Sensitivity of Reservoir Design to the Generating Mechanism of Inflows

J. R. WALLIS
IBM Research Center, Yorktown Heights, New York 10598

N. C. MATALAS

Abstract. The design capacity of a reservoir depends on the generating mechanism of the inflows. If one uses the sequent peak algorithm in a deterministic manner, the minimum capacity $x$ required to meet a specified level of demand $\alpha$ is shown to depend on $\alpha$ and the values of the lag 1 serial correlation coefficient $\rho$ and the Hurst coefficient $h$ of the inflows. Although these experimental results may not realistically describe the ranges of the values $\alpha$, $\rho$, and $h$ over which other objective functions can be expected to be sensitive to the flow-generating model, they do indicate that such ranges exist and that one may have to consider carefully both the list of moments to preserve and the means for their unbiased preservation.

In the past, synthetic streamflow sequences have been used with simulation models of a water resource system to assess the response of the system to unknown future inflows. Such assessments are based on a set of system responses corresponding to a set of synthetic flow sequences, the lengths of which equal the economic time horizon of the system. The synthetic sequences are generated so that they bear some degree of resemblance to the corresponding historical sequences. For the most part, the resemblance has been limited to the low order moments: means, variances, skewnesses, and lag 1 serial correlations. For multi-season, multisite situations, the resemblance has included correlations in and between seasons and sites.

To generate synthetic sequences, a model must be chosen that will allow the specified degree of resemblance between the synthetic and historical sequences to be achieved. The mathematical structure of the model assures that the particular set of moments, i.e., the specified degree of resemblance, will be maintained statistically in the synthetic sequences. Customarily, algorithms are chosen that yield the values of the moments for infinitely long synthetic sequences that equal the estimated values of the corresponding moments for the finite historical sequences. The justification for this choice is somewhat obscure, since in expectation the moments are known to differ between the historical and synthetic sequences whose lengths are equal [Wallis and Matalas, 1971].

Although maintaining moments may be thought of as a necessary condition for model acceptability, it is not a sufficient condition. Maintaining moments is useful insofar as it provides a basis for mathematically structuring a model. But from an operational viewpoint, it is desired that, on the average, long synthetic sequences reflect more extreme events than short historical sequences. In this context, note that events may refer to the magnitude of individual flows or to the sum of flows over specified time spans.

The Markovian model has been used extensively to generate synthetic flow sequences. Although this model allows resemblance to be achieved in terms of the low order moments, it frequently fails to generate events more extreme than the historical ones. A possible explanation for this apparent failure is that the Markovian model, being a mathematical abstraction of a short memory process, cannot account for the long-term persistence that is evidenced in historical sequences by the tendency of the estimated values of the Hurst coefficient $h$ to be
larger than $\frac{1}{2}$. To account for long-term persistence, Mandelbrot and Wallis [1969] considered fractional noise processes, which unlike Markovian processes lie outside the Brownian domain of attraction, where $h$ may assume values other than $\frac{1}{2}$. On the basis of discrete parameter models approximating fractional noise processes, Matalas and Wallis [1971] set forth analytical procedures for generating synthetic sequences that maintain the historical estimates of $h$ as well as the historical estimates of the low order moments.

Askew et al. [1970] have noted that discrete parameter models approximating fractional noise processes tend to generate synthetic sequences with more extreme events than the sequences generated by Markovian models. This tendency lends some credibility to the explanation offered above for the apparent shortcoming of the Markovian model. Even so, to what extent one model is better than the other for operational purposes remains questionable.

To gain some insight into the relative merits of Markovian models and models approximating fractional noise processes, an investigation of the effect of persistence on the reservoir design capacity was considered. The sequent peak algorithm [Thomas and Burden, 196] was used to derive the minimum reservoir capacity required to meet various levels of demands for the flow sequences generated by these models.

STOCHASTIC FLOW MODELS

A sequence of flows $\{X(j) : j \in J\}$, where $J = \{1, \ldots, N\}$ denotes a set of equally spaced time points, represents one realization of a stochastic process the generating mechanism of which is unknown. Two models are considered for approximating this mechanism. The Markovian model has been used widely for a number of years to generate synthetic flow sequences. The fractional noise model was introduced to hydrology by Mandelbrot and Wallis [1968] and adapted to synthetic flow generation by Wallis and Matalas [1971] and Matalas and Wallis [1971].

