A Bootstrap Approach and Reliability of Time Varying Design

L. Logan, K. Ponnambalam and R. Zhou

A pollution control strategy encourages better management of wastewater discharges to the receiving streams. This is necessary to protect the aquatic species against high levels of toxic contaminants. The primary concern is to prevent violation of water quality objectives. Protection has been based on controls using frequency of extreme minimum flows and a constant design load. A statistical method to evaluate reliability of a stream under varying design loads is proposed. In it, we applied a Bootstrap compute-intensive technique to investigate the use of extreme-value probability distribution on seasonal flows. The Weibull Type-3 model is demonstrated to be suitable for design purposes. It was possible to link this approach to a system reliability concept, to assess the vulnerable nature of a stream.

20.1 Introduction

Streamflow quantity dilutes and plays a major role in assimilating wastewater. These phenomena are congruent with satisfying requirements of water quality objectives. Essentially, streamflow events are stochastic and non-stationary, and by tradition a design flow for waste load allocations relies on probability distribution of extreme minimum flows. Gumbel (1963) and Engelhardt and Bain...
(1975) describe different types of probability models. A typical asymptotic unbounded exponential type distribution is:

\[ F(x) = \exp(-\exp(-a(x - b))), \ a > 0 \]

or the Cauchy initial type:

\[ F(x) = \exp(-b/x^{a}), \ x \leq 0, \ a > 0, \ b < 0. \]

Likewise, there is a limited-initial and bounded exponential type:

\[ F(x) = \exp(-(x - c)/(b - c)^{a}), \ x \geq c, \ a > 0, \ b > c \]

that is used frequently. Other distributions are Gamma, Log Normal, and Log-Pearson. A series of technical documents on designs for waste load allocations and for making water quality-based decisions have been completed by US-EPA (1983, 1984, 1991).

In Ontario, a guideline for design load allocations relies on \(7Q_{20}\) instead of the \(7Q_{10}\) frequently used in the USA. Special attention is given to waste disposals under acute conditions that focus on mixing zones (US-EPA 1994). These approaches depend on the researcher’s ability to characterize the frequency of the flow events. This is done in conjunction with some established pollution criteria that are needed to compute the loading requirements. A stream that shows some resilience, or one that is not vulnerable, often demonstrates a site that might be adequate for waste load allocations. Other studies have looked at efficient ways to dispose treated and diluted toxic effluent to stream, (US-EPA, 1987) while investigators have covered methods to evaluate the performance of a system under defined set of operations; such methods are described by Fiering 1982, Hashimoto, et al., 1982, and Logan, 1992.

This chapter evaluates some aspects of varying design flows using reliability analysis.

### 20.2 Design

A pollution control strategy is acceptable if the characteristics of the stream are well understood and we are able to construct a design flow that is statistically sound. Several probability models are available for extremes. For example, Bobee and Ashkar (1991) examined Log-Pearson and Gamma Distributions. Also, Tasker (1987), and MOEE (1990) compared methods for estimating low flow characteristics, while Perzyna and Gottschalk (1994) derived frequency distributions of low flows. Similarly, Lawal and Watt (1996) analyzed frequency of
20.2 Design

low flows using the Akaike format and also examined the non-zero limit. Early works of Condie and Nix (1975) looked at probability distributions of low flows, while Biswas and Bell (1984) evaluated design flows.

Basically, the generalized extreme-value distributions of the Weibull Types are the most widely used. In one of his early works Gumbel (1963) adopted the Weibull Type-3 asymptotic distribution to evaluate low flow characteristics:

\[
F(x) = \exp\{-[(x - \gamma)/\beta - \gamma]\alpha\}, \quad x \geq \gamma, \; \alpha > 0, \beta > \gamma, \text{ and } \gamma \geq 0
\]  

(20.1)

where:

\[
F(x) = \text{the probability distribution},
\]

\[
x = \text{the sample value},
\]

\[
\alpha, \beta, \text{ and } \gamma = \text{location, scale, and shape factor parameter, respectively.}
\]

The density function \( f(x) \) is \( F'(x) \):

\[
f(x) = -{\alpha}/(\beta - \gamma) y^{(\alpha - 1)/\alpha} \exp(-y), \quad x \geq \gamma
\]

\[
= 0, \quad x < \gamma
\]  

(20.2)

where: \( y = [(x - \gamma)/\beta - \gamma]\alpha \).

Subsequently, another useful indicator is a reliability function (Antle and Bain, 1981):

\[
R(x) = \exp(-y), \; \alpha < \gamma < \infty
\]  

(20.3)

In addition, let us consider an observation examined at a threshold value \( x_1 \). In this case, for an event \( x \) occurring, with regards to a recurrence period \( \tau \), there either is a greater than or equal probability \( p = P[X \geq x_1] \) or less than probability \( q = P[X < x_1] = 1 - p \) possibilities.

