Flood frequency censoring errors associated with daily-read flood observations

Tom Micevski, George Kuczera, and Stewart W. Franks

Discipline of Civil, Surveying and Environmental Engineering, University of Newcastle, Callaghan, New South Wales, Australia

Received 8 December 2004; revised 18 March 2005; accepted 12 April 2005; published 6 July 2005.

[1] The effects of daily-read discharges, discharges derived from a single daily observation (say at 9 AM), are examined in the context of flood frequency analysis. These discharges underestimate the true peak discharge because the actual peak is unlikely to occur at the single observation time. The resultant error in discharge estimates can be quite large, up to a factor of 10. Several maximum likelihood methods are proposed to deal with these daily-read, or censored, discharges. Three methods are evaluated: censored as gauged (CAG), which treats daily-read discharges as the true peak, binomial censoring (BC), which uses binomial censoring, and random dependent censoring (RDC), which incorporates the dependence between the the daily-read and true peak discharges. The performance of the methods is evaluated using a Monte Carlo study. The RDC method is found preferable to the CAG method because it provides better performance (in terms of root mean squared error and bias) and is theoretically sound. The BC method performs very poorly and is not recommended.


1. Introduction

[2] Data collection forms an important part of flood frequency analysis. Often, the data that is sought for analysis is the annual maximum flood series; that is, the peak instantaneous discharge over a (water) year. This is normally obtained as the maximum of the recorded daily discharges for that year. However, in long gauged records, the discharges in the early part of the record may in fact be the discharge associated with a specific reading time; such a discharge is referred to as the daily-read discharge. This is not an uncommon occurrence. For example, some annual maximum flood discharges from New South Wales (NSW), Australia, (especially pre-1970s) are actually daily-read discharges; in such cases, river levels, and thus flood peaks, were manually recorded at, typically, 9 AM [Department of Land and Water Conservation, 2000]. This practice is now virtually nonexistent due to the introduction of continuous (fixed interval) data loggers at almost every stream gauging location. Nonetheless, the older data will remain affected by this practice.

[3] This study assesses the impact of daily-read data on the integrity of the flood frequency analysis. Daily-read flows can be classified as censored-from-below observations or, alternatively, as random censoring thresholds, where censoring is on the right. The study will examine the data errors that result from this censoring. It will then explore the consequences of these errors within the context of flood frequency analysis. A new technique, based on the method of maximum likelihood, is developed to rigorously deal with daily-read data.

2. Relationship Between True Peak and Daily-Read Flows

[4] An understanding of the relationship between true peak and daily-read flows is developed empirically through examination of the continuous portion of several flood gauging records. Data from NSW, Australia, was used for this investigation and was obtained from the Pinneena database [Department of Land and Water Conservation, 2000]. The data extracted consisted of river levels and rating curves.

[5] For each day with recorded continuous river levels, the peak and “daily reading time” water levels were extracted. Selection of the daily reading time is somewhat arbitrary, with 9 AM being used because it corresponds to the typical reading time used in the earlier periods of the data set. The appropriate rating curve was then used to convert these water levels into discharges. The maximum values of the peak flow and daily-read flow, for each year, were then recorded and the ratio \( R \) of these two flows calculated, where \( R = (\text{true peak flow})/(\text{daily-read flow}) \). These two maximum flows do not necessarily correspond to the same flood event; however, in practice, it was found that they were in correspondence about 85% of the time. This analysis was performed for several catchments of varying size and the results are summarized in Table 1, with the results for two sites displayed in Figure 1. Note that the dashed line in Figure 1 is the “perfect correction line”; the closer that the points are plotted to this line, the closer that the daily-read flows are to the true peak flows.

[6] Table 1 reveals that the statistics of the ratio \( R \) are quite variable across the sites. No general rule, such as the ratio (\( R \)) reducing as catchment area increases, is evident, suggesting that catchment specific (local) con-
ditions exert a strong influence. It is shown that daily-read flows can significantly underestimate the true peak flow, with two sites, 203002 and 210006, displaying some very large differences, some greater than a factor of 10. These two sites have vastly different areas, with 203002 being quite small (62 km²) while 210006 relatively large (3340 km²), further indicating the influence of catchment specific conditions. Also, site 422001 has a ratio of one indicating that there is no difference between daily-read and true peak flows. This occurs because the catchment is very large and responds slowly. In such cases, it is appropriate to treat the daily-read flows as true peak flows. Note that selecting a different reading time, other than 9 AM, would cause the points on the graphs to be shifted horizontally; the daily-read flows would either increase or decrease, while the peak flows remain the same. This should not significantly alter the analysis presented herein.

