***OSW Informational and Technical Note* 2014.43** **September 30, 2014**

Subject: Weighting Estimates of Peak-Flow-Frequency Statistics when Using the Expected Moments Algorithm for At-Site Flood-Frequency Estimation

This technical note updates Office of Surface Water (OSW) Technical Memorandum 2010.05 with information for recent updates of PeakFQ and WREG.

Bulletin 17B (IACWD, 1981), the federal guidelines for flood-frequency estimation, notes that the uncertainty of peak flow statistics, such as the one-percent annual exceedance probability (AEP) flow at a streamgage (or site), can be reduced by combining the at-site estimate with an independent regional estimate to obtain a weighted estimate of the flow statistic at the site. As discussed in previous memo 2010.05, the program Weighted Independent Estimates (WIE) (Cohn and others, 2012) was developed to assist in this process. Newer versions of PeakFQ version 7.1 (Veilleux and others, 2014) and WREG version 1.05 (Eng and others, 2009) now exist and provide additional output to make the weighting calculations more straightforward. This technical note shows how to calculate such weighted estimates in practice without use of the WIE program. The OSW support of WIE will be phased out over the next few years.

Weighting Equation

For both frequency analysis and regression, the peak flow statistic is commonly transformed using base-10 logarithms:

$$Y_i = log_{10}\mathrm{Q}_i \tag{1}$$

Where Q_i is the peak flow statistic in cubic feet per second (cfs) at site i, and Y_i is the log-transformed variable. All subsequent operations are performed on the transformed variable, Y_i. As described in Appendix 8 of Bulletin 17B (IAWCD, 1982) (with slightly different notation), the weighted estimate can be calculated using variances as:

$$Y_{weighted,i} = \frac{Y_{Site,i}*V_{Reg,i}+Y_{Reg,i}*V_{Site,i}}{V_{Site,i}+V_{Reg,i}} \tag{2}$$

Where all Y and V variables are in log_{10} units, $Y_{weighted,i}$ is the weighted estimate for site *i*; $Y_{Site,i}$ is the at-site estimate at site *i*; $Y_{Reg,i}$ is the regression estimate at site *i*; $V_{Site,i}$ is the variance of the at-site estimate at site *i*; and $V_{Reg,i}$ is the variance of the regression estimate at site *i*. For independent $Y_{Site,i}$ and $Y_{Reg,i}$, the variance of the weighted estimate for site *i* can be calculated as (with all V variables in log_{10} units):

$$V_{weighted,i} = \frac{V_{Site,i}*V_{Reg,i}}{V_{Site,i}+V_{Reg,i}} \tag{3}$$

A 95 percent confidence interval on the weighted estimate can be calculated as:

$$95\%_CI_{upper,i} = 10^{\left(Y_{weighted,i}+1.96\sqrt{V_{weighted,i}}\right)} \tag{4}$$
$$95\%_CI_{lower,i} = 10^{\left(Y_{weighted,i}-1.96\sqrt{V_{weighted,i}}\right)}$$

Note that $Y_{weighted,i}$,$V_{weighted,i}$, and $95\%_CI_i$ must be calculated separately for each site *i* for each annual exceedance probability of interest.

Application

Application of this weighting technique is straightforward if the quantile estimates ($Y_{Reg,i}$ and $Y_{Site,i}$) and their variances ($V_{Reg,i}$ and $V_{Site,i}$) are known. Because the $V_{Site,i}$ was not specifically reported by PeakFQ for a flood frequency analysis using Bulletin 17B methods, the Weighted Independent Estimates (WIE) program was developed to compute this statistic and automate the weighting process described above. However, when performing a flood-frequency analysis using the Expected Moment Algorithm (EMA) (Cohn and others, 2001) in the new PeakFQ version 7.1, $V_{Site,i}$ is now reported. Also, WREG now reports $V_{Reg,i}$ for the individual sites used to develop the regional regression equation. Thus, WIE should no longer be used compute weighted estimates when performing a flood-frequency analysis using EMA; instead the steps provided below should be followed.

Data Sources for At-Site Estimates:

The parameters necessary for the weighted annual exceedance probabilities are calculated in PeakFQ version 7.1. When performing an EMA analysis in PeakFQ 7.1 the following parameters will be provided in the output:

A) $Y_{Site,i}$: The EMA at-site estimate for the peak streamflow statistic for a set of annual exceedance probabilities. These values can be found in the Annual Frequency Curve – Discharges at Selected Exceedance Probabilities table in the output file.
B) $V_{Site,i}$: The variance corresponding to the EMA at-site estimate for the peak streamflow statistics. These values can be found in the Annual Frequency Curve – Discharges at Selected Exceedance Probabilities table in the output file. Note that the variances given in the table are already in log-space.