For the case of the smallest values, the probability of \( \tau \), for the observation \( x \) is the product of the probability of \( \tau - 1 \) occurrences times the probability of the recovery event following. The expected value of \( \tau \) is as follows (Chow, et al., 1988):

\[
E[\tau] = \sum_{\tau=1}^{\infty} \tau (1-p)^{\tau-1} \cdot p = p[1 + 2(1-p) + 3(1-p)^2 + \ldots]
\]  

(20.4)

Equation 20.4, a power series of the form \((1+x)^n\), with \( x = -(1-p) \) and \( n = -2 \), gives the probability for the smallest values, \( P[X < x\tau] = 1/E[\tau] \). This states that the probability of occurrence of an extreme minimum event under such a condition equals the reciprocal of \( E[\tau] \). Hence, the design flow is equated to the return period \( (T) \):
A Bootstrap Approach and Reliability of Time Varying Design

\[ E[\tau] = T = 1/p \]  

(20.5)

We note that the characteristics of the instantaneous failure rate or the hazard function:

\[ h(x) = \frac{\gamma(\beta - \gamma)^{\alpha}}{(x - \gamma)^{\gamma + 1}}, \text{ for } x > \alpha \]  

(20.6)

The hazard function allows us to interpret the system's responses, for example, if \( \gamma < 1 \) then \( h(x) \) is decreasing, however, if \( \gamma > 1 \) then \( h(x) \) is an increasing function. In addition, if \( \gamma \) is known, then estimates of the remaining parameters exist and are unique. If all three parameters are unknown, then the likelihood function is unbounded, because \( \alpha \rightarrow x_{\text{min}} \). When the value of \( \gamma \geq 1 \), we have a likelihood function that is bounded. We conclude that some knowledge of the characteristic of \( h(x) \) is useful, seeing that it accounts for the vulnerable nature of the stream.

20.3 A Bootstrap Approach

The Bootstrap or Jackknife technique is a non-parametric computer-intensive approach demonstrated by Efron and Gong 1983, Efron and Tibshirani 1993, Allen 1997, and Buzas 1997. In this analysis, we applied the Bootstrap approach to produce a random replacement from the original data, thereby producing realization of a Bootstrap's sample distribution. Let \( (x_i) \) be a Bootstrap sample from the data set:

\[ B: \{X(t), t \geq 0\} \sim F_i \left( x_j \right), \text{ } i = 1, 2, \ldots, m, \text{ } j = 1, 2, \ldots, n^* \]  

(20.7)

Here, \( \{X(t), t \geq 0\} \) represents a stochastic series of observations analyzed at different windows on any interval \( 0 \) to \( \infty \). In this analysis we assume that the process is confined to separate intervals of successive points. Each process contains random variables that are positive and identically distributed (Parzen, 1962, Bunday, 1991). The intervals are defined:

\[ T_1 = t_1 - t_0, \text{ } T_2 = t_2 - t_1, \text{ } T_3 = t_3 - t_2, \ldots, \text{ } T_m = t_m - t_{m-1} \]  

(20.8)

where: \( T \) is an interval on \( 0 < t_0 < t_1 < t_2 < t_m \).

For our case we consider the Bootstrap variable \( x^*_{i,j} \) to be a streamflow event confined at \( T_{i,j} \), where \( i = 1, m \) denotes a step stage within a year, and \( j = 1, n^* \) denotes the sample replication. In other words, the interval contains random variables assumed to be stationary and ergodic. For example, if \( i = \) a month, then
we have a matrix of twelve sets of data each containing $n^*$ Bootstrap replications. The replacement sampling delivers random variables with an empirical distribution. We note $\omega^*$ as a Bootstrap estimator of the statistics $\omega(F)$:

$$\omega^* = \omega(x_{1}^*, x_{2}^*, \ldots)$$ (20.9)

with a defined mean $\omega^*$, and variance of the estimate $\sigma^2$:

$$\omega^* = 1/n^* \sum_j \omega_j$$ (20.10)

$$\sigma^2 = 1/(n^* - 1) \sum_j (\omega_j - \omega^*)^2$$ (20.11)

In this case, we apply the Weibull Type-3 process probability model as a predictor for the phenomena that are changing over time. From each Bootstrap series we determine estimates $\omega^*$ and $\sigma^2$ for $\alpha^*$, $\beta^*$, and $\gamma^*$. The purpose here is to see whether the Bootstrap replications provide the required statistics of normality for the populations (Snedecor and Cochran 1980).