The stream water level and discharge time series for site 210006 is presented in Figure 2 for the event where the peak flow was 10 times greater than the daily-read (9 AM) flow. It shows that the reason for the large anomaly between the censored and peak flows is that daily (9 AM) readings are far too infrequent to accurately capture the dynamic response of this catchment. Indeed, this peak would require readings spaced at intervals of around 2 hours to provide an adequate approximation.

Additional examination of the ratios using the reduced ratio (\( r = R - 1 \)) suggests that there is a structure within these ratios. The coefficients of variation (CV = (standard deviation)/mean) of \( r \) are listed in Table 1 and show that this statistic is quite stable (approximately one) for most sites, except for the two sites that experience extremely large differences between true peak and daily-read flows. These two sites obviously have greater variability; thus the increased CV is not unexpected. Furthermore, quantile-quantile (Q-Q) plots, shown for two sites in Figure 3, suggest that the reduced ratio follows a lognormal distribution; that is, the relationship between the true peak flow and the daily-read flow can be transformed to approximate a common statistical distribution. This is referred to as the censoring distribution. Note that the Q-Q probability limits were derived using a parametric bootstrap procedure.

### 3. Evaluation of Methods

It was shown that estimating peak flows using daily-read flows can result in large errors. Four methods, that can be used to deal with these errors, are now examined. Evaluation of these methods will be performed using a Monte Carlo experiment.

These methods, all based on the maximum likelihood approach, are explained in greater detail in sections 3.1–3.5 and summarized as follows: (1) exclude erroneous data (identify and exclude the daily-read flows, analyzing only the true peak flows), (2) ignore data errors (treat the daily-read flows as if they were true peak flows, ignoring the associated data errors), (3) binomial censoring (use the binomial censoring procedure of Stedinger and Cohn [1986] with multiple random thresholds), and (4) random dependent censoring (use a new procedure incorporating random censoring thresholds and the statistical dependence between the daily-read and true peak flows).

#### 3.1. Monte Carlo Experiment

The flood distribution selected for the Monte Carlo experiment was the lognormal distribution. This distribution was selected because previous studies have shown that the lognormal distribution provides an adequate fit to IPO (Interdecadal Pacific Oscillation) stratified flood data from NSW [Franks and Kuczera, 2002; Micevski et al., 2003]. The IPO is a climate index of long-term Pacific Ocean sea surface temperature anomalies [Folland et al., 1999] that has been shown to modulate the effects of El Niño–Southern Oscillation within Australia [Power et al., 1999; Kiem et al., 2003].

The experiment describes the following scenario involving a total record length of \( n \) flow observations consisting of \( g \) true peak flows \( Q = \{Q_1, \ldots, Q_g\} \) and \( c \) daily-read flows \( q = \{q_{g+1}, \ldots, q_n\} \), where \( n = g + c \); it is stressed that \( g \) and \( c \) are known constants. The true peak flows are assumed to follow a lognormal distribution,
log \( Q \sim N(\mu, \sigma^2) \), and the true peak and daily-read flows are related by the censoring distribution, \( \log r \sim N(m, s^2) \), where \( r = R - 1 = Q/q - 1 \) is the reduced ratio. Note that the terms gauged and censored flows are used interchangeably for true peak and daily-read flows respectively. The steps used within the Monte Carlo experiment are as follows.

[13] 1. Randomly generate the true peak flow \( Q \) over the entire record length \( n \), \( \log Q_i \sim N(\mu, \sigma^2) \), for \( i = 1, \ldots, n \); where \( \sim \) denotes “is sampled from.”

[14] 2. Randomly censor \( c \) flows to obtain the daily-read flow \( q_j \), for \( j = g + 1, \ldots, n \). (1) Randomly generate the censoring ratio \( R = Q/q = r + 1 \), \( \log r \sim N(m, s^2) \), where \( r \) is the reduced ratio from section 2. (2) Compute the daily-read flow, \( q_j = Q_j/R_j \).

[15] 3. Estimate the parameters of the lognormal flood distribution \((\mu, \sigma)\) using the four methods.

[16] 4. Repeat steps 1–3 \( N = 1000 \) times.

