Sensitivity of regional confidence intervals for floods

Robert Whitley

Department of Mathematics, University of California, Irvine, CA 92717, USA

T. V. Hromadka II

Williamson and Schmid, 17782 Sky Park Blvd, Irvine, CA 92714, USA

A problem arising in the use of regionalized data for the estimation of the effects of future urbanization on a catchment is the construction of confidence intervals for estimates of peak discharge for various durations. This discharge is modeled by a log Pearson III distribution in which both the skew and standard deviation are obtained from a pool of regional data and so are known to a much higher accuracy than the mean. The sensitivity of the confidence intervals to the values chosen for the mean standard deviations is discussed.

Key Words: *T*-year flood, confidence interval, log Pearson III distribution, regionalization, sensitivity, skew, standard deviation.

INTRODUCTION

In 'Regional Confidence Intervals for Floods', by Whitley and Hromadka II[1], confidence intervals are given for the estimation of *T*-year values for values of discharge which are modeled by a log Pearson III distribution for which *the skew and standard deviation are given while* the mean is estimated from data. In the problem studied, estimates for the skew and standard deviation were obtained from regional data and so were estimated from a much larger data pool than the mean (see Ref. 2) and the model assumption that these two values are known is intended to reflect the presumed higher accuracy implicit in their determination from regional data.

To assume that the skew and standard deviation are known does not, of course, mean that in practice estimates of them do not vary: in fact, the selection of sites to include in the set of regional data, as well as various ways of estimating regional skew, from contour 'skew maps' as well as from the data, can cause significant variation in the values used. Consequently it is important to see how sensitive the confidence intervals obtained are to changes in these values.

DISCUSSION

The problem discussed in Ref. 1 is as follows: Discharges Q for a set of durations are modeled by means of a log Pearson III distribution, the three unknown parameters of which can be determined from the mean, standard deviation, and skew of $X \approx \log Q$. It is desired to estimate the T-year values of the discharge for various values of T, i.e. for $p = 1 - 1/T$, to estimate the pth quantile value y_p for X, defined by

$$P(X \leqq y_p) = p$$

Given the value σ of the standard deviation and the value γ of the skew, and having an estimate $\hat{\mu}$ for the unknown mean μ, $100q\%$ confidence intervals for y_p were computed in Ref. 1; they are of the form

$$y_p \leqq \hat{\mu} + \sigma C(m, p, q, \gamma) \tag{1}$$

where $C = C(m, p, q, \gamma)$ is a function of the number of years of record m used in estimating μ, the confidence level q desired, $p = 1 - 1/T$, and the skew γ. A $100q\%$ confidence interval for the T-year value of the discharge Q is then

$$Q \leqq 10^{\hat{\mu}} 10^{\sigma C} \tag{2}$$

The specific estimate for $\hat{\mu}$ enters into the confidence interval in a simple way, so it will be enough to look at the values of the factor $10^{\sigma C}$ which multiplies $10^{\hat{\mu}}$ in equation (2). The data given in Table 1 for peak Q gives an indication of the range of variation in σ and γ that should be considered.

Table 1. Peak discharge data ($10^{\hat{\mu}}$ in cfs) for catchments

Name	$\hat{\mu}$	$10^{\hat{\mu}}$	m	σ	γ
Alhambra	3.442	2270	46	0.206	−0.189
Arcadia	3.142	1390	16	0.258	0.309
Brown	2.381	240	11	0.663	−0.056
Compton 1	3.435	2720	47	0.212	−0.159
Compton 2	3.330	2140	20	0.171	0.830
Dominguez	3.932	8550	10	0.185	−0.890
Eaton	3.354	2260	13	0.228	0.503
Rubio	3.269	1860	41	0.201	−0.660

The regional values used, $\sigma = 0.21$ and $\gamma = -0.30$, were obtained from the data of Table 1 as well as similar data for the durations 5, 15 and 30 minutes and 1, 2, 3, 6, 12 and 24 hours.

The factors $10^{\sigma C}$ of equation (2) were computed for a set of values $m = 10, 50$ and $T = 10, 100$, for the range of values $\sigma = 0.1, 0.2, 0.3$ and skew values of $\gamma = -2, -1, -, 1, 2$. These tables are included as an appendix.

