Magnitude and Frequency of Floods in Rural Basins of South Carolina, North Carolina, and Georgia

In cooperation with the South Carolina, Georgia, and North Carolina Departments of Transportation and the North Carolina Floodplain Mapping Program (Division of Emergency Management)

Why estimate the magnitude and frequency of floods?

Cost-Effective Design of Roads and Bridges

USGS
science for a changing world

Design of dams, levees, culverts and other flood-control structures

Flood-plain management and risk assessment

Log-Pearson Type III Distribution

U.S. Water Advisory Committee on Water Data (1982) recommended the log-Pearson Type III Distribution for flood frequency analyses.

The log-Pearson Type III distribution is calculated using the general equation:

$$\text{Log } Q_T = \log X_{mean} + KS$$

where

Q_T is the flood discharge for the specified return interval T;

$\log X_{mean}$ is the mean of the log x discharge values;

K is a frequency factor; and

S is the standard deviation of the log x values.

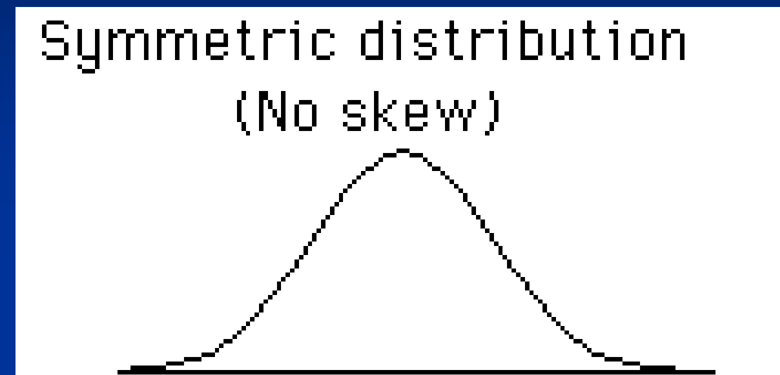
What about this K value?

$$\text{Log } Q_T = \log X_{mean} + KS$$

K, the frequency factor, is a function of skew and return interval.

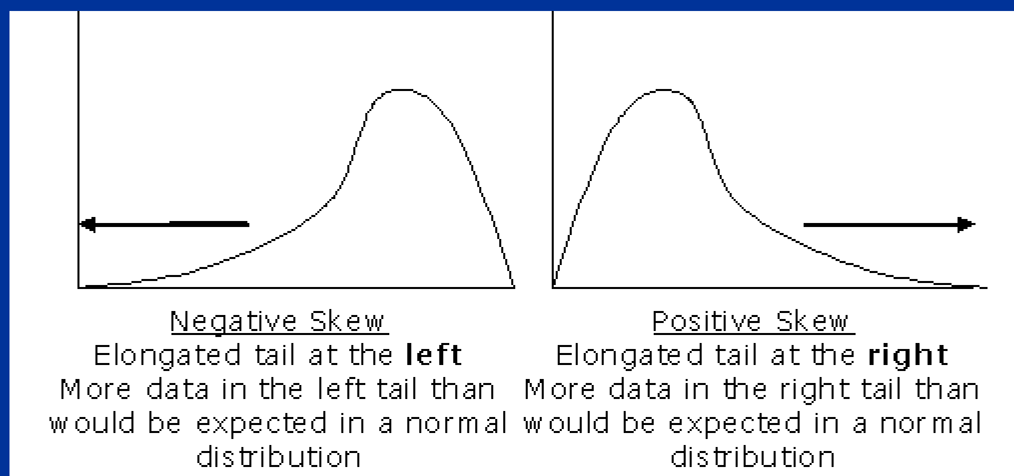
Let's talk about skew

Mean identical or close to median



Skew describes the symmetry (or asymmetry) in a data sample

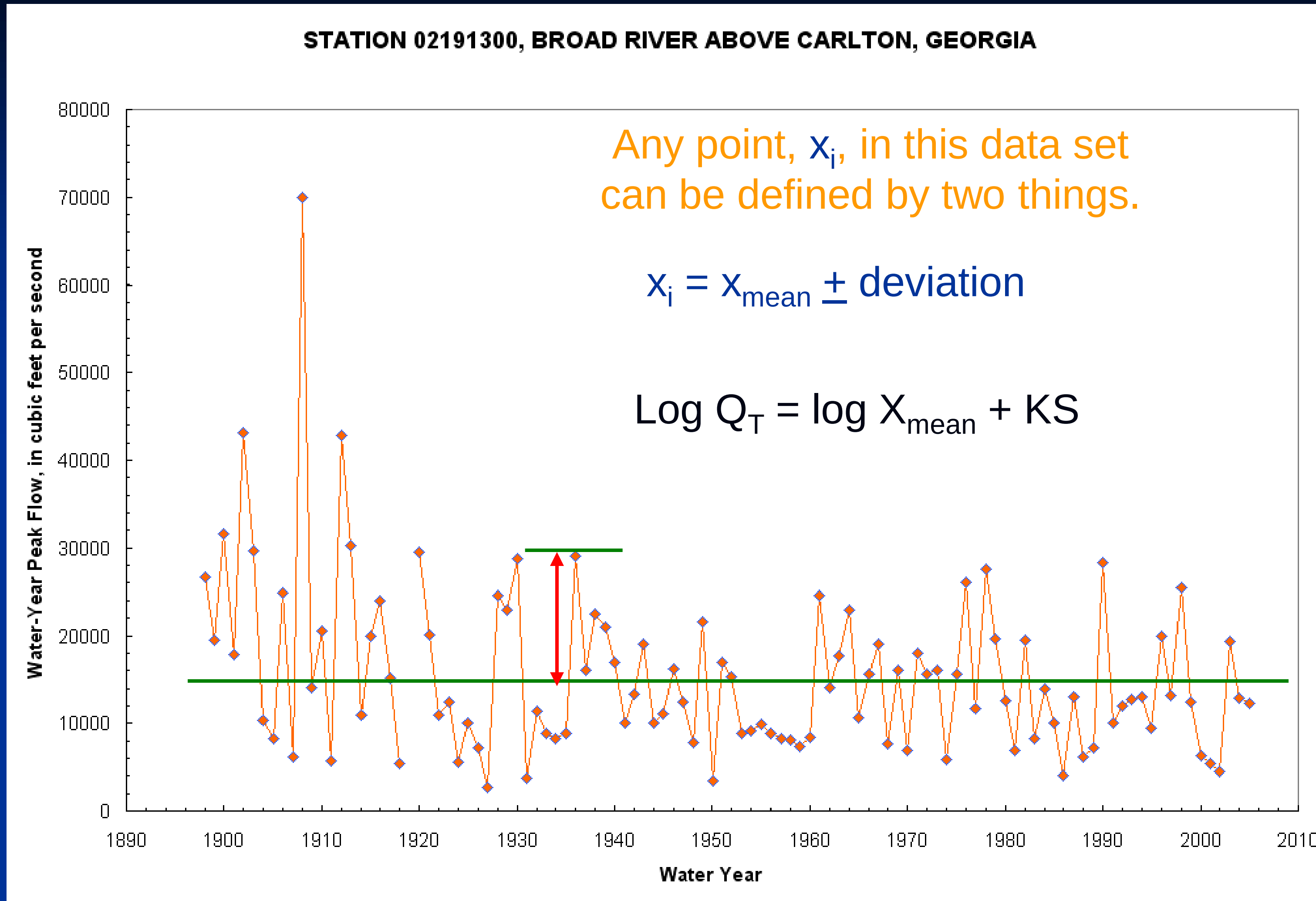
Mean less than median



Mean greater than median

Negative skew

Positive skew



So, peak-flow values possess two important properties: (1) the tendency to deviate from the mean; and (2) the frequency of occurrence. The K value is a function of skew and recurrence interval and acts as an adjustment to the standard deviation based on recurrence interval.

Why is a Regional Skew Desirable?

