

Calculating Weighted Estimates of Peak Streamflow Statistics

According to the Federal guidelines for flood-frequency estimation, the uncertainty of peak streamflow statistics, such as the 1-percent annual exceedance probability (AEP) flow at a streamgage, can be reduced by combining the at-site estimate with the regional regression estimate to obtain a weighted estimate of the flow statistic. The procedure assumes the estimates are independent, which is reasonable in most practical situations. The purpose of this publication is to describe and make available a method for calculating a weighted estimate from the uncertainty or variance of the two independent estimates.

Data Sources

The at-site estimate of the logarithm of the peak streamflow statistic, Q_{as} , typically is calculated from the record of annual peak flows at the streamgage, according to methods described in Bulletin 17B (Interagency Committee on Water Data, 1981). A corresponding estimate, Q_{re} , is obtained from regression equations derived from a multivariate analysis of at-site values and explanatory basin characteristics from multiple streamgages in a region (Stedinger and Tasker, 1986).

If the mean square errors (MSE) of the two independent estimates are known, then an optimal weighted average is inversely proportional to the uncertainty of the individual estimates and can be calculated by the following:

$$\log(\hat{Q}) = \frac{\text{MSE}[\log(Q_{re})] \cdot \log(Q_{as}) + \text{MSE}[\log(Q_{as})] \cdot \log(Q_{re})}{\text{MSE}[\log(Q_{as})] + \text{MSE}[\log(Q_{re})]}$$

$\text{MSE}[\log(Q_{as})]$ and $\text{MSE}[\log(Q_{re})]$ represent the uncertainty corresponding to each estimate. When the MSE corresponding to $\log(Q_{as})$ is large, the uncertainty is also large, and so the optimal weight assigned to $\log(Q_{as})$ is relatively small. Conversely, when the MSE is small, the variance is also small, and so the optimal weight is correspondingly large.

Application of this weighting technique is easy if estimates of the quantiles, Q_{as} and Q_{re} , and the associated uncertainties, $\text{MSE}[\log(Q_{as})]$ and $\text{MSE}[\log(Q_{re})]$ are available. Because the $\text{MSE}[\log(Q_{as})]$ is not specifically reported at present (2012) from PEAKFQ, the Weighted Independent Estimates (WIE) program was developed by

the U.S. Geological Survey Office of Surface Water to compute this statistic and automate the process described above. It is anticipated that these statistics will be reported in future versions of PEAKFQ (Flynn and others, 2006) and WREG (Eng and others, 2009).

The WIE Program

The WIE program calculates the following:

- A. The variance of the at-site estimate using the Expected Moments Algorithm (Cohn and others, 2001, eq. 65). The log-Pearson Type III (LP3) distribution parameters that can be obtained from the PEAKFQ-generated output file must be supplied by the user. These parameters include the number of years of systematic record, skew type, MSE of the regional skew used, and the mean, standard deviation, and skew in log units for the streamgage.
- B. The variance of the regression estimate (Stedinger and Tasker, 1986) from either
 1. The variance at a specific streamgage, or
 2. The average variance of the regional regression model.

In either case, results from the regression model generated by WREG must be supplied by the user. These results include the standard error of the model, the number of explanatory variables, and the covariance matrix $[(U) \text{ or } (X^T \Lambda^{-1} X)]$ of the explanatory variables used in the regression equation.

C. The optimal weighted estimate, $\log(\hat{Q})$, and the corresponding MSE.

The WIE computer program is available as a package that includes an executable program file, detailed instructions for use, and other supporting files. The package containing the compressed files for the WIE program is available for download at <http://water.usgs.gov/usgs/osw/swstats/freq.html>.