Paper accepted September 1989. Discussion closes January 1991.

The variation of the factors with σ is an easy direct consequence of equation (2), since C does not depend on σ. From the appendix:

Table 2. Factors $10^{\sigma C}$ of equation (2)

Confidence level $q=85\%$
Skew $\gamma=0$

	$\sigma=0.1$			$\sigma=0.2$			$\sigma=0.3$	
	$m=10$	$m=50$		$m=10$	$m=50$		$m=10$	$m=50$
$T=10$	1.45	1.39	$T=10$	2.10	1.93	$T=10$	3.04	2.68
$T=100$	1.84	1.77	$T=100$	3.40	3.12	$T=100$	6.26	5.52

For example, to compute the 85% confidence value for the 100 year value ($T=100$) of the peak discharge with $\sigma=0.2$ and $m=10$ years of record, multiply the value $10^{\hat{\mu}}$ by the factor 3.40. Because $\{10^{\cdot 1C}\}^2=10^{\cdot 2C}$ and $\{10^{\cdot 1C}\}^3=10^{\cdot 3C}$, the second table is, to within roundoff, the square of the first table and the third is the cube of the first; e.g. if σ is twice what you thought it was, then the factors you multiply by to get confidence intervals are really the square of what you thought they were.

The variation in skew involves a more complex dependence on γ. To get an idea of this, use the number in the appendix to obtain:

Table 3. Ratio of factors for $\sigma=0.2$

(a) Factor for $\gamma=1$ divided by the factor for $\gamma=-1$

	$q=0.15$			$q=0.50$			$q=0.85$	
	$m=10$	$m=50$		$m=10$	$m=50$		$m=10$	$m=50$
$T=10$	1.10	1.10	$T=10$	1.12	1.10	$T=10$	1.10	1.10
$T=100$	1.93	1.93	$T=100$	1.96	1.94	$T=100$	1.93	1.93

(b) Factor for $\gamma=2$ divided by the factor for $\gamma=-2$

	$q=0.15$			$q=0.50$			$q=0.85$	
	$m=10$	$m=50$		$m=10$	$m=50$		$m=10$	$m=50$
$T=10$	1.19	1.19	$T=10$	1.19	1.20	$T=10$	1.19	1.19
$T=100$	3.32	3.33	$T=100$	3.43	3.35	$T=100$	3.32	3.33

For example, from the appendix, the factor you multiply by to get an 85% confidence interval endpoint for $T=100$, $m=10$, $\sigma=0.2$, and $\gamma=-1$ is 2.430. If instead of $\gamma=-1$, the true value of γ were $+1$, then by Table 3 (a) the actual value to multiply by is $2.430\times 1.93=4.69$.