[17] 5. Calculate the relative bias and rmse (root mean squared error) of selected flow quantiles, expressed as a percentage of the true quantiles (\( \hat{\theta} \) and \( \theta \) denote the true value and estimate of a quantile):

\[
\%\text{bias} = \frac{100}{N\theta} \sum \left( \hat{\theta} - \theta \right)
\]

\[
\%\text{rmse} = \frac{100}{\theta} \sqrt{\sum \left( \hat{\theta} - \theta \right)^2 / N}.
\]

[18] For NSW flood data, based on approximately 50 sites, the average length of an IPO stratified flood record is about 30 and 40 years for IPO negative and positive epochs respectively. Hence the total record length used in the analysis is 40 years. Additionally, these stratified data sets have a mean, \( \mu \), and standard deviation, \( \sigma \), of approximately 10 and 1.2 respectively (log flow in ML/d).

3.2. Method 0: Gauged Only

[19] The gauged only (GO) method does not make any attempt to handle censored data; instead, the censored flows are excluded, with only the gauged flows being analyzed. This method is used to provide a basis on which to judge the benefits provided by including the censored flows. The likelihood function is

\[
p(Q, q|\mu, \sigma) = \prod_{j=1}^{g} p(\log Q_j|\mu, \sigma) \prod_{j=g+1}^{n} p(\log q_j|\mu, \sigma^2)
\]

where \( p(\log x|\mu, \sigma^2) \) denotes a lognormal pdf evaluated at \( x \) with known parameters \( \mu \) and \( \sigma \).

3.3. Method 1: Censored as Gauged

[20] The censored as gauged (CAG) method ignores the error associated with the censoring and treats the daily-read flow as if it were the true peak flow. The likelihood function simply becomes

\[
p(Q, q|\mu, \sigma) = \prod_{j=1}^{g} p(\log Q_j|\mu, \sigma^2) \prod_{j=g+1}^{n} p(\log q_j|\mu, \sigma^2)
\]

where \( p(\log x|\mu, \sigma^2) \) denotes a lognormal pdf evaluated at \( x \) with the underlying normal distribution having mean \( \mu \) and standard deviation \( \sigma \).

![Figure 2](image2.png)

**Figure 2.** Stage and hydrograph of an event showing a factor of 10 difference between the daily-read (9 AM) and the true peak discharges for site 210006.

![Figure 3](image3.png)

**Figure 3.** Quantile-quantile plots, with 90% probability limits, of the censoring ratio \( R = r + 1 \) for two sites assuming a lognormal censoring distribution, \( \log r \sim N(m, s^2) \) (top and bottom sites are 210001 and 210006).
3.4. Method 2: Binomial Censoring

A daily-read flow understimates the true peak flow because it is unlikely that the true peak flow occurred at the daily reading time, say 9 AM. This situation is analogous to that of censored data where the daily-read flow is the censoring threshold, with the true peak flow being unobserved. Indeed, each daily-read flow is independent of another, resulting in what is known as random censoring [e.g., Schneider, 1986, pp. 2–3]. This suggests that the binomial censoring procedure of Stedinger and Cohn [1986] may be applicable; each daily-read flow becomes binomial censoring (BC) method. Significantly, however, it accounts for the dependence between the threshold (daily-read flow) and true peak flow. The added complexities of the multiple random thresholds present in the data, as in the random censoring (RC) method, involves a new approach for censored flows. It incorporates the random dependent censoring (RDC) method.

The likelihood for a single daily-read flow can be written as

\[ p(q|\mu, \sigma) = \int_{-\infty}^{\infty} p(q|Q, \mu, \sigma)p(Q|\mu, \sigma)dQ \]

where \( p(q|\mu, \sigma) \) is the conditional pdf of the daily-read flow given the true peak flow, and \( p(Q|\mu, \sigma) \) is the flood pdf. Note that the formulation of this likelihood is totally general; the relation used for \( p(q|Q) \) is arbitrary and the assumed flood distribution \( p(Q|\mu, \sigma) \) can also be freely chosen (with an associated change in the model parameters).

The full likelihood function, consisting of both true peak and daily-read flows, becomes

\[ p(Q, q|\mu, \sigma) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(q|Q, \mu, \sigma)p(Q|\mu, \sigma)dQ \]

3.5. Method 3: Random Dependent Censoring

The random dependent censoring (RDC) method involves a new approach for censored flows. It incorporates the multiple random thresholds present in the data, as in the binomial censoring (BC) method. Significantly, however, it accounts for the dependence between the threshold (daily-read flow) and true peak flow. The added complexities of this approach result in the likelihood function having to be evaluated using numerical integration.

The likelihood for a single daily-read flow can be expressed, using the total probability theorem, as

\[ p(q|\mu, \sigma) = \int_{-\infty}^{\infty} p(q|Q, \mu, \sigma)p(Q|\mu, \sigma)dQ \]

where \( p(q|Q) \) is the conditional pdf of the daily-read flow given the true peak flow, and \( p(Q|\mu, \sigma) \) is the flood pdf. Note that the formulation of this likelihood is totally general; the relation used for \( p(q|Q) \) is arbitrary and the assumed flood distribution \( p(Q|\mu, \sigma) \) can also be freely chosen (with an associated change in the model parameters).