The Markovian process is defined as

$$X(j) = \rho X(j - 1) + \epsilon(j) \quad (1)$$

where $\rho$ is the lag 1 serial correlation coefficient and $\epsilon(j)$ is a random component independent of $X(j - 1)$. The lag $u$ serial correlation coefficient $\rho_u$ is related to $\rho_u = \rho^{(u)}$ as follows:

$$\rho_u = \rho^{(u)} \quad (2)$$

A basic property of this process is that $h = \frac{1}{2} \lor \rho$.

Mandelbrot and Wallis [1968] proposed the following discrete parameter model which is referred to as type 2, for approximating fractional noise processes:

$$Y(j) = (h - \frac{1}{2}) \sum_{v=1}^{p-1} (j - v)^{3/2} \epsilon(v) \quad (3)$$

where $\epsilon(v)$ is a random component independent of $Y(j) \forall j > v$. To reduce the dominance of the low-frequency components that characterize this process, Matalas and Wallis [1971] proposed a filtered type 2 process defined as

$$X(j) = (h - \frac{1}{2}) \sum_{i=p}^{p-1} (pu - i)^{3/2} \epsilon(i) \quad (4)$$

where $p \geq 1$ is an integer and $X(j) = Y(pj)$ is the event generated by (3) at time $pj$. The mean, variance, and skewness for $X(j)$ and $Y(j)$ are the same, but

$$\rho_u(X) = \rho_{pu}(Y) \quad (5)$$

For the processes defined by (3) and (4), $h$ may assume values in the range $(0, 1)$ with the exception of $\frac{1}{2}$.

Without any loss of generality, assume that $E[X(j)] = 0$ and $E[X(j)]^2 = 1$. The short memory or short-term persistence of the Markovian process is mathematically described by

$$E[X(j) X(j - k) | X(j - 1), \ldots, X(j - k + 1)] = 0 \quad \forall k \geq 2 \quad (6)$$

Continuous parameter fractional noise processes are characterized by infinite memories. A discrete parameter approximation of the process requires the memory length $M$ to be finite but large, so that (7) holds for $k > M$. Note that the processes defined by (3) and (4) belong to the Brownian domain of attraction with
estimates of \( h \to \frac{1}{2} \) as \( N \gg M \). Operationally, the filtered fractional noises should use \( M \) greatly in excess of the design period \( N \).

The Markovian and fractional noise processes do not exhaust all possibilities for approximating the generating mechanism of streamflow sequences. Nonetheless, they represent two well-documented, tractable processes that are available for modeling the extremes of hydrologic persistence, the effect of which on reservoir design is described below.

**SEQUENT PEAK ALGORITHM**

Let \( \{ x(j) : j \in J \} \) and \( \{ D(j) : j \in J \} \) denote time sequences of flows and demands, where \( J = \{1, \ldots, N\} \) refers to the set of equally spaced time points over which a reservoir must operate. If

\[
\sum_{j=1}^{N} x(j) = \sum_{j=1}^{N} D(j) \quad (8)
\]

the flow is said to be fully developed; in such a case all demands spanned by \( J \) will be met. If there is partial development,

\[
\sum_{j=1}^{N} x(j) > \sum_{j=1}^{N} D(j) \quad (9)
\]

then there will be a sequence of wastes \( \{ W(j) : j \in J \} \), where

\[
\sum_{j=1}^{N} W(j) = \sum_{j=1}^{N} x(j) - \sum_{j=1}^{N} D(j) \quad (10)
\]

The total demand \( D = \sum_{j=1}^{N} D(j) \) is related to the total flow by

\[
D = \alpha \sum_{j=1}^{N} x(j) \quad (11)
\]

where \( 0 < \alpha \leq 1 \) denotes the level of development. The minimum capacity required to meet all demands is \( C_m \). If \( C(0) \) denotes the initial storage required to avoid storage deficiencies, the storage at the end of the \( j \)th time period \( C(j) \) is given by

\[
C(j) = \min_i \{ C_m [x(j) - D(j) + C(j - 1)] \} \quad (12)
\]

and the waste water at \( j \), \( W(j) \), is given by

\[
W(j) = \max_i \{ 0, [x(j) - D(j) - C_m + C(j - 1)] \} \quad (13)
\]