Data Sources for Regional Estimates:

Results from the regression model generated by WREG are also needed to generate weighted estimates. The following parameters are provided or can be calculated from the WREG output:

1) If using results from WREG v1.05 or newer:
   A) $Y_{Reg,i}$: The regional estimate of the peak streamflow statistic for site i.
   B) $V_{Reg,i}$ for gages used in the regression model: The variance of predication (VP) for each site used in the regional regression model should be used as the $V_{Reg,i}$ in the weighting equation for those gage sites used in the regional regression model. Note the variances are reported in the WREG output in log-space.
   C) $V_{Reg,i}$ for gages *not* used in the regression model: The average variance of prediction (AVP) should be used as the $V_{Reg,i}$ in the weighting equation for sites not used in the regional regression model. Note the AVP is reported in the WREG output in log-space. Also note that if the site was not used in development of the regression equations because it has characteristics unlike all other sites in the regression, then the regression should not be used to estimate $Y_{Reg,i}$ or $V_{Reg,i}$.

2) If using results from an older version of WREG:
   A) $Y_{Reg,i}$: The regression estimate of the peak streamflow statistic, $Y_{Reg,i}$, for each site is not output from WREG and must be calculated outside of WREG by substituting the dependent variables for site i into the regional quantile model generated by WREG.

   B) $V_{Reg,i}$ for gages used in the regression model: The $V_{Reg,i}$ for each site is not output from WREG. Thus, the user has two options. The first, and simplest option, is to use the average variance of prediction (AVP) as an estimate for the $V_{Reg,i}$ for each site i used in the regional regression (see C below for details on how to determine AVP). The second, more complicated option is to calculate the variance of prediction (VP) for each site used in the regional regression model. The VP is an accurate representation of $V_{Reg,i}$ for each site i used in the regional regression. The VP for each site is not output from older versions of WREG and must be calculated outside of WREG using the following equation.

$$VP_i = \sigma_\delta^2 + x_i(X^T \Lambda^{-1} X)^{-1} x_i^T \quad (5)$$

   Where VP_i is the variance of prediction for the annual exceedance probability of interest at site *i*, σ_δ^2 is the model error variance for the annual exceedance probability of interest, x_i is the (1 x *k*) vector of independent variables for the ith site augmented by a value of one, $(X^T \Lambda^{-1} X)^{-1}$ is the (*k* x *k*) covariance matrix of regression parameters output by WREG in the file *InvXLX.txt*, *k* is the number of independent variables in the WREG regression model including the constant.

   Older versions of WREG do not directly output the model error variance; however it can be calculated from the Standard Model Error. By using equation 33 on page 5 of the WREG user's guide (Eng and others, 2009), the Standard Model Error for the annual exceedance probability of interest, S_{me} , can be transformed into the model error variance.

   From equation 33 in WREG manual on page 33

$$S_{me} = 100\left\{e^{[(\ln 10)^2 \sigma_\delta^2]} - 1\right\}^{1/2} \quad (6)$$

   Or rearranged to solve for model error variance, σ_δ^2

$$\sigma_\delta^2 = \frac{\ln\left\{\left(\frac{S_{me}}{100}\right)^2 + 1\right\}}{(\ln 10)^2} \quad (7)$$

   By substituting equation 7 into equation 5, the variance of prediction can be calculated for each site used in the regional regression.

   C) $V_{Reg,i}$ for gages *not* used in the regression model: The average variance of prediction (AVP) for the regional regression model should be used as the $V_{Reg,i}$ in the weighting equation for those sites not used in the regional regression model. However, older versions of WREG do not directly output the model error variance; they do output the Average Standard Error of Prediction. By using equation 33 on page 5 of the WREG user's guide (Eng and others, 2009), the Average Standard Error of Prediction for the

annual exceedance probability of interest, S_P, can be transformed into the Average Variance of Prediction.

From equation 33 on page 33 of the WREG user's guide (Eng and others, 2009)

$$S_P = 100\left\{e^{[(\ln 10)^2 AVP]} - 1\right\}^{1/2} \tag{8}$$

Or rearranged to solve for average variance of prediction, AVP

$$AVP = \frac{\ln\left\{\left(\frac{S_P}{100}\right)^2 + 1\right\}}{(\ln 10)^2} \tag{9}$$

References

Cohn, T. A., W. L. Lane, and Stedinger, J. R., 2001, Confidence intervals for expected moments algorithm flood quantile estimates, *Water Resources Research*, 37(6), 1695– 1706.