3.6. Results and Discussion

In the following sections, the GO method is used as the reference to evaluate the performance of the other methods. The Monte Carlo experiment was performed for the following scenarios: typical lognormal flood distribution parameter \((\mu, \sigma^2)\) values of (8.0, 1.22) and typical censoring distribution \((E[r], CV[r])\) values of (0.2, 1.0) and (1.0, 1.0); where \( E[x] \) and \( CV[x] \) respectively represent the expected value and coefficient of variation of \( x \). Note that the choice of \( \mu \) is arbitrary and does not influence the results, while censoring distribution values are derived from Table 1.

3.6.1. Method 1: CAG

The CAG method was analyzed for typical flood and censoring distributions. The results of one of these analyses,
for a total record length \( n = g + c = 40 \) years, are shown in Figure 4. The solid symbols, interconnected by lines, present the quantile rmse and bias for the CAG method, while the open symbols present the results for the GO method for the stated number of true peak flows. For example, when \( g = 10 \), the CAG method uses both true peak and daily-read flows \( (n = g + c = 10 + 30) \), while the GO method only uses the true peak flows \( (n = g = 10) \). The GO method provides a benchmark to evaluate performance gains from utilizing daily-read flows. The results are shown for three recurrence intervals: 10, 50, and 100 years. Note that all analyses gave similar results to those shown.

The CAG method clearly outperforms the GO method. Indeed, and somewhat surprisingly, the rmse for the CAG method is insensitive to the number of daily-read flows. For \( g \leq 20 \), the CAG bias exceeds that of the GO method, but this bias is insufficient to make any noticeable impact on the rmse. This behavior arises because as \( g \) is decreased, the increase in bias is compensated by a decrease in sampling error. To understand this, consider the first-order variance approximation for the log quantile from the lognormal distribution [Stedinger et al., 1993, equation 18.4.4]

\[
\text{Var}(\log Q_T) = \frac{\sigma^2}{n} \left(1 + \frac{z_T^2}{2}\right) \quad (10)
\]

where \( \text{Var}(x) \) is the variance of \( x \), \( Q_T \) is the \( T \)-year quantile, and \( z_T \) is the standard normal deviate with exceedance probability \( 1/T \). Note that \( n \) and \( z_T \) are fixed; the variance of \( \log Q_T \) only depends on \( \sigma \). In the Monte Carlo experiments, there was minimal bias in \( \sigma \) over all values of \( g \).

\[ \text{Var} Q_T = \left[\frac{E(Q_T)}{Q_T}\right]^2 \text{Var}(\log Q_T) \]

\[ \frac{\text{Var} Q_T}{Q_T^2} = \left[\frac{E(Q_T)}{Q_T}\right]^2 \text{Var}(\log Q_T) \]

As \( g \) decreases, the bias in \( Q_T \) becomes more negative causing \( E(Q_T)/Q_T \) to fall well below one. Since \( \text{Var}(\log Q_T) \) remains virtually constant, the contribution of the sampling error \( \text{Var}(Q_T/Q_T) \) to the relative rmse decreases and, in this case, offsets the contribution of the bias to the relative rmse.

3.6.2. Method 2: BC

The BC method was analyzed for typical flood and censoring distributions. Figure 5 compares the BC and GO methods for a record length \( n = 40 \) years. It is clear the BC method performs very poorly, indeed disastrously, with the bias exploding as \( g \) decreases. The results shown in Figure 5 were typical of the other analyses, with the “errors” becoming more pronounced as the ratio \( (E(r)) \) was decreased. This result becomes intuitive if the reverse case is considered: the “errors” become smaller as the ratio is increased. The increased ratio implies that there is greater variability between the daily-read and true peak flows, resulting in a “flatter” distribution of \( Q \) and \( q \). In the limiting case, this distribution becomes uniform and the variables \( Q \) and \( q \) lose their interdependence; which

![Figure 4](image-url)  

Figure 4. Quantile rmse and bias for the CAG and GO methods for a record length of \( n = 40 \) years: \( \log Q \sim N(8.0, 1.2^2) \), \( E(r) = 1.0 \), \( CV(r) = 1.0 \).

![Figure 5](image-url)  

Figure 5. Quantile rmse and bias for the BC and GO methods for a record length of \( n = 40 \) years: \( \log Q \sim N(8.0, 1.2^2) \), \( E(r) = 1.0 \), \( CV(r) = 1.0 \).
describes the binomial case completely. This situation was alluded to earlier in Section 3.4.

