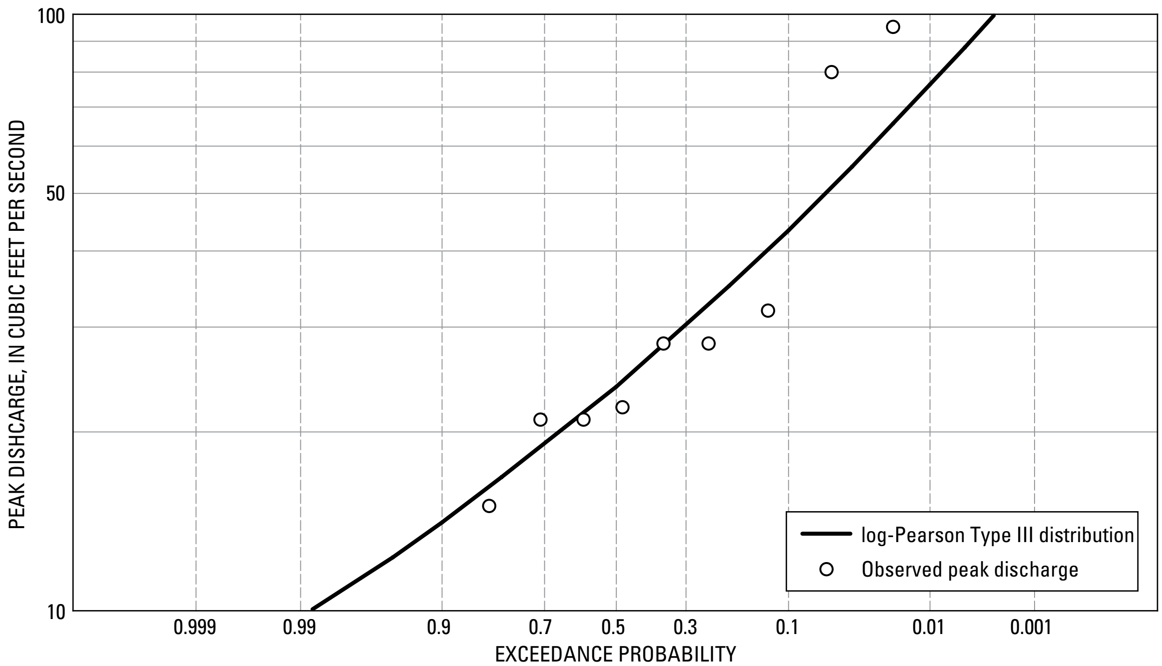

Figure 12. The fitted log-Pearson Type III distribution for gaging station Collawash River Tributary near Breitenbush Hot Springs, Oregon (14208200). The upper end of the distribution is poorly defined, and peak discharges estimated from it are uncertain. This station was one of eight eliminated because of its poor fit to the probability distribution.

The Regression Analysis

A regional regression analysis is based on the assumption that streamflow is related to various physical and climatological characteristics. For example, streamflow increases with watershed size, other factors, such as precipitation, being equal. A 100-square-mile watershed produces more runoff than a 25-square-mile watershed.

As an example, the relationship between 100-year peak discharges and watershed area for Region 1 is shown in figure 13. The line shown on the plot minimizes the sum of the squared vertical differences between the line and the points. The line “models” the relationship between peak discharge and watershed area. It can be used to predict the peak discharge for a watershed in the same region given its area. The variation about the line is due, in part, to other watershed characteristics not included in the model.

Similar relationships exist between peak discharge and other watershed characteristics (table 2), each characteristic accounting for part of the variability in streamflow. These relationships can be quantified in a mathematical form. For this analysis, a linear relationship is assumed between streamflow and watershed characteristics. The linear mathematical model takes the form

$$y = b_0 + b_1x_1 + b_2x_2 + b_mx_m \quad (4)$$

where y represents streamflow and $x_1, x_2, \ldots, x_m$ represent the *m* watershed characteristics. The regression coefficients, $b_1, b_2, \ldots, b_m$ define the relationship among variables and are determined from the data. The data consist of *n* observations of y and x_m, from which *n* equations of the type of Equation 4 can be written. The regression coefficients are determined by minimizing the sum of the squared differences between the actual values of y and the values of y estimated by the *n* equations. The equations resulting from this minimization are called the normal equations.

While regression analysis assumes a linear relationship between the response and predictor variables, the true relationship for peak discharges is nonlinear. A log-transformation of peak discharges and watershed characteristics allows the nonlinear relationship to be modeled by a linear relationship (Riggs, 1968; 1973).