Cohn, T.A., Berenbrock, Charles, Kiang, J.E., and Mason, R.R., Jr., 2012, Calculating weighted estimates of peak streamflow statistics: U.S. Geological Survey Fact Sheet 2012–3038, 4 p., available at *http://pubs.usgs.gov/fs/2012/3038/.*

Eng, Ken, Chen, Yin-Yu, and Kiang, Julie, 2009, User's guide to the weighted-multiple-linear-regression program (WREG version 1.0): U.S. Geological Survey Techniques and Methods, book 4, chapter A8, 21 p., *http://pubs.usgs.gov/tm/tm4a8.*

Interagency Advisory Committee on Water Data (IAWCD), 1982, Guidelines for determining flood-flow frequency: Bulletin 17B of the Hydrology Subcommittee, Office of Water Data Coordination, U.S. Geological Survey, Reston, VA., 183 p., *http://water.usgs.gov/osw/bulletin17b/bulletin_17B.html*.

Stedinger, J. R., and Tasker, G. D., 1986, Regional hydrologic analysis: 2. Model-error estimators, estimation of sigma and log-Pearson type 3 distributions, *Water Resources Research*, 22(10), 1487–1499.

U.S. Geological Survey, 2014, PeakFQ – Peak flow FreQuency program: U.S. Geological Survey, accessed August 29, 2014, at *http://water.usgs.gov/software/PeakFQ/.*

U.S. Geological Survey, 2013, WREG—Weighted-multiple-linear regression program: U.S. Geological Survey, accessed August 29, 2014, at *http://water.usgs.gov/software/WREG/.*

Veilleux, A.G., Cohn, T.A., Flynn, K.M., Mason, R.R., Jr., and Hummel, P.R., 2014, Estimating magnitude and frequency of floods using the PeakFQ 7.0 program: U.S. Geological Survey Fact Sheet 2013-3108, 2 p., *http://pubs.usgs.gov/fs/2013/3108/.*

Example: Calculating Weighted 0.01 Annual Exceedance Probability Estimate at a Site using output from PeakFQ and WREG

At-Site Estimates:
Below is the *Annual Frequency Curve – Discharges at Selected Exceedance Probabilities* table from a PeakFQ output for site 1.

Table E1: Annual Frequency Curve – Discharges at Selected Exceedance Probabilities as output from PeakFQ version 7.1 for site 1.

ANNUAL FREQUENCY CURVE -- DISCHARGES AT SELECTED EXCEEDANCE PROBABILITIES					
			<----------FOR EMA ESTIMATES---------->		
ANNUAL EXCEEDANCE PROBABILITY	EMA W/ REG INFO ESTIAMTE	EMA W/O REG INFO ESTIAMTE	VARIANCE OF EST.	95% CONFIDENCE INTERVALS LOWER	UPPER
0.9950	6.8	7.7	0.1446	0.2	18.5
0.9900	9.4	10.3	0.1118	0.5	22.4
0.9500	21.3	22.4	0.0524	2.6	38.5
0.9000	32.4	33.3	0.0341	5.8	52.5
0.8000	52.4	52.9	0.0201	12.9	78.4
0.6667	80.4	80.3	0.0127	24.3	117.3
0.5000	123.2	122.3	0.0089	62.2	183.3
0.4292	146.2	144.8	0.0082	84.8	220.9
0.2000	267	265.2	0.0085	178.8	448.7
0.1000	387.3	388	0.0112	255.2	767.9
0.0400	562.5	571.7	0.0165	353.4	1467
0.0200	707	727.3	0.0218	423.6	2277
0.0100	861.2	897.2	0.0281	488.8	3414
0.0050	1025	1082	0.0353	549.2	4944
0.0020	1254	1348	0.0464	621.4	7818

$Y_{Site,1}$ for the 0.01 AEP estimate for site 1 is highlighted in yellow in table E1. $V_{Site,1}$ for the 0.01 AEP estimate for site 1 is highlighted in green in table E1. Thus,

$$Y_{Site,1} = log_{10}(861.2) = 2.935 \quad \text{(E1)}$$

$$V_{Site,1} = 0.0281 \quad \text{(E2)}$$

Regional Estimates:

Below is the output from WREG for a regression model for the 0.01 AEP. Two output files are shown, one from WREG v1.05, the newest version, and one from WREG v1.04. The files have been shortened to just focus on the output pertaining to the regression model and the output for observation 1.