3.6.3. Method 3: RDC

[34] The random censoring method was assessed using the typical scenarios, as well as additional cases where the parameters of the flood distribution \( \mu \) and \( \sigma \) and censoring distribution parameter \( E(r) \) were varied to extreme observed levels, or beyond. The results are shown for a record length of \( n = 40 \) years in Figure 6 and are representative of the other analyses performed.

[35] As for the CAG method, the rmse showed little change with \( g \). However, the RDC method exhibited less bias than the CAG, particularly for low \( g \) (compare Figure 4). This arises because \( p(q|Q) \) is compensating for the underestimation of the true peak \( Q \) by the daily-read flow \( q \).

[36] Figure 6 shows that the rmse remains quite constant for all values of \( g \), and remains under the GO case; while the bias is quite similar to the GO case, except for small values of \( g \) where it is significantly less. This effect is not intuitive because one would expect that the rmse would decrease by a more significant amount as \( g \) is increased; however, the explanation for this observed effect is that the variability present within the flood distribution (\( \sigma \)) overwhelms the data errors associated with the censored data.

[37] This effect is demonstrated in Figure 7 where an unusually “tight” flood distribution is considered, with \( \sigma \) lowered to an atypical level of 0.2 (versus the typical observed value of about 1.2). In this case, the rmse decreases as \( g \) increases, and largely follows the GO curve. This indicates that the censored data has little useful information in this case; the censoring error has overwhelmed the flood distribution’s variability, negating any
possible benefits. Note that despite the loss of information, the RDC method manages to control the bias.

[38] In comparison, Figure 8 shows that the censored data errors overwhelm the CAG method causing both the rmse and bias to significantly increase, resulting in much poorer performance than both the RDC and GO methods.

3.6.4. Misspecification Errors

[39] The effect of censoring distribution misspecification errors is examined in this section. This is achieved by generating the censored data using an assumed, true, \( E(r) \) value and then using another, erroneous, \( E(r) \) value for the likelihood calculations. Obviously, these misspecification errors only affect the RDC method. The RDC method was analyzed for the typical scenarios, as well as extra cases where the flood distribution parameter \( E(r) \) was increased to larger than typical values. The results of two RDC analyses are presented in Figures 9 and 10.

[40] When the erroneous \( E(r) \) value is less than the true value, shown in Figure 9, the RDC method exhibits little change in rmse, but an increasing bias as \( g \) decreases. Though the bias deteriorates, it is insufficient to have much impact on the rmse.

[41] When the erroneous \( E(r) \) value is greater than the true value, as in Figure 10, both the rmse and bias deteriorate for the RDC method as \( g \) decreases. Indeed, the rmse performance of the RDC method is little different from the GO method.

[42] Thus, if the true censoring distribution cannot be determined, it is preferable to use a censoring distribution with a small \( E(r) \). This is because, with small \( E(r) \), the RDC method behaves like the CAG method, which has robust rmse performance for typical flood data.

4. Conclusion

[43] It has been demonstrated that daily-read flows, flows derived from a single daily (say, 9 AM) observation, can underestimate the true peak flows by significant amounts, with factors of 10 times being possible. Several methods to deal with these daily-read, or censored, flows were evaluated using a Monte Carlo study.

[44] The BC method was found to perform very poorly because the dependence between the true peak and daily-read flows is contrary to the assumption of independence made in method’s derivation.

[45] Under typically encountered flood and censoring distributions, both the CAG and RDC methods were found to improve flood quantile estimates over the GO case. The RDC method is preferable over the CAG method because it is theoretically sound, provides a better performance when the censoring distribution is known, and is quite robust when the censoring distribution is misspecified. However, due to the additional computational issues of the RDC method, the CAG method may be used in these situations without a huge loss in performance over the RDC method.

[46] Under atypical flood and censoring distributions, the CAG method’s performance may degrade considerably below that of the GO method, while the RDC method, at worst, performs similarly to the GO method.

[47] The results presented here are of particular importance given the use of established databases in large-scale applications.
flood studies. In particular, this study demonstrates the clear need to rigorously evaluate the meaning of data, especially where changes in monitoring practices have occurred.

Acknowledgments. The assistance with data provided by Alex Springall is gratefully acknowledged. This research was funded by an Australian Research Council discovery grant. The authors wish to thank the three anonymous reviewers and the associate editor for their helpful comments.

References
Department of Land and Water Conservation (2000), PINNEENA 7 users guide, Sydney, N. S. W., Australia.