The executable program *wie.exe* calculates the variances from the two independent estimates, the weighted average peak streamflow, and the associated variance. The program must be executed in a disk operating system (DOS) or Microsoft Windows system. To execute the WIE program after downloading, enter the program's name (*wie.exe*) in a DOS window or double-click the left mouse button. The user will be prompted to input data from the LP3 analysis of the at-site streamgage (results from PEAKFQ) and from a regional regression equation (results from WREG). Program results are displayed on the screen. All program results and inputs also are saved in a single output file named *wie_out.txt* and filed in the same directory or folder as the location of the executable program file.

Example Sessions for the WIE Program

In the subsequent paragraphs, two example sessions are given for the WIE program. The first example addresses the variance at a specific streamgage (Task B.1), and the second example addresses the average variance of the regional regression model (Task B.2). Data from the same streamgage are used in both examples.

A computer session for calculating the estimate using the variance at a specific streamgage is shown in figure 1; the numbers shown in red are entries specified by the user and are required to execute the program. The program begins by collecting data needed to calculate the variance from the PEAKFQ analysis. Next, WIE calculates and displays the peak streamflow estimate and the variance on the basis of the at-site streamgage data and the statistics that were entered manually. Finally, the program requests information needed from the regional regression equation to calculate the variance at the given streamgage (Task B.1). For this task, the user is asked to enter (1) the peak flow estimate at the streamgage from the regression model, (2) the average standard error of the regression model, (3) the number of explanatory variables in the regression model, (4) the covariance matrix ($X^T \Lambda^{-1} X$) of the regression model, row by row, and (5) the values of

the explanatory variables in log space. After these data are entered, the variance at the given streamgage is calculated as well as the weighted average peak streamflow estimate and the associated variance (weighted average). The upper and lower 95-percent confidence intervals also are calculated and displayed.

The second example is similar to the first except that the average variance of the regional regression model (Task B.2) is used. Figure 2 displays a computer session for the second example. Data from the PEAKFQ analysis in example 1 were also used in the second example; however, data that reflect calculations using the average variance of the regional regression model were entered. For the second example, the user is asked to enter the peak streamflow estimate from the regional regression equation and the standard error of prediction. The program then calculates the variance of the regional regression model and the weighted average peak streamflow estimate and the associated variance. Confidence intervals for the upper and lower 95-percent also are calculated and displayed.

References

- Cohn, T.A., Lane, W. L., and Stedinger, J.R., 2001, Confidence intervals for Expected Moments Algorithm flood quantile estimates: Water Resources Research, v. 37, no. 6, p. 1695–1706.
- Eng, Ken, Chen, Yin-Yu, and Kiang, J.E., 2009, User's guide to the Weighted-Multiple-Linear Regression Program (WREG version 1.0): U.S. Geological Survey Techniques and Methods, book 4, chap. A8, 40 p.
- Flynn, K.M., Kirby, W.H., and Hummel, P.R., 2006, User's manual for program PeakFQ, annual flood-frequency analysis using Bulletin 17B guidelines: U.S. Geological Survey Techniques and Methods, book 4, chap. B4, 42 p.
- Interagency Advisory Committee on Water Data, 1981, Guidelines for determining flood flow frequency: Bulletin 17B of the Hydrology Subcommittee, Office of Water Data Coordination, U.S. Geological Survey, Reston, Va., revised March 1982, 183 p.
- Stedinger, J.R., and Tasker, G.D., 1986, Regional hydrologic analysis, 2. Model-error estimators, estimation of sigma and log-Pearson Type III distributions: Water Resources Research, v. 22, no. 10, p. 1487–1499.

By Timothy A. Cohn, Charles Berenbrock,
Julie E. Kiang, and Robert R. Mason, Jr.

For questions about calculating weighted estimates, please contact
h2osoft@usgs.gov

```

~~~~~

Weighting of Independent Estimates from
PeakFQ and Regional Regression Model

~~~~~

~~~~~

Enter PeakFQ information
-----
Enter station number
13186000
Enter number of observations
53
Enter fitted parameters of M, S, G (Log units)
3.628 0.1789 -0.361
Enter skew type (W=Weighted, S=Station, G=Generalized)
Please enter a W, S, or G
W
Enter mse of generalized skew
0.150
Enter non-exceedance probability to be estimated
(for example, 0.01 for the 1-percent chance flood)
0.010

estimate = 9923. (cfs)
log of estimate = 3.9967
variance of estimate (log) = 0.22620E-02

-----
Enter Regional Regression Equation info
-----
Enter regression estimate (cfs)
11451.