The nonlinear model of the relationship between streamflow and watershed characteristics looks like this:

$$y = 10^{b_0} x_1^{b_1} x_2^{b_2} \cdots x_m^{b_m} \quad (5)$$

A logarithmic transformation of Equation 5 yields the linear relationship

$$\log_{10}(y) = b_0 + b_1\log_{10}(x_1) + b_2\log_{10}(x_2) + \cdots + b_m\log_{10}(x_m) \quad (6)$$

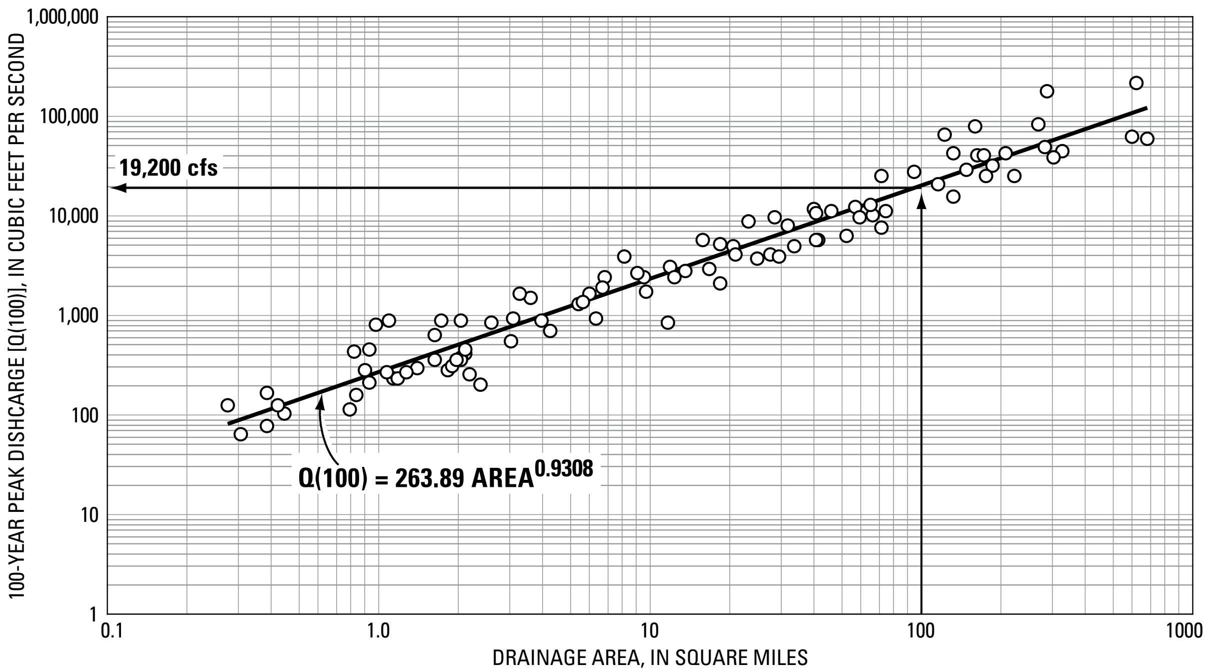

Figure 13. A simple regional regression model. 100-year peak discharges are plotted against watershed area for Region 1, coastal watersheds. The line (i.e., the model) through the data points was fitted by ordinary least squares regression analysis and is represented mathematically by the equation shown on the graph. Based on this model, a watershed of 100 square miles has a 100-year peak discharge of about 19,200 cfs (cubic feet per second).

Previous studies in Oregon have used ordinary least-squares regression to derive the prediction equations. Ordinary least-squares regression assumes that peak discharge records are equally reliable, i.e., of the same length and variance, and that concurrent flows at any pair of stations are independent. These conditions are seldom met in practice.

Tasker and Stedinger (1989) proposed an operational generalized least-squares model for deriving prediction equations for streamflow characteristics such as peak discharge. This model accounts for the unequal lengths and variances of streamflow records and cross-correlation between series of streamflow characteristics. Tasker and others (1986) showed that generalized least squares, compared to ordinary least squares, provides (1) estimates of regression parameters with smaller mean square errors, (2) relatively unbiased estimates of the variance of the regression parameters, and (3) a more accurate estimate of the model error. The prediction equations in this study were derived using generalized least-squares regression.

Defining the Prediction Equations

Only some of the 15 watershed characteristics are correlated with peak discharge. Since only correlated watershed characteristics can explain the observed variability in peak discharges, there is no benefit to including all characteristics in a prediction equation. The goal, then, is to find the prediction equation that explains as much of the observed variability in peak discharges as possible with the fewest number of watershed characteristics.

With 15 watershed characteristics, the number of possible prediction equations is $2^{15} - 1$ or 32,767. Rather than test all possible prediction equations, a backward-step analysis may be used to determine the best prediction equation. In a backward step analysis, a regression is done using all candidate watershed characteristics. The characteristic that has the least significant coefficient is eliminated and the regression is run again. This process is repeated until only one characteristic remains.

Each regression is associated with a set of watershed characteristics and their respective coefficients, and each set of characteristics and coefficients represents a candidate prediction equation. The best prediction equation generally is considered the combination of watershed characteristics that gives the smallest model error while its regression coefficients are all significantly different from zero. The significance of the regression coefficients is determined by a statistical test (Student's t-test was used).