REFERENCES

1 Whitley, R. and Hromadka II, T. V. Regional confidence intervals for floods, to appear
2 Hromadka II, T. V. and McCuen, R. H. Investigation of mitigation needs for changes in duration floodflows due to development, prepared for Coastal Community Builders, 1989, in-review

```
      program mu2
c     Robert Whitley, Dept. of Math., Univ. of CA, Irvine, CA 92717
c     T.V. Hromadka II, Williamson and Schmidt, 17782 Sky Park Blvd.,
c     Irvine, Ca 92714
c
c     This program calculates 15%,50%,and 85% confidence intervals
c     for return periods T=10 and 50 years for m=10 and 50 years of
c     record, for three values of standard deviation .1,.2,and .3, and
c     for five values of skew -2,-1,0,1 and 2; all for log base 10
c     values. This is done to test the sensitivity of the model, which
c     is also used in the program muonly, to variations in the standard
c     deviation and the skew. The factors the program calculates are
c     multiplied times the estimated discharge (for the duration under
c     consideration) in order to get the desired confidence intervals.
c     The distribution of values is modeled by a log Pearson III
c     distribution where the skew and the standard deviation are given
c     by regional data, and the entire statistical variation is assumed
c     to come from the variation in the sample mean. the confidence
c     intervals for the log data have the form   sample mean +B(p,q)),
c     where B(p,q) is a function of the number of years of record, the
c     regional skew, the regional standard deviation, p=1-1/T, and
c     confidence level q. Consequently the discharge values for a given
c     sample mean are just 10 to the power (sample mean) times 10 to the
c     power B(p,q). The factors in the output are 10 to the power
c     B(p,q).

      external C
      real B(2,6),C,p(2),q(3),sigma(3),skew(5),tmp
      integer i1,i2,i3,i4,i5,m(2)
      character*10 ans
c     percentiles q for confidence intervals
      q(1)=.15
      q(2)=.50
      q(3)=.85
```

```
c       values for T year floods, T=10,100, p=1-1/T
        p(1)=.90
        p(2)=.99

c       values for standard deviation
        sigma(1)=.1
        sigma(2)=.2
        sigma(3)=.3

c       values for skew
        skew(1)=-2.0
        skew(2)=-1.0
        skew(3)=0.0
        skew(4)=1.0
        skew(5)=2.0
c       values for number of years of record m
        m(1)=10
        m(2)=50
        open(unit=2,file='prn')

        do 50 i1=1,3
c       q = confidence level do above
        do 40 i2=1,5
c       skew value do above
        do 30 i3=1,3
c       sigma value do above
        do 20 i4=1,2
c       T-year value do above
        do 10 i5=1,2
c       number of years of record do above
           tmp=C(m(i5),skew(i2),sigma(i3),p(i4),q(i1))
           B(i4,i5+2*(i3-1))=exp(log(10)*tmp)
   10   continue
   20   continue
   30   continue
           write(2,35) 100*q(i1),skew(i2)
   35   format(1x,"confidence level = ",F4.0,"%","      skew = ",F6.2)
        call printdata(B)
   40   continue
        print *,"adjust paper. input y"
        read *,ans
   50   continue
        close(unit=2,file='prn')
        end
c       program end

        real function normal(s)

c       Uses table lookup to return the value t such that for Z a N(0,1)
c       random variable, P(Z<t)=s

        real s,small

        small=.0005
        if (abs(s-.05) .le. small) then
             normal=-1.64485
        else if (abs(s-.15) .le. small) then
             normal=-1.03643
        else if (abs(s-.50) .le. small) then
             normal=.0
        else if(abs(s-.80) .le. small ) then
             normal=.84162
        else if(abs(s-.85) .le. small ) then
             normal=1.03643
```

```
      else if(abs(s-.9Ø) .le. small ) then
           normal=1.28155
      else if(abs(s-.95) .le. small ) then
           normal=1.64485
      else if(abs(s-.96) .le. small ) then
           normal=1.75Ø69
      else if(abs(s-.98) .le. small ) then
           normal=2.Ø5375
      else if(abs(s-.99) .le. small ) then
           normal=2.32635
      else if(abs(s-.995) .le. small ) then
           normal=2.57583
      else
           print *,"normal percentile not in table"
           stop
      endif
      return
      end

      real function WH(b,x)
c     Uses the Wilson-Hilferty approximation in order to compute the
c     x-th percentile for the distribution Gamma(b,1); if Y has such a
c     distribution, then WH(b,x)=t is a value with, approximately,
c     Pr(Y<t)=x. See Kendall and Stuart, The Advanced Theory of
c     Statistics, Vol. I, page 4ØØ and see Mathur, A Note on Wilson-
c     Hilferty Transformation of Chi-squared, Bull. of the Calcutta
c     Stat. Assoc. 1Ø(1961) 1Ø3-1Ø5 for a discussion of the error.

      real b,temp,x

      temp=1-(1/(9*b))+x*sqrt(1/(9*b))
      temp=exp(3*log(temp))
      WH=b*temp

      return
      end

      real function C(mm,sskew,ssigma,pp,qq)

c     For a given confidence level pp and T year value qq, returns a
c     value C so that C plus the estimated mean is the left end point of
c     the 1ØØ pp % confidence interval for the return frequency
c     qq=1-1/T.


      real a,b,normal,pp,qq,sskew,ssigma,small,WH
      external normal,WH
      integer mm

      small = .Ø5

      if (abs(sskew) .le. small) then
           C=ssigma*normal(pp)-ssigma*(-normal(qq))/sqrt(mm)
      endif

      if (sskew .gt. small) then
           b=4/(sskew*sskew)
           a=ssigma/sqrt(b)
           C=a*WH(b,normal(pp))-a*WH(mm*b,-normal(qq))/mm
      endif

      if (sskew .lt. -small) then
           b=4/(sskew*sskew)
```

```
            a=-ssigma/sqrt(b)
            C=a*WH(b,-normal(pp))-a*WH(mm*b,normal(qq))/mm
      endif
      return
      end
      end

      subroutine printdata(B)
c     prints a set of confidence factors, i.e. factors which you
c     multiply estimated discharge by to obtain confidence intervals.
c     For given values of confidence level q and skew, tables are
c     printed of these factors for T year values, T=1Ø,1ØØ, number of
c     years of record m=1Ø,5Ø, and standard deviations .1, .2, and .3.

c     main program needs to open unit=2, file='prn'

      integer i,T(2)
      real B(2,6)
      character*8Ø  head1,head2

      T(1)=1Ø
      T(2)=1ØØ

      head1="              sigma =.1           sigma =.2           sigma =.3"
      head2="              m=1Ø    m=5Ø        m=1Ø   m=5Ø         m=1Ø    m=5Ø"
      write(2,5) head1
    5 format(1x,A)
      write(2,6) head2
    6 format(1x,A)

      do 2Ø i=1,2
         write(2,1Ø) T(i),B(i,1),B(i,2),B(i,3),B(i,4),B(i,5),B(i,6)
   1Ø    format(1x,"T=", F4.Ø,6(F8.3))
   2Ø continue
      write(2,'(1x,/)')
      return
      end
```