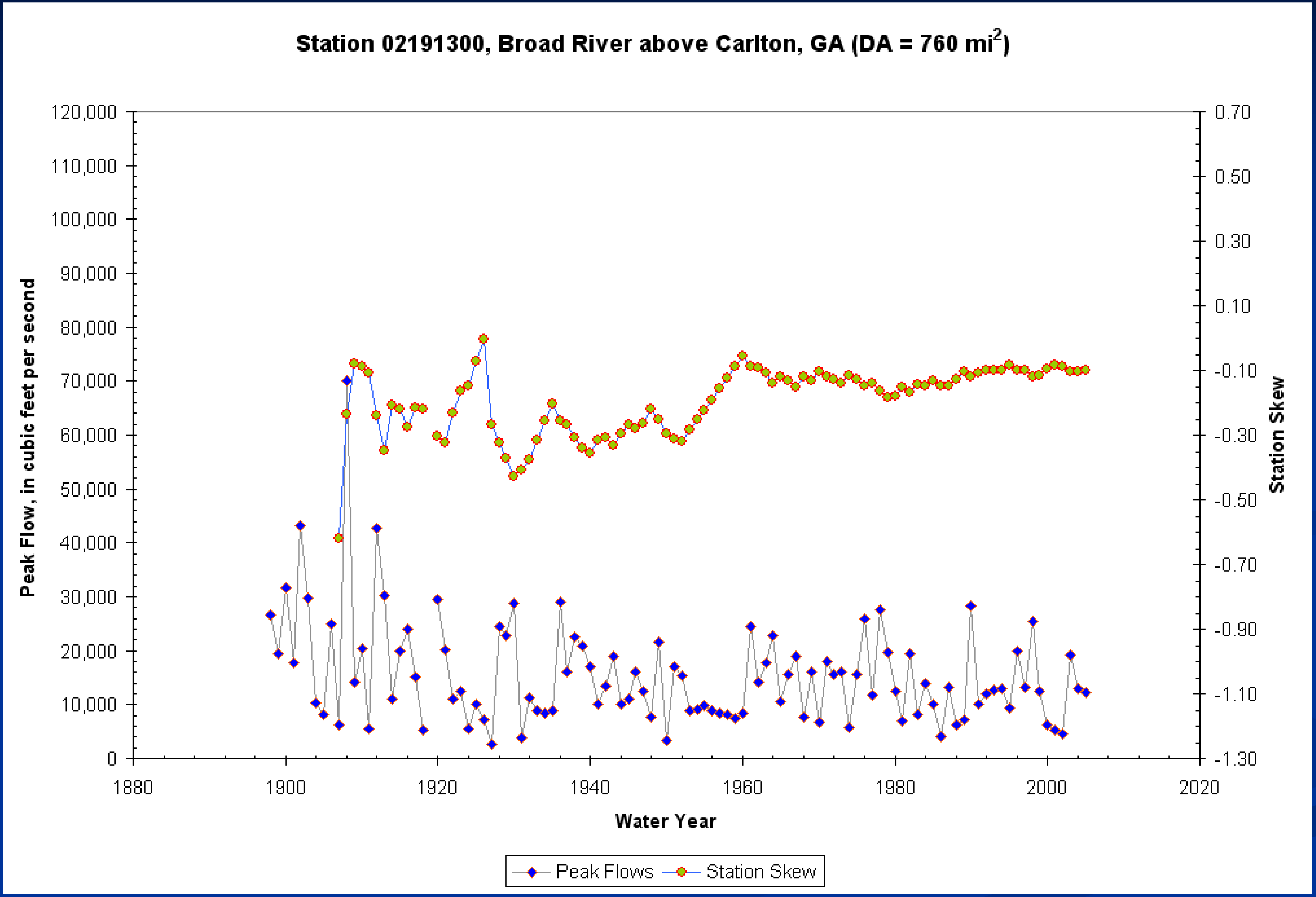
$$\text{Log } Q_T = \log X_{mean} + KS$$

Remember K is a function of the skew and the recurrence interval.

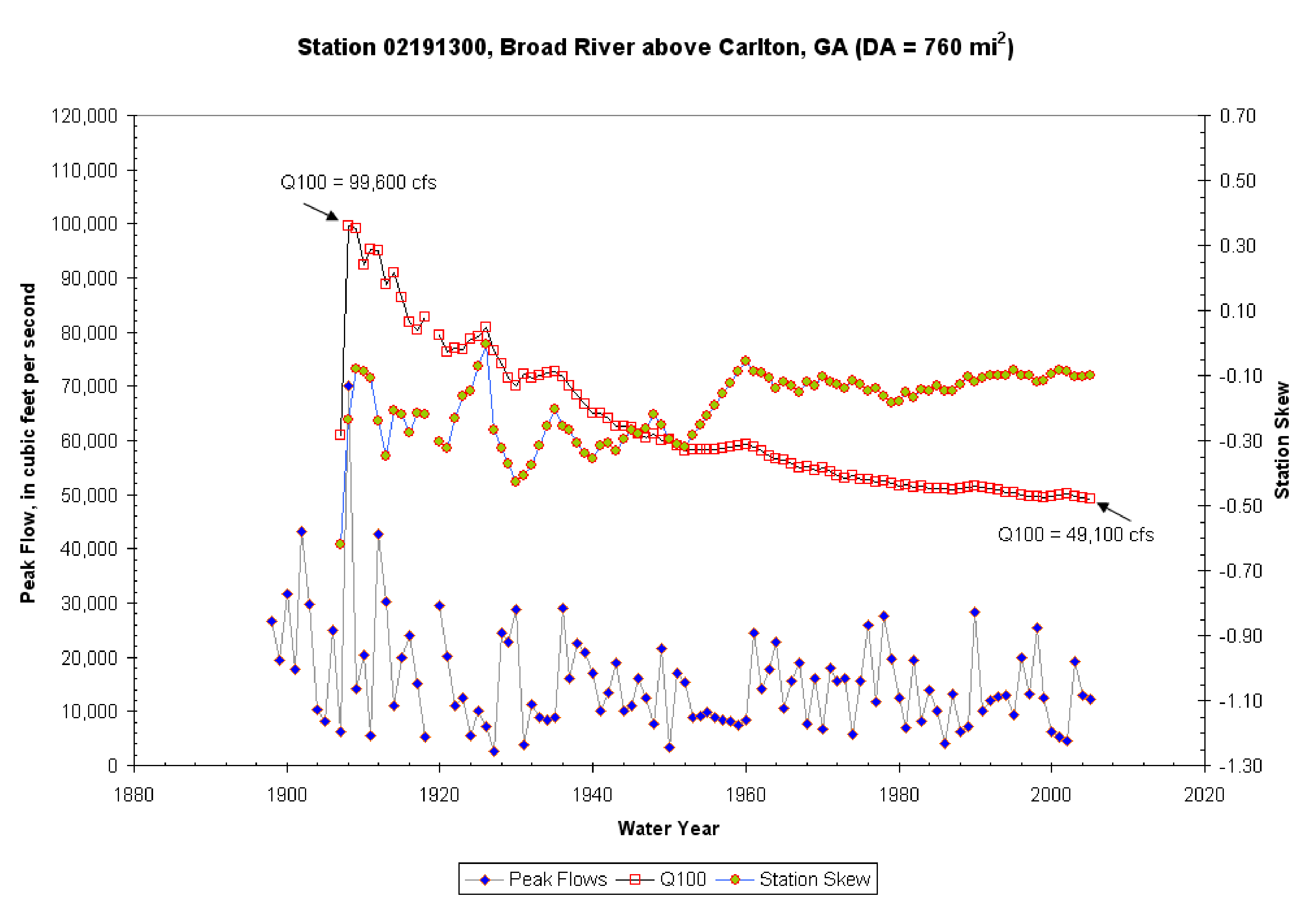
The skew coefficient for a given station is sensitive to extreme events making it difficult to obtain an accurate skew estimate from small samples.

The accuracy of the estimated skew can be improved by weighting the station skew with a generalized skew estimated by pooling information from numerous stations.