Let

\[
Z(i) = \sum_{j=1}^{i} [x(j) - D(j)] \quad (14)
\]

\( i \in I = \{1, \cdots, N\} \)

If \( \{ z(i) : i \in I \} \), the minimum and maximum values of \( Z(i) \) are \( Z' \) and \( Z'' \), respectively. The range of cumulative departures \( R \) is defined as

\[
R = |Z'' - Z'| \quad (15)
\]

for \( \alpha = 1 \), \( C_m = R \), and \( C(0) = Z' \).

The solution for \( C_m \) may be obtained by the sequent peak algorithm [Thomas and Burden, 1963], which is as follows. The two time series \( \{ X(j) : j \in J \} \) and \( \{ D(j) : j \in J \} \) are assumed to be cyclic. This strictly operational assumption is imposed to carry out the solution for \( C_m \). The solution requires the use of only two cycles; in such a case the time span for the two time series is defined as \( J' = \{1, \cdots, 2N\} \), where \( X(j) = X(j + N) \) and \( D(j) = D(j + N) \) for \( j = 1, \cdots, N \). The following steps lead to \( C_m \).

1. Calculate \( X(j) - D(j) \forall j \in J' \).
2. Calculate \( Z(i) \forall i \in I' = \{1, \cdots, 2N\} \).
3. In \( \{ Z(i) : i \in I \} \) locate the sequence of peaks \( \{ P_r : r \in \mathcal{R} \} \), where \( \mathcal{R} = \{1, \cdots, m\} \), such that \( P_1 < P_2 < \cdots < P_m \).
4. Between sequent peaks locate the sequence of troughs \( \{ T_s : s \in S \} \), where \( S = \{1, \cdots, m - 1\} \).
5. Form the sequence \( \{ (P_r - T_s) : s \in S \} \).

The minimum design capacity \( C_m \) is given by

\[
C_m = \max_r (P_r - T_s) \quad (16)
\]

The values of \( C(j) \) and \( W(j) \) are given by (12) and (13), \( C(0) \) being determined as follows. Set \( C(0) = 0 \), and then from (12) determine \( C(j) \forall j \in J \). If \( C(j) \geq 0 \forall j \in J \), then \( C(0) = 0 \). If \( C(q) < 0 \forall q \in Q \leq J \), then \( C(0) = \max_q |C(q)| \).

**EXPERIMENTAL DESIGN**

On the basis of a sequence \( \{ X(j) : j \in J \} \), where \( J = \{1, \cdots, N\} \) with \( N = 100 \), \( C_m \) was determined by using the sequent peak algorithm for various levels \( \alpha \) of development. A given sequence represented a realization of a normal stochastic process. Two sequence-generating mechanisms were considered. The first, the Markovian process, is described as
Flow Modeling

\[ X(j) = \mu(1 - \rho) + \rho X(j - 1) + (1 - \rho^2)^{1/2} \varepsilon(j) \]  

(17)

where \( X(j) \sim N(\mu, \sigma) \) and \( \varepsilon(j) \sim N(0, 1) \) \( \forall j \). The second, the filtered type 2 process, is defined as

\[ X(j) = \mu + \sigma \left[ \sum_{i=0}^{M-1} (M - i)^{2k-3} \right]^{-1/2} \sum_{i=0}^{M-1} (M - i)^{2k-3/2} \varepsilon(pj - M + i) \]  

(18)

where \( X(j) \sim N(\mu, \sigma) \) \( \forall j \) and \( \varepsilon(pj - M + i) \sim N(0, 1) \) \( \forall pj - M + i \) [Matalas and Wallis, 1971].