The null hypothesis, H_o, is that the coefficient in question is equal to zero. The statistical test determines the probability, P, that the coefficient is *not* different from zero. H_o is rejected, and the coefficient retained, for small values of P. In this analysis, H_o is rejected for P less than 0.05.

The computer program used to do the generalized least squares regressions (GLSNET, version 2.5), limits the number of predictor variables to 9, so the set of 15 watershed characteristics had to be reduced to 9 or fewer for each region. First, highly correlated pairs of watershed characteristics ($r \geq 0.8$) were identified for each region. A regression was done for each characteristic from each pair. Only the characteristic with the most significant regression coefficient was retained. Second, regressions were done using ordinary least squares analysis to determine the characteristics most likely to be significantly correlated to peak discharge from among the remaining characteristics.

When the set of nine or fewer characteristics was determined for each region, a backward step analysis was done using the 100-year peak discharges. The results of the backward-step analyses for Regions 1, 2A, and 2B are shown in tables 7, 8, and 9, respectively.

The set of characteristics determined for the 100-year peak discharges was used for all frequencies. If a backward step analysis is done independently at each frequency, the resulting prediction equations may incorporate different predictor variables. While this may lead to the smallest model errors for each equation, it may lead to undesirable results overall. Specifically, flood magnitude may not vary smoothly with frequency—a plot of magnitude versus frequency likely will show discontinuities. It is even possible that the magnitude of a high frequency event will exceed the magnitude of a low frequency event. For example, the 10-year event could be larger than the 25-year event.

The final prediction equations are shown by region in tables 10, 11, and 12. Maps of all of the characteristics used in the prediction equations are shown in figures 14, 15, 16, 17, 18, and 19. These maps are for illustration only. It is strongly recommended that estimates of watershed characteristics be made from the digital grids and coverages described in table 2 using GIS techniques.

Accuracy of the Prediction Equations

Measures of the accuracy of the prediction equations are average prediction error (Wiley and others, 2000) and equivalent years of record (Hardison, 1971). These measures are reported in tables 10, 11, or 12 for all prediction equations developed in this analysis. The average prediction error ranged from 25.3 to 39.1 percent over the three flood regions. Equivalent years of record varied from 2.0 to 13.6 years. Flood Regions 2A and 2B had the highest average prediction errors, and Region 1, the lowest.

The average prediction error is the square root of the sum of the squared standard error of the model and the average squared standard error of sampling, in log units. Model error is the uncertainty due to a model that does not account for all the variability in peak discharges. Sampling error is the uncertainty due to estimating model parameters from a sample, i.e., not from the whole population (Tasker and Stedinger, 1989). For the prediction equations, the average error of prediction is within 3.5 percentage points of the model error in all cases. Sampling error is a small part of the total error.

In practical terms, the small sampling error compared to the large model error means increasing the length of record available for estimating the peak discharges at gaged watersheds will not significantly decrease the average error of prediction. More benefit would result from improving the models by increasing the accuracy with which current watershed characteristics are estimated or by adding new characteristics to account for previously unaccounted for variability. The preceding comment does not mean that estimates of peak discharge at individual gaging stations could not be improved by additional years of record. Estimates at short record stations likely would be improved by additional record.

An equivalent number of years of record is the number of years of actual record required to give the same average prediction error as the regression. It is also used as a weighting factor in estimating peak discharges at gaging stations (Equation 9—discussed later). Hardison (1971) describes the calculation for estimating an equivalent number of years of record.

Transition Zone between Regions 2A and 2B

Although watersheds with mean watershed elevations above and below 3,000 feet are assigned to different flood regions (2A and 2B), the effect of elevation on peak discharge should change smoothly as elevation increases through 3,000 feet. Ideally, then, there should be a smooth transition of peak discharge estimates from one flood region into the other. In fact, there is often a discontinuity. For a watershed with a mean elevation near 3,000 feet, calculation of peak discharges by prediction equations for both Regions 2A and 2B generally do not yield the same result.

To ensure a smooth transition between Flood Regions 2A and 2B, peak discharges for watersheds with mean elevations near 3,000 feet are estimated by a weighted average of peak discharges estimated by prediction equations for both regions. For watersheds with mean elevations within a given transition zone, the following equation assumes that there is a linear change in peak discharges from one region into the other.

$$Q_T = Q_{2b}\left(\frac{3,000 + W/2 - E}{W}\right) + Q_{2a}\left(\frac{E - 3,000 + W/2}{W}\right) \qquad (7)$$

where

Q_T = the weighted discharge of the watershed in the transition zone,
Q_{2a} = the discharge estimated by the prediction equation for Region 2A,
Q_{2b} = the discharge estimated by the prediction equation for Region 2B,
W = the width of the transition zone in feet of elevation, and
E = the mean elevation of the watershed.