Skew Sensitivity



Skew and Q100 Sensitivity



USGS
science for a changing world

Updated regional skew

A. Gruber and J. R. Stedinger

Cornell University
School of Civil and Environmental Engineering

- Regional skew was updated based on analyses of 342 sites (after screenings) across South Carolina, Georgia, North Carolina, and surrounding states

- Applied Bayesian GLS statistical methods in efforts to develop the "best" model of regional skew based on explanatory variables (25 basin characteristics)

- When completed, determined that a constant skew was applicable to entire 3-state study area

Regional skew = -0.0186, MSE = 0.0831, equivalent record length 69 years (compared to B17 skew map's equivalent record length of 17 years (MSE = 0.3025))

Stations Included In The Regional Regression

64 stations in South Carolina

303 stations in North Carolina

310 stations in Georgia

20 stations in Alabama

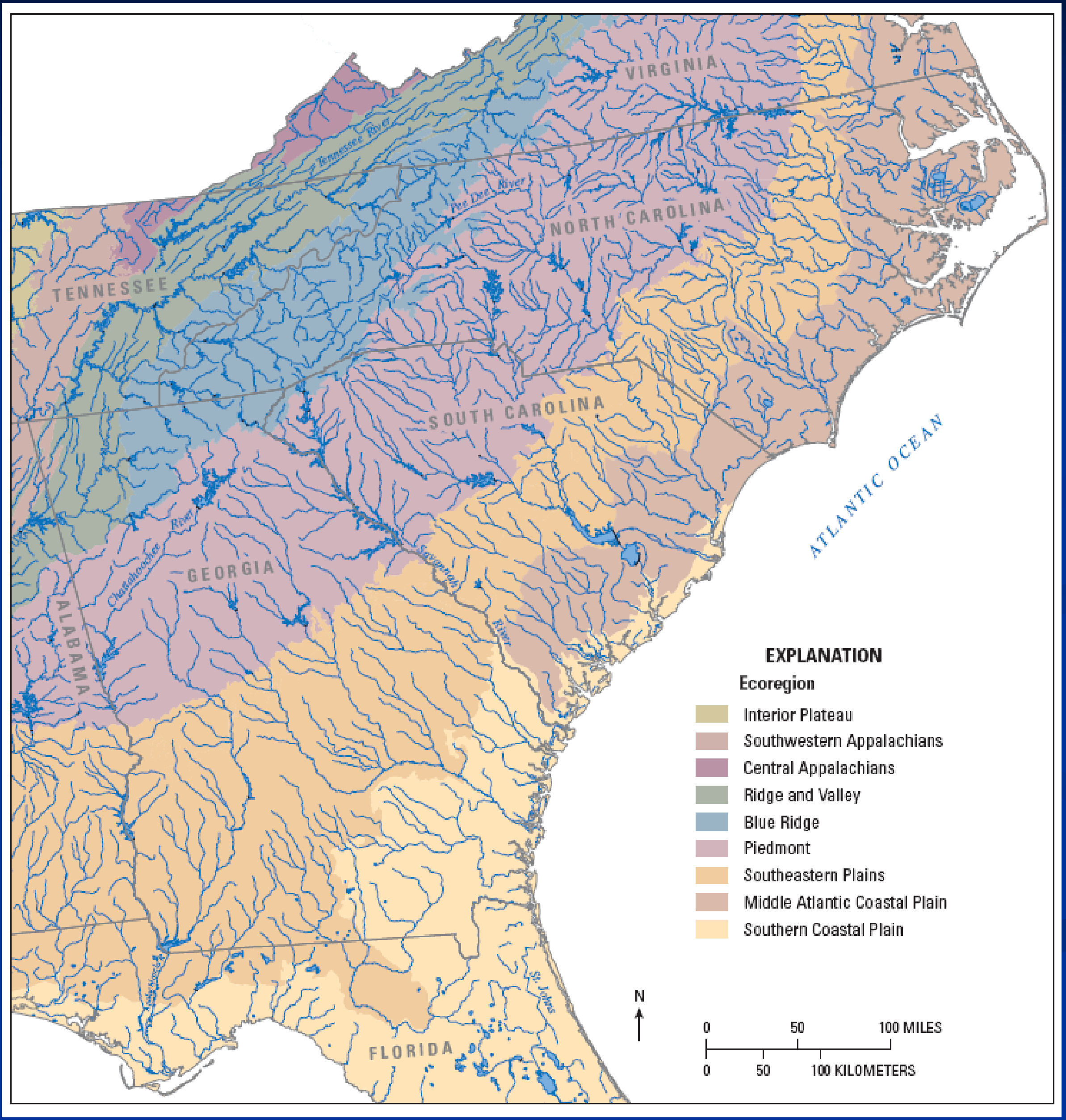
23 stations in Florida

40 stations in Tennessee

68 stations in Virginia

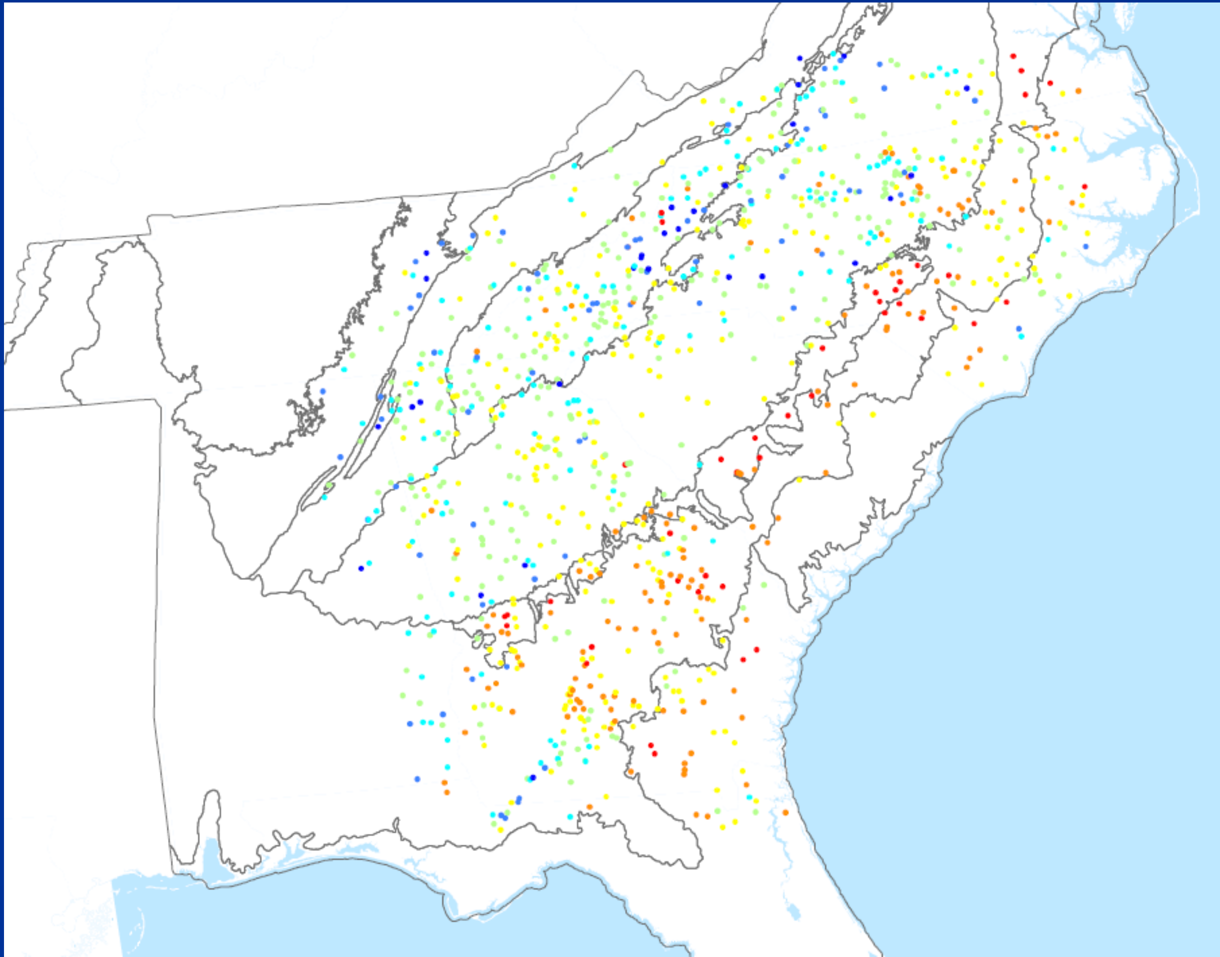
Total of 828 sites

EPA Level III Ecoregions

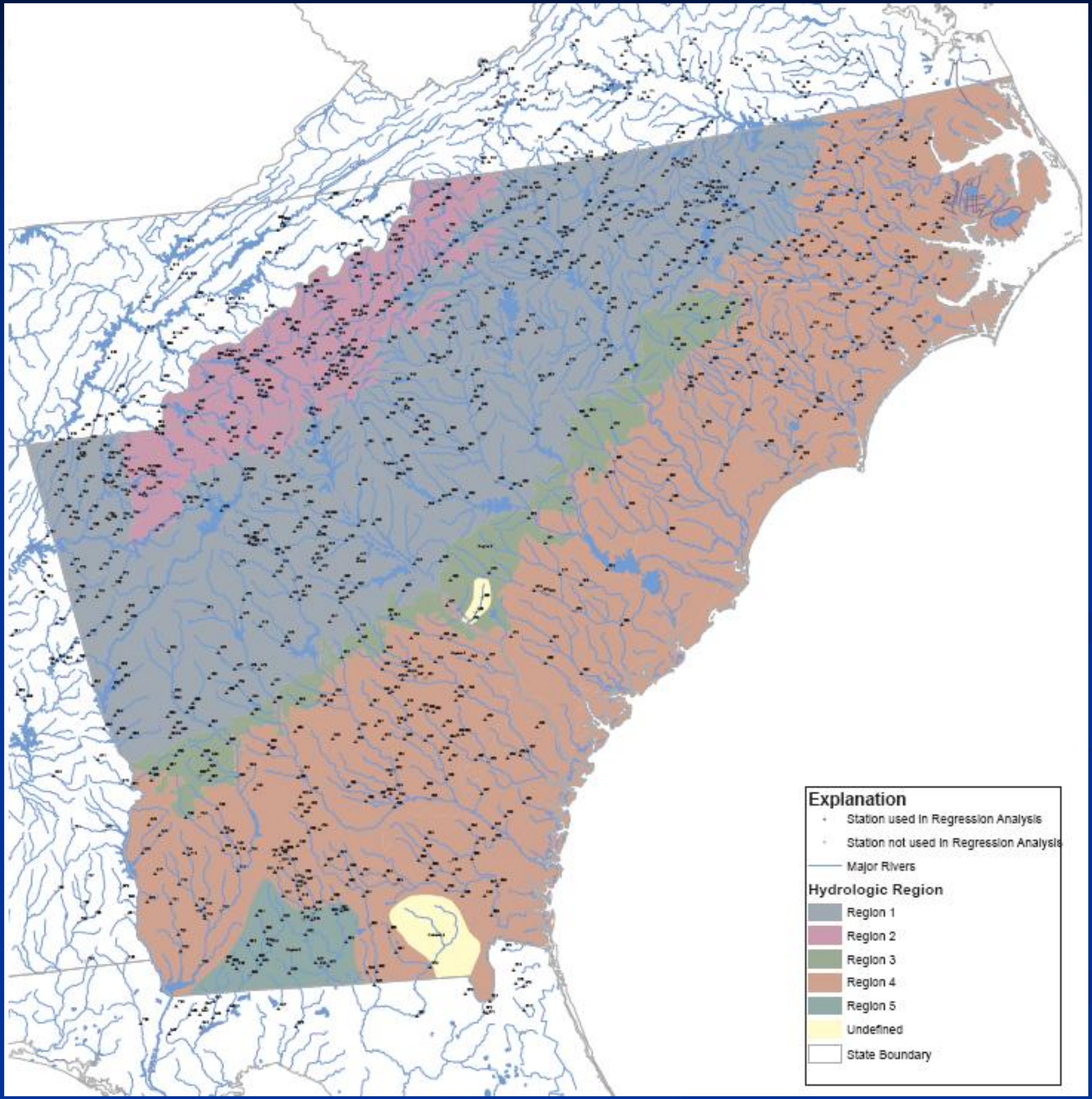


Hydrologic Regions

Based on initial regressions and assessing residuals, several regions were found to react similarly with respect to floods and therefore, were grouped together.



Hydrologic Regions



Region 1: Ridge/Valley and Piedmont

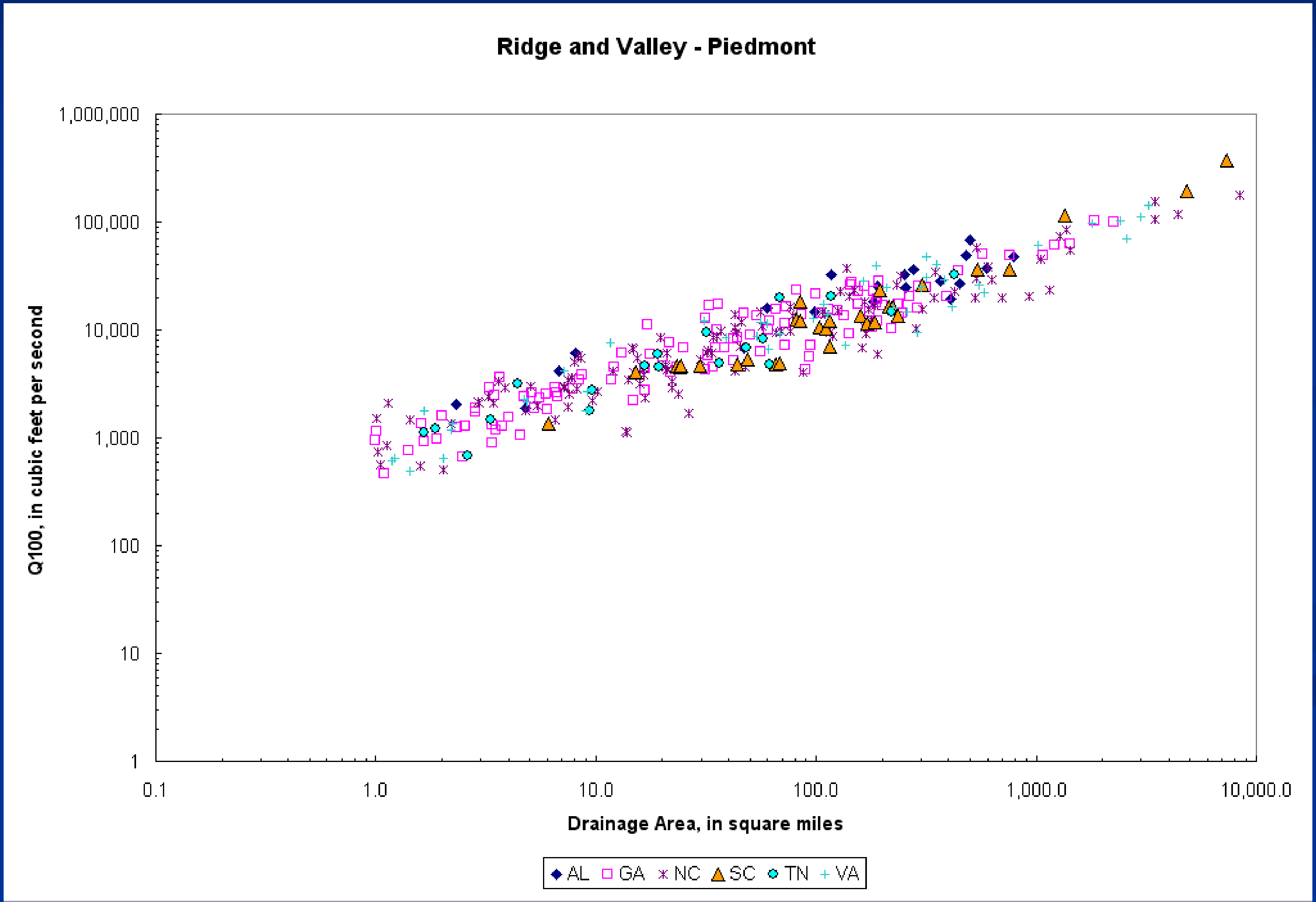
Region 2: Blue Ridge

Region 3: Sandhills

Region 4: Coastal

Region 5: Southwest Georgia

Let's take a look at the Q100 data by Hydrologic Regions (stations draining at least 75% from one region).