The value of \( C_m \) corresponding to the sequence \( \{X(j) : j \in J\} \) depends on \( \mu \) and \( \sigma \). The estimate of \( \mu \) given by

\[ \bar{X} = \left( \frac{1}{N} \right) \sum_{i=1}^{N} X(j) \]  

(19)

has the expectation

\[ E(\bar{X}) = \mu \]  

(20)

However, the estimate of \( \sigma \) given by

\[ S = \left( \frac{1}{N} \right) \sum_{i=1}^{N} [X(j) - \bar{X}]^2 \]  

(21)

has the expectation

\[ E(S) = \sigma^2(N, \rho) \]  

(22)

and is therefore biased.

To compare values of \( C_m \) in relation to the two models, sequences were generated with \( \mu = 10 \) and \( \sigma = E(S)/f(N, \rho) \), such that \( E(S) \) for Markovian sequences equals \( E(S) \) for filtered type 2 sequences. Two cases were considered.

\[ E(S) = 1 \] and \( E(S) = 3 \). Because the analytical form of \( f(N, \rho) \) is unknown, values of \( \sigma \) corresponding to specific values of \( E(S) \) could not be assigned. Thus the following sequence-generating procedure was used.

With \( \sigma = 1 \) and \( \mu = 10 \), 500 sequences of length \( N = 100 \) were generated for various values of \( \rho \) with each of the two models. For a given value of \( \rho \), \( C_m \) and \( S \) were determined for various levels of development \( \alpha \) for each of the 500 sequences, and from these values the means \( \bar{C}_m \) and \( \bar{S} \) were obtained. Only the values of \( C_m \) and \( \bar{C}_m \) depend on \( \alpha \). The value \( \bar{S} \) is the estimate of \( E(S) \), and the ratio \( \bar{S}/\sigma \) provides an estimate of \( f(N, \rho) \). The mean storage corresponding to sequences where \( E(S) = 1 \) was defined as

\[ \lambda_m = \bar{C}_m(\sigma/\bar{S}) \]  

(23)

Similarly, 500 sequences of length \( N = 100 \) with \( \sigma = 3 \) and \( \mu = 10 \) were generated for various values of \( \rho \) to obtain values of \( \lambda_m \) corresponding to \( E(S) = 3 \).

For the filtered type 2 process, the basic parameters are \( h \), \( p \), and \( M \), and therefore denoting \( f(N, \rho) \) as \( f(N, h, p, M) \) is perhaps more appropriate. For this process, the sequences were generated for various values of \( h \), \( p \), and \( M \), where \( M/p = 1000 \). The variable \( \rho \) is related to \( h \), \( p \), and \( M \) as defined by (5) and (6). The estimated values of \( f(N, \rho) \) and \( f(N, h, p, M) \) for the Markovian and filtered type 2 processes are given in the tabulation below and in Table 1. Note that two estimates of \( f(N, \rho) \) and \( f(N, h, p, M) \) were obtained, one related to \( \sigma = 1 \) and the other related to \( \sigma = 3 \). The entries in the tabulation below and in Table 1 are averages of the two values.

Wallis and Matalas [1971] have pointed out that estimates of \( \rho \) are biased. To take this bias into account, approximate values of \( E(\rho) \), denoted as \( \bar{E}(\rho) \), corresponding to \( \rho \) for \( N = 100 \) are given in the tabulation below and Table 2 for the Markovian and filtered type 2 processes, respectively. When \( \lambda_m \) for the Markovian and filtered fractional worlds are compared, the comparisons should be based on equal \( \bar{E}(\rho) \) rather than on the theoretic \( \rho \).