Depending on availability of RRE data, select the
appropriate method to use to calculate the
variance of the RRE

(1) Station (At-Site) predictors
Need: U matrix, Xo matrix, and SEM

(2) Basin wide average predictors
Need: SEP

Enter 1 or 2:
1

Enter the average standard error of model (SEM):
0.1588

What is the type and/or units of SEM?
1--log units (average)
2--percent (average)
3--percent (upper or lower)
1

Calculated variance of model (SEM) = 0.25217E-01

How many explanatory variables are in the RRE?
3
Input all U matrix values for row 1
0.33276 -0.10035E-01 -0.18763 -0.20337E-01
Input all U matrix values for row 2
-0.10035E-01 0.21217E-02 0.53802E-02 -0.22203E-02
Input all U matrix values for row 3
-0.18763 0.53802E-02 0.11659 -0.38839E-03
Input all U matrix values for row 4
-0.20337E-01 -0.22203E-02 -0.38839E-03 0.20716E-01

Input the explanatory values for the site in log units
Input explanatory value 1 (log units)
2.8073
Input explanatory value 2 (log units)
1.5406
Input explanatory value 3 (log units)
1.3522

Calculated variance of sampling error (log) = 0.26747E-02
Total variance of prediction of RRE (log) = 0.27892E-01

-----
Results from weighting
-----
Weighted estimate (cfs) = 10030.
log(weighted estimate) = 4.001

upper 95% confidence interval (cfs) 12330.
lower 95% confidence interval (cfs) 8160.

variance of estimate (log) = 0.20923E-02

Press "Enter" to end program

```

Figure 1. A WIE screen session for calculating the weighted estimate of peak streamflow-frequency statistics using the variance at a specific streamgage. (Input data are in red print.)

```

~ ~ ~ ~ ~
Weighting of Independent Estimates from
PeakFQ and Regional Regression Model
~ ~ ~ ~ ~

~ ~ ~ ~ ~
Enter PeakFQ information
-----
Enter station number
13186000
Enter number of observations
53
Enter fitted parameters of M, S, G (Log units)
3.628 0.1789 -0.361
Enter skew type (W=Weighted, S=Station, G=Generalized)
Please enter a W, S, or G
W
Enter mse of generalized skew
0.150
Enter non-exceedance probability to be estimated
(for example, 0.01 for the 1-percent chance flood)
0.010

estimate = 9923. (cfs)
log of estimate = 3.9967
variance of estimate (log) = 0.22620E-02

-----
Enter Regional Regression Equation info
-----
Enter regression estimate (cfs)
11451.

Depending on availability of RRE data, select the
appropriate method to use to calculate the
variance of the RRE

(1) Station (At-Site) predictors
Need: U matrix, Xo matrix, and SEM

(2) Basin wide average predictors
Need: SEP

Enter 1 or 2:
2
Enter the average standard error of prediction (SEP):
48.4

What is the type and/or units of SEP?
1--log units (average)
2--percent (average)
3--percent (upper or lower)
3

Calculated variance of prediction (log) = 0.29390E-01

-----
Results from weighting
-----
Weighted estimate (cfs) = 10025.
log(weighted estimate) = 4.001

upper 95% confidence interval (cfs) 12329.
lower 95% confidence interval (cfs) 8152.

variance of estimate (log) = 0.21003E-02

Press "Enter" to end program

```

Figure 2. A WIE screen session for calculating the weighted estimate of peak streamflow-frequency statistics using the average variance of the regional regression model. (Input data are in red print.)