USGS
science for a changing world

Journal of Hydrology (2007) 344, 82–95

available at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/jhydrol

The use of GLS regression in regional hydrologic analyses

V.W. Griffis [a,*], J.R. Stedinger [b,1]

[a] Department of Civil and Environmental Engineering, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931-1295, USA
[b] School of Civil and Environmental Engineering, Cornell University, Hollister Hall, Ithaca, NY 14853-3501, USA

Received 27 October 2006; received in revised form 2 May 2007; accepted 30 June 2007

KEYWORDS
Generalized least squares regression;
Flood frequency analysis;
Regional skew;
Log-Pearson type 3 distribution

Summary To estimate flood quantiles and other statistics at ungauged sites, many organizations employ an iterative generalized least squares (GLS) regression procedure to estimate the parameters of a model of the statistic of interest as a function of basin characteristics. The GLS regression procedure accounts for differences in available record lengths and spatial correlation in concurrent events by using an estimator of the sampling covariance matrix of available flood quantiles. Previous studies by the US Geological Survey using the LP3 distribution have neglected the impact of uncertainty in the weighted skew on quantile precision. The needed relationship is developed here and its use is illustrated in a regional flood study with 162 sites from South Carolina. The performance of a pooled regression model is compared to separate models for each hydrologic region: statistical tests recommend an interesting hybrid of the two which is both surprising and hydrologically reasonable. The statistical analysis is augmented with new diagnostic metrics including a condition number to check for multicollinearity, a new pseudo-$\bar{R}^2$ appropriate for use with GLS regression, and two error variance ratios. GLS regression for the standard deviation demonstrates that again a hybrid model is attractive, and that GLS rather than an OLS or WLS analysis is appropriate for the development of regional standard deviation models.

Introduction

An important problem in hydrology is estimation of flood quantiles for ungauged locations, or sites with very short records. Regional generalized least squares (GLS) analyses are commonly used to estimate such statistics using physiographic characteristics of a catchment such as drainage

* Corresponding author. Tel.: +1 906 487 1079; fax: +1 906 487 2943.
E-mail addresses: vgriffis@mtu.edu (V.W. Griffis), jrs5@cornell.edu (J.R. Stedinger).
1 Tel.: +1 607 255 2351; fax: +1 607 255 9004.

0022-1694/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2007.06.023

In the regression analysis from the previous South Carolina flood-frequency investigations, we have only included stations draining at least 75% from one region (physiographic province), which is standard practice.

In 2007, Stedinger and Griffis published a paper using the peak-flow data base from the previous South Carolina flood-frequency investigation. In that paper, they used a pooled regression analysis in which a qualitative variable was included for physiographic region. This is similar to what Feaster and Tasker did for the Piedmont and upper Coastal Plain.

USGS
science for a changing world

Stedinger noted that for regions with relatively few sites, pooling the data allows for development of a more accurate model.

The pooled approach allows for a common slope with different intercepts, which makes sense if one believes the basin time of concentration should scale with area to a power.

The different intercepts allow for differences in runoff volume due to differences in soil characteristics, land cover, storage area, slope, etc.

Journal of Hydrology (2007) 344, 82–95

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

ELSEVIER

JOURNAL OF HYDROLOGY

The use of GLS regression in regional hydrologic analyses

V.W. Griffis [a,*], J.R. Stedinger [b,1]

[a] Department of Civil and Environmental Engineering, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931-1295, USA
[b] School of Civil and Environmental Engineering, Cornell University, Hollister Hall, Ithaca, NY 14853-3501, USA

Received 27 October 2006; received in revised form 2 May 2007; accepted 30 June 2007

KEYWORDS
Generalized least squares regression;
Flood frequency analysis;
Regional skew;
Log-Pearson type 3 distribution

Summary To estimate flood quantiles and other statistics at ungauged sites, many organizations employ an iterative generalized least squares (GLS) regression procedure to estimate the parameters of a model of the statistic of interest as a function of basin characteristics. The GLS regression procedure accounts for differences in available record lengths and spatial correlation in concurrent events by using an estimator of the sampling covariance matrix of available flood quantiles. Previous studies by the US Geological Survey using the LP3 distribution have neglected the impact of uncertainty in the weighted skew on quantile precision. The needed relationship is developed here and its use is illustrated in a regional flood study with 162 sites from South Carolina. The performance of a pooled regression model is compared to separate models for each hydrologic region: statistical tests recommend an interesting hybrid of the two which is both surprising and hydrologically reasonable. The statistical analysis is augmented with new diagnostic metrics including a condition number to check for multicollinearity, a new pseudo-$\overline{R}^2$ appropriate for use with GLS regression, and two error variance ratios. GLS regression for the standard deviation demonstrates that again a hybrid model is attractive, and that GLS rather than an OLS or WLS analysis is appropriate for the development of regional standard deviation models.

Introduction

An important problem in hydrology is estimation of flood quantiles for ungauged locations, or sites with very short records. Regional generalized least squares (GLS) analyses are commonly used to estimate such statistics using physiographic characteristics of a catchment such as drainage

* Corresponding author. Tel.: +1 906 487 1079; fax: +1 906 487 2943.
E-mail addresses: vgriffis@mtu.edu (V.W. Griffis), jrs5@cornell.edu (J.R. Stedinger).
1 Tel.: +1 607 255 2351; fax: +1 607 255 9004.

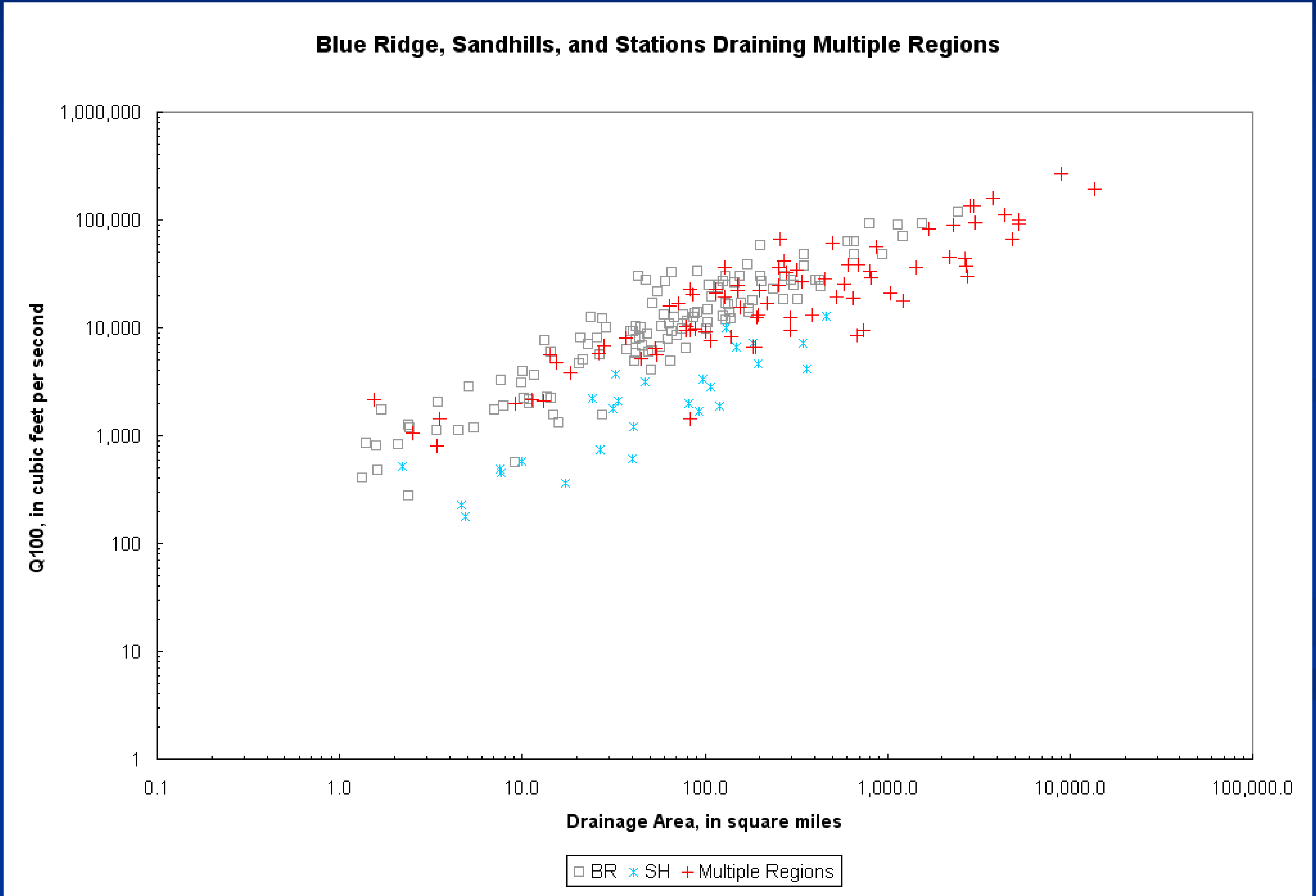
0022-1694/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2007.06.023

In our current study, we have done something similar. Instead of using a qualitative variable, we included percent region as a variable. Consequently, along with drainage area, slope, main channel length, etc., we have variables for %BR, %RV-PD, %SH, etc.