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( f(N, \rho) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.1</td>
<td>0.99</td>
</tr>
<tr>
<td>0.2</td>
<td>0.99</td>
</tr>
<tr>
<td>0.3</td>
<td>0.99</td>
</tr>
<tr>
<td>0.4</td>
<td>0.98</td>
</tr>
<tr>
<td>0.5</td>
<td>0.98</td>
</tr>
<tr>
<td>0.6</td>
<td>0.97</td>
</tr>
<tr>
<td>0.7</td>
<td>0.95</td>
</tr>
<tr>
<td>0.8</td>
<td>0.94</td>
</tr>
<tr>
<td>0.9</td>
<td>0.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( p )</th>
<th>( h = 0.6 )</th>
<th>( h = 0.7 )</th>
<th>( h = 0.8 )</th>
<th>( h = 0.9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.90</td>
<td>0.88</td>
<td>0.80</td>
<td>0.70</td>
</tr>
<tr>
<td>5</td>
<td>0.96</td>
<td>0.94</td>
<td>0.90</td>
<td>0.80</td>
</tr>
<tr>
<td>10</td>
<td>0.98</td>
<td>0.96</td>
<td>0.92</td>
<td>0.84</td>
</tr>
<tr>
<td>20</td>
<td>0.98</td>
<td>0.96</td>
<td>0.92</td>
<td>0.84</td>
</tr>
</tbody>
</table>
TABLE 2. Values of \( \rho \) and \( E(\rho) \) for the Filtered Type 2 Processes

<table>
<thead>
<tr>
<th>( h = 0.6 )</th>
<th>( h = 0.7 )</th>
<th>( h = 0.8 )</th>
<th>( h = 0.9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>( E(\rho) )</td>
<td>( \rho )</td>
<td>( E(\rho) )</td>
</tr>
<tr>
<td>1</td>
<td>0.66</td>
<td>0.57</td>
<td>0.73</td>
</tr>
<tr>
<td>5</td>
<td>0.35</td>
<td>0.29</td>
<td>0.44</td>
</tr>
<tr>
<td>10</td>
<td>0.24</td>
<td>0.19</td>
<td>0.33</td>
</tr>
<tr>
<td>20</td>
<td>0.16</td>
<td>0.13</td>
<td>0.24</td>
</tr>
</tbody>
</table>

An estimate of \( h \) is provided by

\[
K = \log \left( \frac{R/S}{\log (N/2)} \right)
\]  

(24)

where for small \( N \) and either process, \( E(K) \neq h \) [Matalas and Huzzen, 1967; Wallis and Matalas, 1970]. Approximate values of \( E(K) \), denoted as \( \bar{E}(K) \), for \( N = 100 \) are given in the tabulation below and in Table 3 for the Markovian and filtered type 2 processes, respectively.

An estimate of \( h \) is provided by

\[
K = \log \left( \frac{R/S}{\log (N/2)} \right)
\]  

(24)

where for small \( N \) and either process, \( E(K) \neq h \) [Matalas and Huzzen, 1967; Wallis and Matalas, 1970]. Approximate values of \( E(K) \), denoted as \( \bar{E}(K) \), for \( N = 100 \) are given in the tabulation below and in Table 3 for the Markovian and filtered type 2 processes, respectively.

\[
\begin{array}{c|c|c|c|c|c}
\rho & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \\
\hline
\bar{E}(\rho) & 0.09 & 0.18 & 0.28 & 0.38 & 0.47 \\
\end{array}
\]

Values of \( \bar{E}(K) \) can form another basis for the objective comparison of \( \lambda_m \) for different worlds of synthetic hydrology.

**Experimental Results**

The storage \( \lambda_m \) is a function of \( \alpha \) and the parameters that define the sequence-generating

![Graph](image)

Fig. 1. The minimum storage versus the expected value of \( \rho \), first order serial correlation for sequences of length \( N = 100 \). The level of development \( \alpha = 0.80 \), \( \mu = 10 \), and \( E(S) = 3 \). The solid line represents the Markovian process, where \( h = \frac{1}{4} \). The dashed line represents the filtered fractional noise process, where \( h > \frac{1}{2} \).
Apart from the mean and standard deviation, \( \rho \) is the basic parameter of the Markovian process, and \( h, p, \) and \( M \) are the basic parameters of the filtered type 2 process. For the Markovian process, \( h = \frac{1}{2} \vee \rho \), whereas for the filtered type 2 process, \( h \) may assume values in the range \((0, 1)\) with the exception of \( \frac{1}{2} \).

From (5) and (6), \( \rho \) is seen to be a function of \( h, p, \) and \( M \) for the filtered type 2 process. To facilitate discussions, \( \lambda_m \) is denoted as \( \lambda_m' \) and \( \lambda_m'' \) for the Markovian and filtered type 2 processes, respectively.