Table E2: WREG output from versions 1.05 and 1.04 for a regression model for the 0.01 AEP.

WREG v1.05
last modified April 19, 2013

Regression Model for Q1%

Performance Metrics		AVP
MSE of unweighted residuals	0.086	(log units)
Avg Variance of Prediction	0.071	(log units)
Avg Standard Error of Prediction	67.38	(percent)
Pseudo R2	81.63	(percent)
Model Error Variance	0.062	(log units)
Standard Model Error	62.55	(percent)

Coefficients of Model

		Std Error	T value	P>|T|
Constant	1.166	0.78	1.50	0.142
DRNAREA	0.515	0.05	10.93	0.000
PRECIP	0.005	0.00	2.00	0.053

Model Coeffs

Transformations		C1	C2	C3	C4
Q1%	log10	1.00	1.00	0.00	1.00
DRNAREA	log10	1.00	1.00	0.00	1.00
PRECIP	1	10.00	1.00	0.00	1.00

Residuals, Leverage, and Influence of Observations

Leverage Limit	0.154
Influence Limit	0.103

Observation	Residual	Leverage	Influence
1	0.137	0.268	0.041
.	.	.	.
.	.	.	.
.	.	.	.

Variance of prediction for streamgages used in regression

Observation	ObsValue	PredValue	VarianceofPrediction
1	2.994	2.856	0.085
.	.	.	.
.	.	.	.
.	.	.	.

$Y_{Reg,1}$ $VP_1=V_{Reg,1}$

WREG v1.04
last modified February 9, 2012

Regression Model for Q1%

Performance Metrics		
MSE of unweighted residuals	0.086	(log units)
Avg Standard Error of Prediction	67.38	(percent)
Pseudo R2	81.63	(percent)
Standard Model Error	62.55	(percent)

S_p S_{me}

Coefficients of Model

		Std Error	T value	P>|T|
Constant	1.166	0.78	1.50	0.142
DRNAREA	0.515	0.05	10.93	0.000
PRECIP	0.005	0.00	2.00	0.053

Model Coeffs

Transformations		C1	C2	C3	C4
Q1%	log10	1.00	1.00	0.00	1.00
DRNAREA	log10	1.00	1.00	0.00	1.00
PRECIP	1	10.00	1.00	0.00	1.00

Residuals, Leverage, and Influence of Observations

Leverage Limit	0.154
Influence Limit	0.103

Observation	Residual	Leverage	Influence
1	0.137	0.268	0.041
2	0.288	0.133	0.092
.	.	.	.
.	.	.	.
.	.	.	.

A) Regional 0.01 AEP Estimate

1) If using WREG v1.05, then the regional estimate is provided in the output for all sites used to develop the regional regression model. As shown in Table E2, $Y_{Reg,1}$ is highlighted in blue.

$$Y_{Reg,1} = 2.856 \tag{E3}$$

2) If using WREG v1.04 or earlier, or determining a regional estimate for a site not used to develop the regional regression model, then $Y_{Reg,i}$ must be computed outside of WREG by plugging the basin characteristics for the site of interest into the regional regression model developed by WREG. The computations to determine $Y_{Reg,1}$ are shown below.

   Use regional model for 0.01 AEP from WREG and site 1 basin characteristics to determine $Y_{Reg,1}$ for site 1. For site 1, DA = 6.368 and Precip = 26.231. The regional regression coefficients for the regional model are provided in the WREG output and are shown in Table E2 highlighted in pink.

   Regional Model:

$$Y_{Reg,1} = 1.166 + 0.515[log_{10}(DA)] + 0.005[10(Precip)] \quad \text{(E4)}$$

Substitute in Site 1 dependent variable values:

$$Y_{Reg,1} = 1.166 + 0.515[log_{10}(6.368)] + 0.005[10(26.231)] \quad \text{(E5)}$$

$$Y_{Reg,1} = 2.892 \quad \text{(E6)}$$

(Note: The $Y_{Reg,1}$ value calculated from WREG v1.04 above is different from the $Y_{Reg,1}$ value from WREG v1.05 due to rounding.)