We also tested for statistically significant differences in the slopes of the regional curves and found that the Blue Ridge and Sandhills had slopes that were statistically different from the other regions. Consequently, we added a cross product of %BRxDA and %SHxDA. Those variables allow for a difference in the slopes of those regions.

What this allows us to do now is take advantage of a much larger range of hydrologic experiences while still accounting for the regional differences.

This study includes 83 stations that drain from multiple regions. In the past, these stations would not have been included in the regression analyses.



What do the preliminary equations look like.

For Q100:

$$Q100 = 10^{(0.02912*PCTRVPD + 0.02775*PCTBR + 0.02046*PCTSH + 0.02602*PCTCOAST + 0.02858*PCTSWGA)} \times DA^{(0.590 + 0.00120*PCTBR + 0.00139*PCTSH)}$$

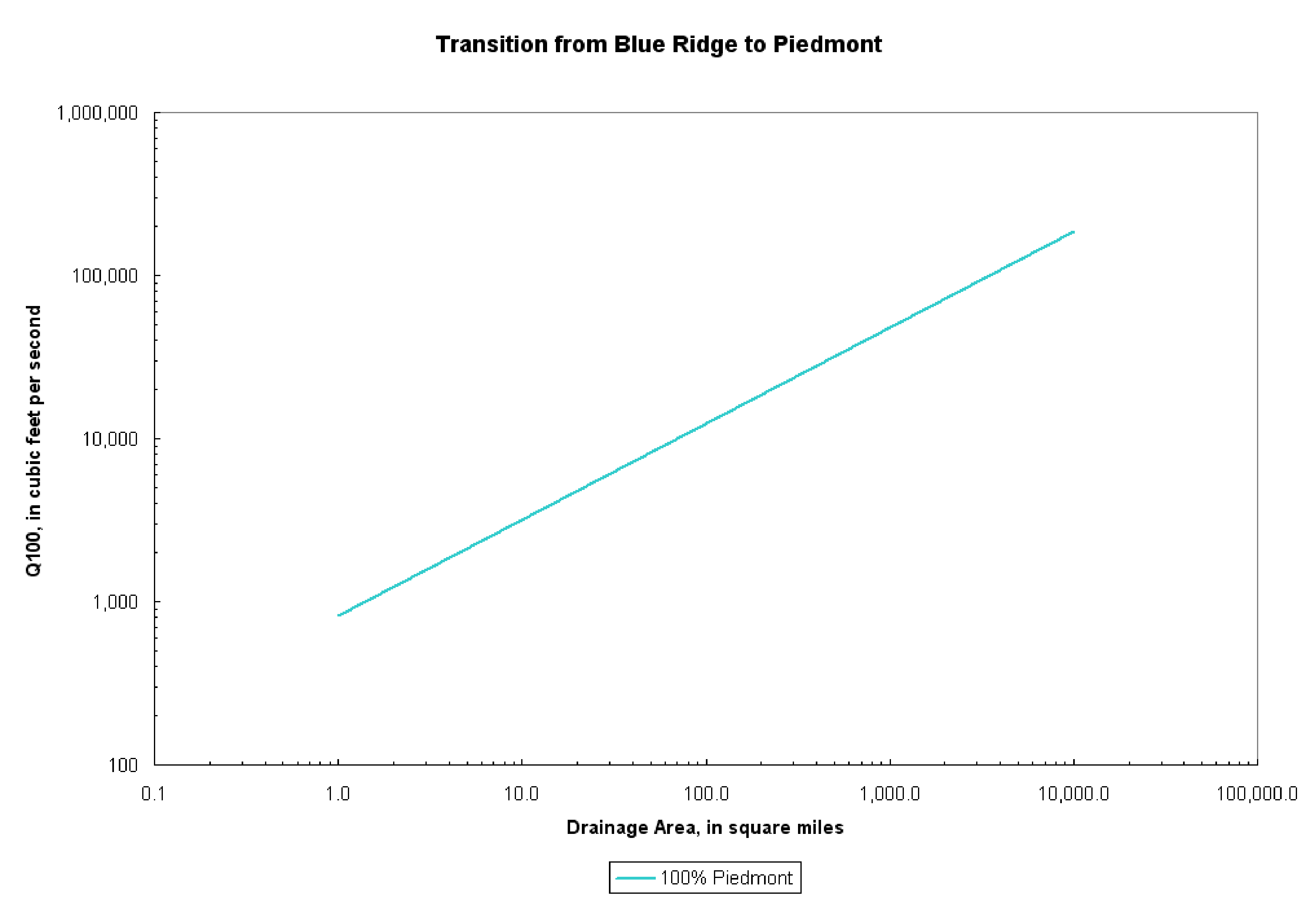
It looks a little scary but it’s really not that bad.

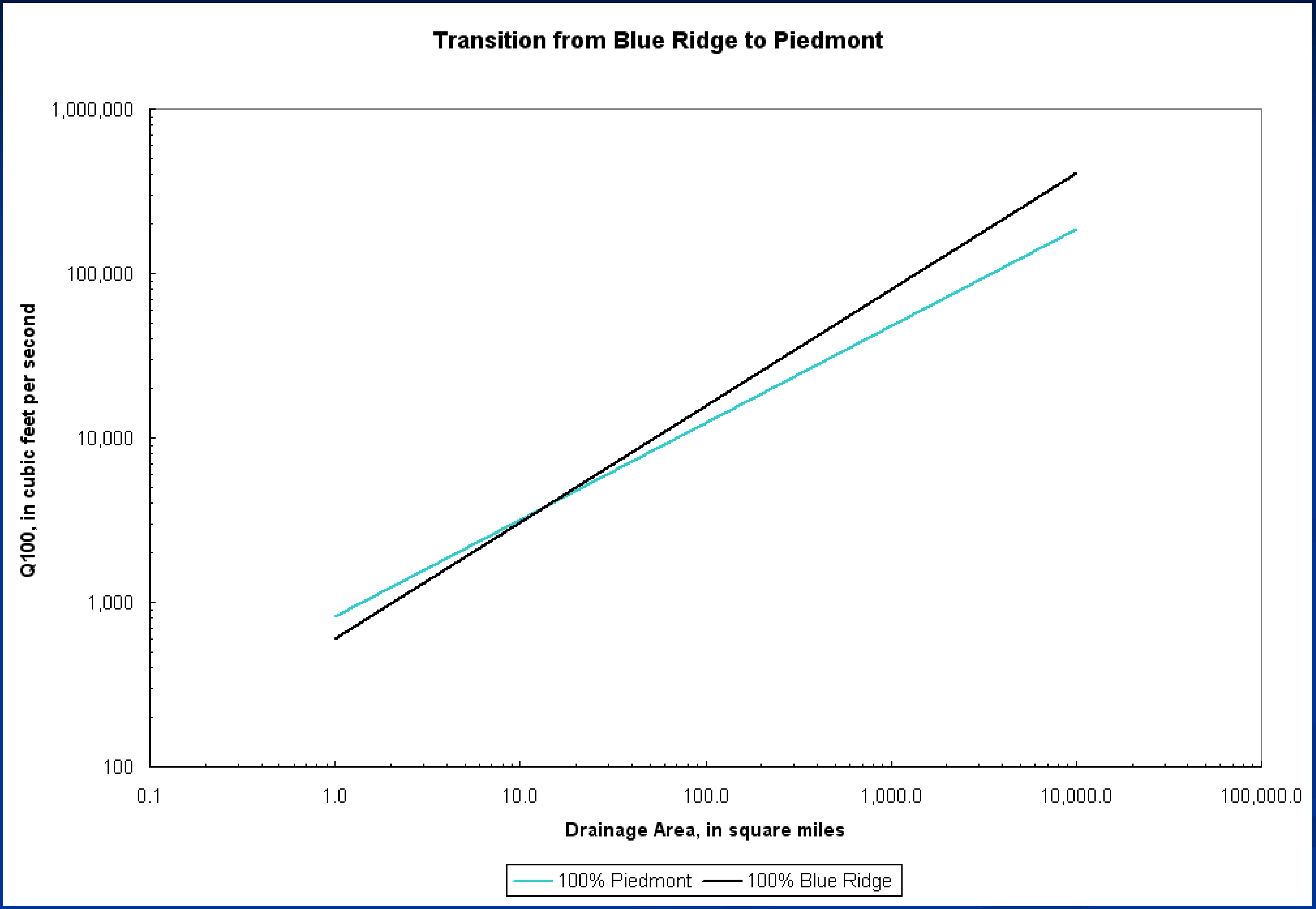
For 100% in the RV-PD region, Q100 collapses to:

$$Q100 = 817\ DA^{0.590}$$

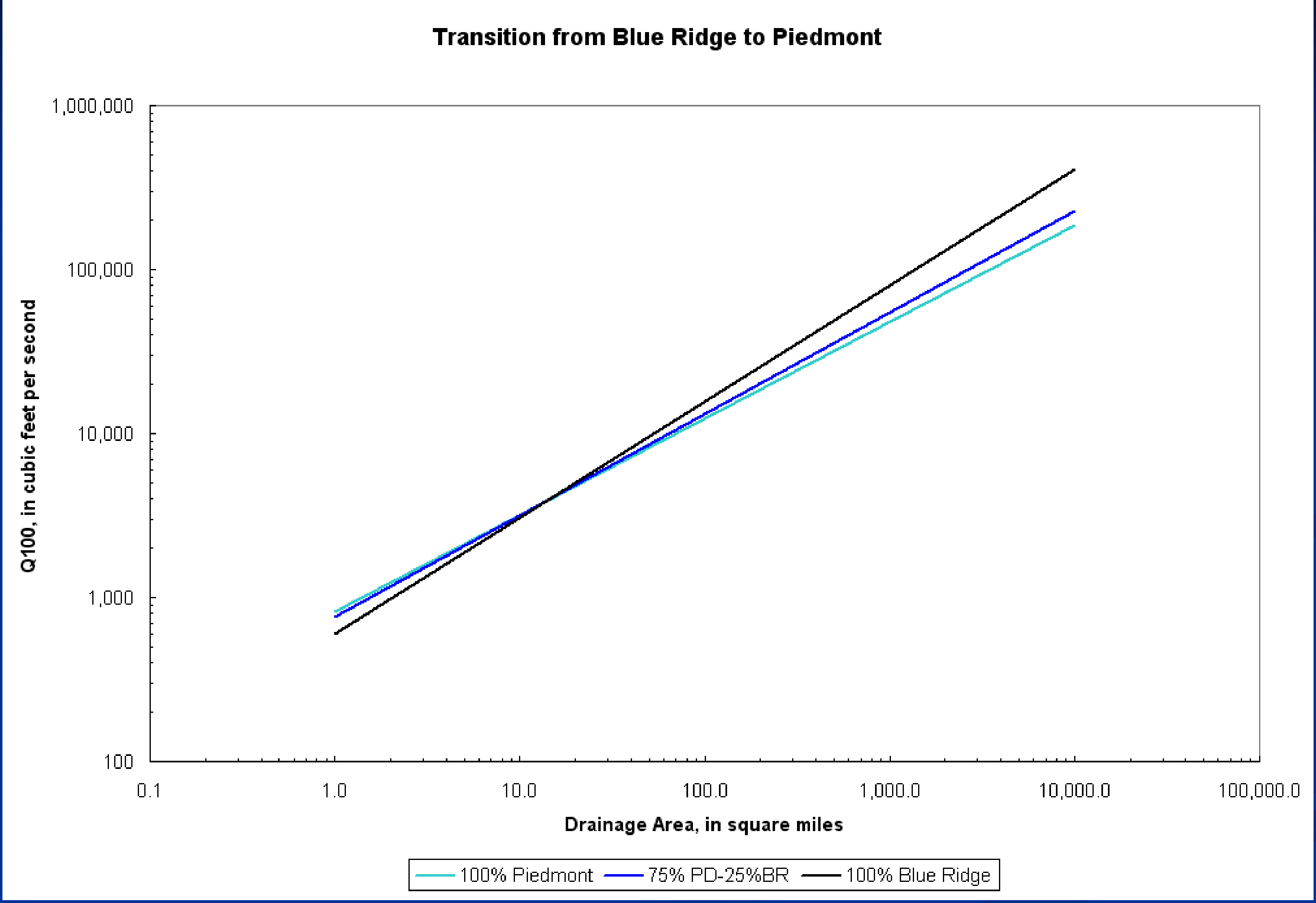
Muuuch Better!

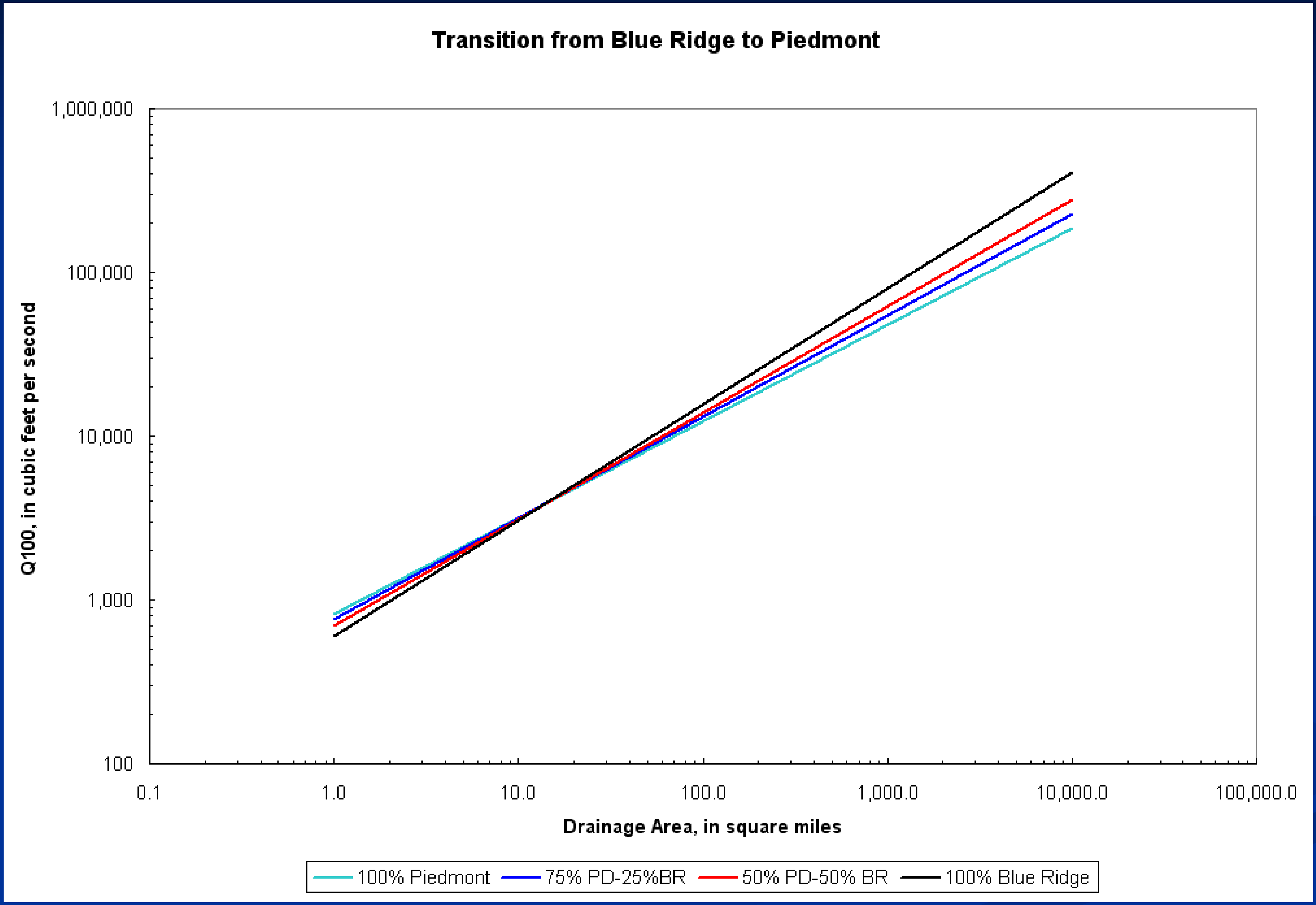
So how does the equation work?