For \( E(S) = 3 \), \( \lambda_m \) as a function of \( \bar{E}(\rho) \) is shown in Figures 1, 2, and 3 for \( \alpha = 0.80, 0.90, \) and 0.99, respectively. For \( E(S) = 1 \), \( \lambda_m \) as a function of \( \bar{E}(\rho) \) is shown in Figure 4 for \( \alpha = 0.99 \).

There is a one-to-one relation between \( \lambda_m' \) and \( \bar{E}(\rho) \) for the Markovian process. However, for the filtered type 2 process, there is a family of functions dependent on the parameter \( p \). From these figures, note that \( \lambda_m' \approx \lambda_m'' \vee \rho \), where \( \alpha \leq 0.80 \). However, as \( \alpha \to 1 \), \( |\lambda_m' - \lambda_m''| \) increases rapidly with \( \rho \), where \( \lambda_m' < \lambda_m'' \).

For any value of \( \alpha \), \( \lambda_m'' \) is not sensitive to values of \( h \) for \( \rho \leq 0.3 \), although \( |\lambda_m' - \lambda_m''| \) may be substantial. As is to be expected, the insensitivity is more acute for \( E(S) = 1 \) than for \( E(S) = 3 \). For very small values of \( E(S) \) relative to \( \mu \), the target values are quite likely to be met by any generating process.

In Figures 5, 6, and 7, \( \lambda_m \) as a function of \( \bar{E}(K) \) is shown for \( \alpha = 0.80, 0.90, \) and 0.99, respectively, where \( E(S) = 3 \). For both processes, there is a one-to-one relation between \( \lambda_m \) and \( \bar{E}(K) \). In general, \( |\lambda_m' - \lambda_m''| \) increases as \( \bar{E}(K) \) increases. However, for \( \alpha \geq 0.90, \lambda_m' \approx \lambda_m'' \) for \( \bar{E}(K) \leq 0.75 \), and for \( \alpha = 0.99, \lambda_m' = \lambda_m'' \vee \bar{E}(K) \).

**SUMMARY AND CONCLUSIONS**

The preceding results indicate that over certain ranges of values of \( \alpha, \rho, \) and \( h \), \( \lambda_m' \approx \lambda_m'' \), such that \( \lambda_m \) is insensitive to the generating
Fig. 4. The minimum storage versus the expected value of $p$, first order serial correlation for sequences of length $N = 100$. The level of development $\alpha = 0.99$, $\mu = 10$, and $E(S) = 1$. The solid line represents the Markovian process, where $h = \frac{1}{2}$. The dashed line represents the filtered fractional noise process, where $h > \frac{1}{2}$.

For $\alpha < 0.80$, $\lambda_m$ depends primarily on $p$, and the Markovian model may be operationally useful even if the real world is more accurately described by a filtered type 2 process. On the other hand, for $\alpha \geq 0.80$, $\lambda_m$ depends primarily on $h$. For such values of $\alpha$, the Markovian model can be used as follows. From an observed flow sequence, estimates of $p$ and $h$ are obtained. A value of $p$ is selected so that $E(K) = K$, where $K$ is the estimate of $h$. With this value of $p$, synthetic sequences are generated with the Markovian model. Thus the Markovian model yields, on the average, $\lambda_m' \approx \lambda_m''$.

Wallis and Matalas [1971] have pointed out that the synthetic sequences generated with a Markovian model such that $E(K) = K$ display correlograms distinctly different from those for historical sequences. Whether this distortion of the temporal structure of the flows would prove detrimental if a more realistic objective function were used remains to be determined.

Note that the sequent peak algorithm was applied in a deterministic manner by using the observed sample means. In practice, a target draft might be used; in such a case account would need to be taken of the probabilities of failing to meet the target and of the associated losses. These
Although these experimental results may not realistically describe the ranges of values of \( \alpha \), \( \rho \), and \( h \) over which other objective functions can be expected to be sensitive to the flow-generating model, they do indicate that such ranges exist and that one may have to consider rather carefully both the list of moments to be preserved and the means for their unbiased preservation.

**REFERENCES**


(Manuscript received January 12, 1972.)