The computational procedure is as follows:
1. For each interval, initialize the process with sample estimates $\alpha^*$, $\beta^*$, and $\gamma^*$ from the existing data set.
2. Generate Bootstrap replicates $(T_i, i=1, m)$, $j=1, n^*$, for a parameter, while keeping the other parameters constant, in turn.
3. The random-simulations for each $T_i$ produce $n^*$ replications for each parameter in question.
4. Determine Bootstrap estimators $\alpha^*$, $\beta^*$, and $\gamma^*$ and checks for normality.
5. Carry out frequency analysis for designs.

### 20.4 Performance Indicators

Investigations of Kingston and Patel (1981), Ascher and Feingold (1984) evaluated the reliability of the Weibull distribution, while Evans and Nigin (1980) examined the Bayesian predictions for the Weibull Type-2 model. In short, we expect the flow events to meet a specified threshold level. If the system is operating satisfactorily, then the reliability of the system is:

$$R(x) = P[x \geq x_T, x \in S]$$
$$= 1 - P[x < x_T, x \in F]$$ (20.12)

where $x \in S$, and $x \in F$ denotes the system is in the satisfactory or failure state, respectively.
Normally, the system recovers if a failure does occur. The probability of an observation violating the design flow is equivalent to the reciprocal of expected duration of failure:

$$\zeta = \frac{1}{E[x \in F]} = \frac{P[x_i \in F]}{P[x_i \in S | x_{i+1} \in F]}$$  \hspace{1cm} (20.13)$$

where $\zeta$ denotes the resiliency. In addition, there is a consequence of failure, which we can infer from the vulnerability $v$, which is frequently derived from the annual most severe consecutive failure events:

$$v = \sum_k p_k D_k$$  \hspace{1cm} (20.14)$$

where $p_k$ denotes the probability of failure of a severe event, and $D_k$ is the deficit at the time of failure.

### 20.5 Illustration

Bootstrap replications containing $n^* = 1000$ samples at each interval provide estimators $\alpha^*$, $\beta^*$, and $\gamma^*$. The Bootstrap produces acceptable values for these parameters as shown by three test exercises (Table 20.1). In addition, examples of parameter densities are shown in Figure 20.1. Acceptance is based on closeness to the Normal Distribution. It is evident that the hazard rate is decreasing as demonstrated by the value of the parameter $\gamma < 1$ in all cases. Furthermore, the Bootstrap results compare favorably with the annual minimum values. It is observed that the system became more vulnerable at a higher design load due to the loss of reliability, and by a reduction in resiliency (Figure 20.2b). There are some advantages, in that the designer has more information on the system's responses under different loading options. Examples of a step-stage moving average 7-day, 1:20 design flows are shown in Figure 20.2a, and the corresponding general performance evaluation indicators in Figure 20.2b.

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>Var($\alpha$)</th>
<th>$\beta$</th>
<th>Var($\beta$)</th>
<th>$\gamma$</th>
<th>Var($\gamma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>1.9</td>
<td>0.16</td>
<td>0.25</td>
<td>0.001</td>
<td>0.02</td>
<td>0.001</td>
</tr>
<tr>
<td>Test A</td>
<td>1.3</td>
<td>0.071</td>
<td>1.64</td>
<td>0.069</td>
<td>0.04</td>
<td>0.006</td>
</tr>
<tr>
<td>Test B</td>
<td>1.1</td>
<td>0.039</td>
<td>1.48</td>
<td>0.060</td>
<td>0.17</td>
<td>0.004</td>
</tr>
<tr>
<td>Test C</td>
<td>2.0</td>
<td>0.235</td>
<td>0.70</td>
<td>0.006</td>
<td>0.03</td>
<td>0.004</td>
</tr>
</tbody>
</table>
20.5 Illustration

Figure 20.1 Densities of Weibull Type-3 model parameters: Bootstrap samples (Test A)

Figure 20.2a Varying design flow - 7 day (1:20).

Figure 20.2b Performance indicators.
We conclude that this approach provides an opportune way to investigate stream behavior under design load changes. This is particularly interesting for further studies on the effects of inverse flow distributions on concentration. The disadvantage of this approach, on the other hand, is that time-varying design loads may be difficulty to implement. We anticipate that a real-time design flow operation would be an appropriate procedure to follow.

20.6 Conclusions

We investigated a stream’s behavior under varying design loads, by performing frequency analysis at different seasonal windows. Analysis was based on the assumption that each series is positive, independent and contains random variables. A Bootstrap technique is used in the generation of random samples by replacement, from which tests exercises characterized the Weibull Type-3 as a suitable process model. We were aware of the design implications under a varying control option, in that the behavior of the system would come into question under such changing design controls. Therefore, we used system reliability analysis to evaluate the system performances. These results demonstrate encouraging possibilities. It nonetheless, points out that further work is needed. We plan additional research on violation of water quality conditions under joint and conditional probabilities of flow events.

References

References