APPENDIX

confidence level = 15.% skew = -2.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.143	1.190	1.308	1.416	1.495	1.686
T=100.	1.169	1.217	1.367	1.481	1.599	1.803

confidence level = 15.% skew = -1.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.203	1.253	1.447	1.571	1.740	1.969
T=100.	1.341	1.398	1.799	1.954	2.414	2.731

confidence level = 15.% skew = 0.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.246	1.299	1.552	1.687	1.933	2.190
T=100.	1.584	1.652	2.510	2.729	3.977	4.507

confidence level = 15.% skew = 1.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.261	1.315	1.590	1.728	2.005	2.272
T=100.	1.863	1.943	3.472	3.773	6.469	7.330

confidence level = 15.% skew = 2.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.246	1.298	1.553	1.685	1.935	2.187
T=100.	2.131	2.220	4.542	4.929	9.678	10.943

confidence level = 50.% skew = -2.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.221	1.229	1.492	1.510	1.822	1.856
T=100.	1.249	1.257	1.560	1.579	1.948	1.984

confidence level = 50.% skew = -1.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.291	1.295	1.667	1.678	2.153	2.173
T=100.	1.440	1.445	2.074	2.087	2.987	3.014

confidence level = 50.% skew = 0.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.343	1.343	1.804	1.804	2.424	2.424
T=100.	1.709	1.709	2.919	2.919	4.988	4.988

confidence level = 50.% skew = 1.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.365	1.361	1.863	1.852	2.544	2.520
T=100.	2.017	2.011	4.068	4.043	8.205	8.130

confidence level = 50.% skew = 2.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.353	1.345	1.830	1.808	2.476	2.431
T=100.	2.314	2.300	5.354	5.289	12.387	12.164

confidence level = 85.% skew = -2.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.326	1.273	1.759	1.620	2.332	2.063
T=100.	1.356	1.302	1.839	1.694	2.494	2.206

confidence level = 85.% skew = -1.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.398	1.341	1.954	1.798	2.731	2.410
T=100.	1.559	1.495	2.430	2.236	3.788	3.343

confidence level = 85.% skew = 0.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.449	1.389	2.098	1.930	3.039	2.682
T=100.	1.843	1.767	3.395	3.123	6.255	5.519

confidence level = 85.% skew = 1.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.466	1.406	2.148	1.978	3.148	2.782
T=100.	2.165	2.078	4.689	4.318	10.153	8.973

confidence level = 85.% skew = 2.00

	sigma =.1		sigma =.2		sigma =.3	
	m=10	m=50	m=10	m=50	m=10	m=50
T= 10.	1.445	1.388	2.088	1.928	3.018	2.676
T=100.	2.472	2.375	6.108	5.638	15.097	13.389