B) Variance of Regional 0.01 AEP Estimate

1) If using WREG v1.05, then the variance of the regional estimate is provided in the output for all sites used to develop the regional regression model. As shown in the Table E2, $V_{Reg,1}$ is highlighted in gray.

$$V_{Reg,1} = VP_1 = 0.085 \quad \text{(E7)}$$

If using WREG v1.05, then the AVP should be used as the variance of the regional estimate for a site not used to develop regional regression model. As shown in the Table E2, the AVP is provided and is highlighted in green.

$V_{Reg,j} = AVP = 0.071$ where site *j* is a site not used in the regression model (E8)

2) If using WREG v1.04 or earlier, then $V_{Reg,i}$ must be computed outside of WREG by either computing the AVP or the VP. The AVP should be used as $V_{Reg,i}$ for a site not used to develop the regional regression model and can be used as a computationally simple estimate for $V_{Reg,i}$ for a site used to develop the regional regression model.

The Average Standard Error of Prediction, S_P is provided in the WREG output and is highlighted in Table E2 in orange. By using equation 9, the AVP can be calculated for the 0.01 AEP.

$$AVP = \frac{\ln\left\{\left(\frac{67.38}{100}\right)^2 + 1\right\}}{(\ln 10)^2} = 0.071 \quad \text{(E9)}$$

$$V_{Reg,1} = AVP = 0.071 \quad \text{(E10)}$$

The VP can be used for $V_{Reg,i}$ for sites used in the regional regression model if the hydrologist performs complex computations needed to obtain this accurate estimate.

The computations to determine $Y_{Reg,1}$ from the WREG v1.04 output are shown below. The Standard Model Error Variance, S_{me} is provided in the WREG output and is highlighted in Table E2 in purple. By using equation 7 the model error variance can be calculated.

$$\sigma_\delta^2 = \frac{\ln\left\{\left(\frac{62.55}{100}\right)^2 + 1\right\}}{(\ln 10)^2} = 0.0623 \quad \text{(E11)}$$

WREG also provides the covariance matrix of regression parameters output by WREG in the file *InvXLX.txt*

$$(X^T \Lambda^{-1} X)^{-1} = \begin{bmatrix} 6.0174 \times 10^{-1} & 1.4652 \times 10^{-3} & -1.8692 \times 10^{-3} \\ 1.4652 \times 10^{-3} & 2.2219 \times 10^{-3} & -1.4028 \times 10^{-5} \\ -1.8692 \times 10^{-3} & -1.4028 \times 10^{-5} & 5.8799 \times 10^{-6} \end{bmatrix} \quad \text{(E12)}$$

Then the vector $x_{i,}$needs to be created. This vector contains a leading one for the constant followed by any transformed dependent variables used in the regional model. In this example the regional model uses two transformed variables $log_{10}(DA)$ and $[10(Precip)]$. Thus, vector x_1 is.

$$x_1 = [1 \quad 0.804 \quad 262.31] \quad \text{(E13)}$$

The second term of equation 5 can be computed by substituting in equation E12 and E13.

$$x_1 (X^T \Lambda^{-1} X)^{-1} x_1^T =$$

$$[1 \quad 0.804 \quad 262.31] \begin{bmatrix} 6.0174 \times 10^{-1} & 1.4652 \times 10^{-3} & -1.8692 \times 10^{-3} \\ 1.4652 \times 10^{-3} & 2.2219 \times 10^{-3} & -1.4028 \times 10^{-5} \\ -1.8692 \times 10^{-3} & -1.4028 \times 10^{-5} & 5.8799 \times 10^{-6} \end{bmatrix} \begin{bmatrix} 1 \\ 0.804 \\ 262.31 \end{bmatrix} \quad \text{(E14)}$$

$$x_1 (X^T \Lambda^{-1} X)^{-1} x_1^T = 0.0236 \quad \text{(E15)}$$

By substituting the results from equations E11 and E15 into equation 5, VP_1 is equal to

$$V_{Reg,1} = VP_1 = 0.0623 + 0.0236 = 0.0859 \quad \text{(E16)}$$

(Note: The $V_{Reg,1}$ value calculated from WREG v1.04 above is different from the $V_{Reg,1}$ value from WREG v1.05 due to rounding.)

Weighted Estimates

Now the weighted 0.01 AEP estimate as well as its variance and confidence intervals can be calculated for site 1 by substituting the at-site and regional values calculated above into equations 1-4.

/Signed/

Andrea Veilleux
for the Office of Surface Water

Distribution: GS-W WSC Surface Water Specialists
GS-W OSW Staff All

OSW notes are archived on the Office of Surface Water Sharepoint site,
https://xcollaboration.usgs.gov/wq/osw/OSWNotes/SitePages/Home.aspx