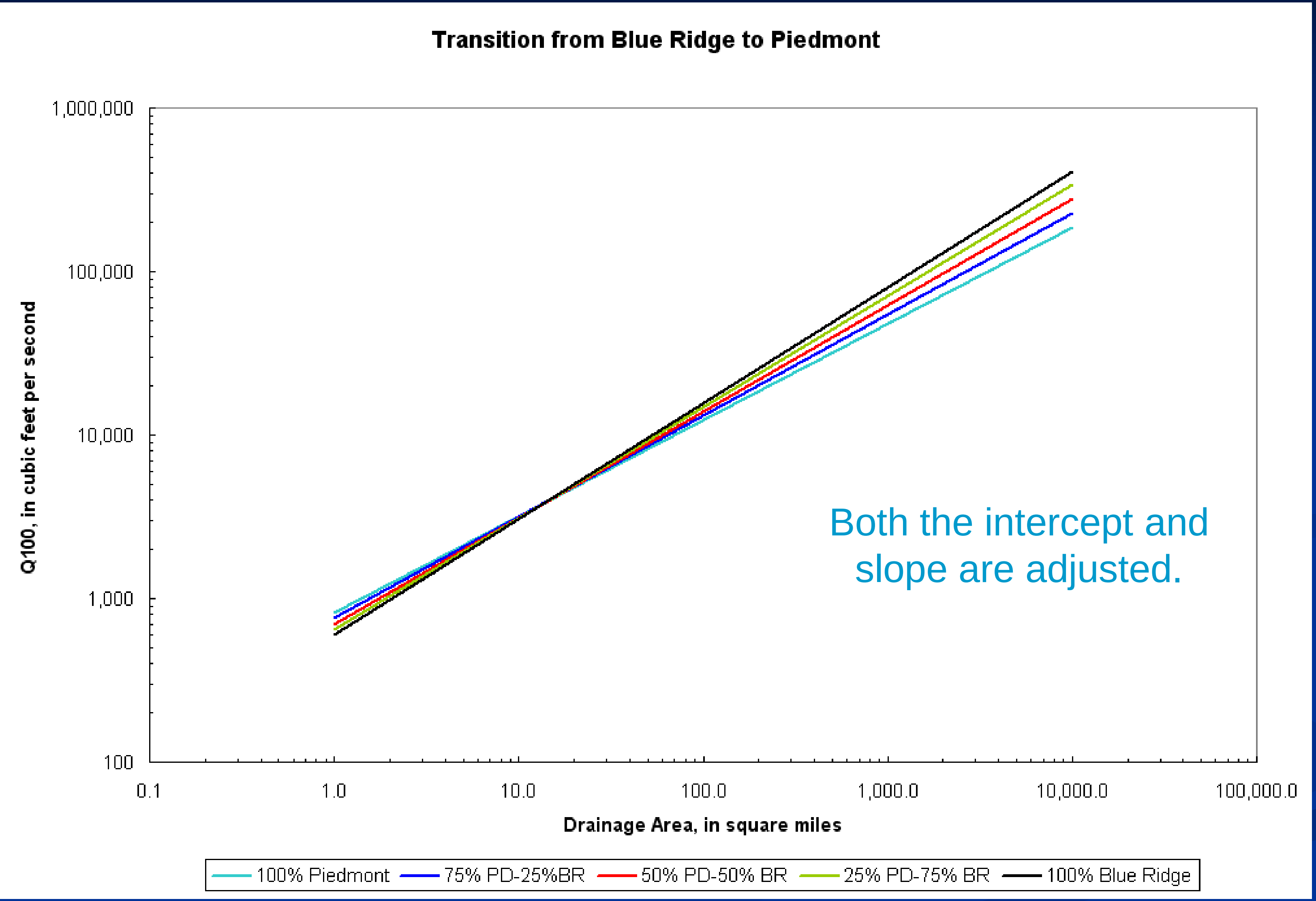



Transition from Blue Ridge to Piedmont
Q100, in cubic feet per second
1,000,000
100,000
10,000
1,000
100
0.1
1.0
10.0
100.0
1,000.0
10,000.0
100,000.0
Drainage Area, in square miles
100% Piedmont
100% Blue Ridge

USGS
science for a changing world

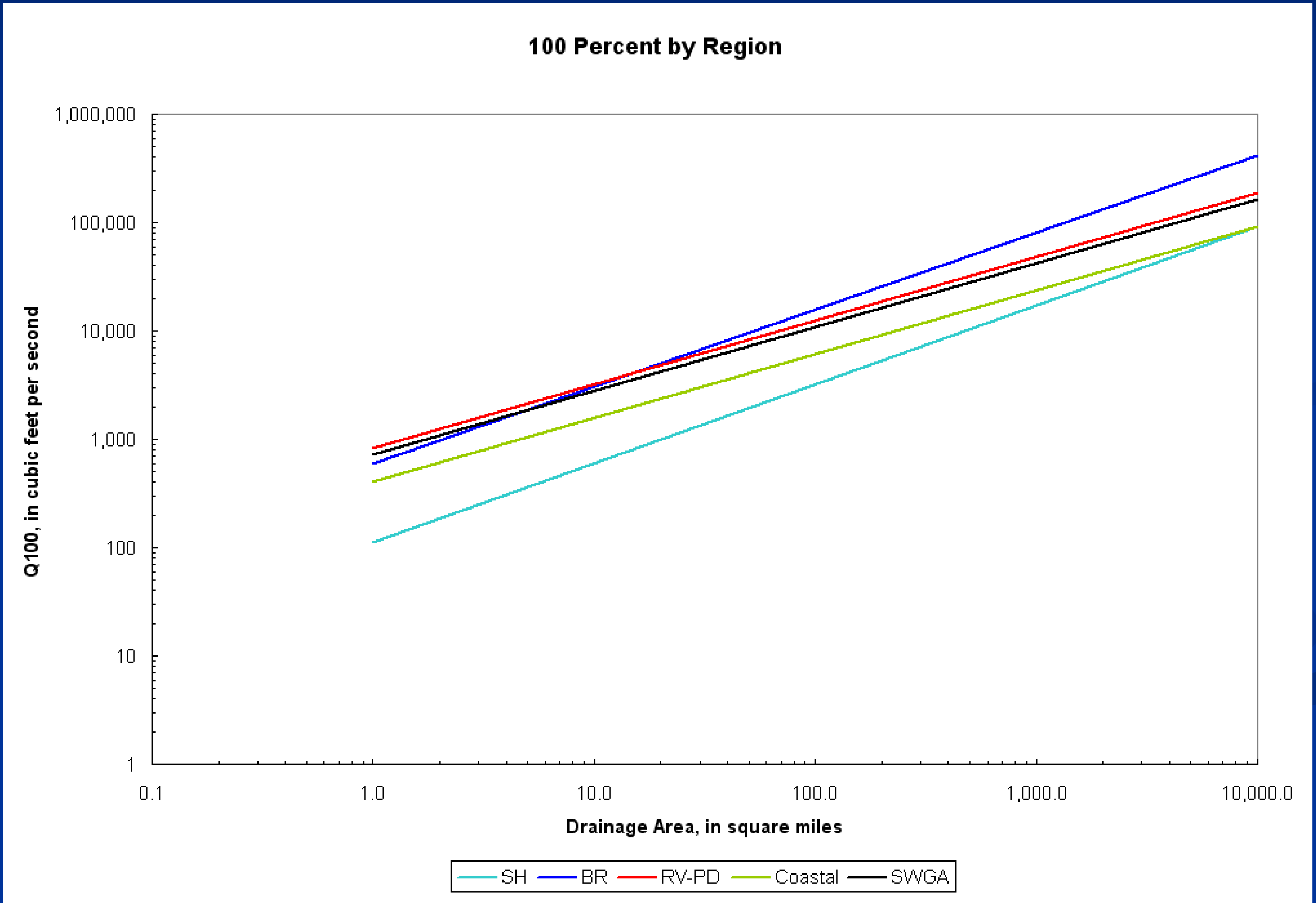



Transition from Blue Ridge to Piedmont
Q100, in cubic feet per second
1,000,000
100,000
10,000
1,000
100
0.1
1.0
10.0
100.0
1,000.0
10,000.0
100,000.0
Drainage Area, in square miles
100% Piedmont
75% PD-25%BR
50% PD-50% BR
100% Blue Ridge



Both the intercept and slope are adjusted.

These are the provisional curves for Q100 when a site drains 100% from each Hydrologic Region.



Acknowledgements

Tony Gotvald, GAWSC

Curtis Weaver, NCWSC

Larry Bohman, USGS-SE

Tim Cohn, USGS-OSW

Jery Stedinger, Cornell University

Andrea Gruber, Cornell University

Questions?