

Article

Reservoir Capacity Estimation by the Gould Probability Matrix, Drought Magnitude, and Behavior Analysis Methods: A Comparative Study Using Canadian Rivers

Tribeni C. Sharma and Umed S. Panu * 

Department of Civil Engineering, Lakehead University, Thunder Bay, ON P7B 5E1, Canada

* Correspondence: uspanu@lakeheadu.ca

Abstract: Among the various methods for estimating reservoir volumes, the Gould probability matrix (GPM) method has been touted as a powerful method for estimating reservoir volumes. The other methods in vogue are the Behavior analysis (BA) with the latest induction of the Drought magnitude (DM) method. A comparison of the above methods in terms of ease, efficiency, and relative merits from each other is currently lacking in the literature. This paper compares the above three methods with a detailed analysis of the GPM method using the monthly flows from 16 Canadian rivers at the draft ratios of 75 and 50% with the probability of failure of 2.5, 5 and 10%. The results reported in this paper indicate that fifteen zones are sufficient in the GPM method to yield the reservoir capacity for the Canadian rivers while requiring no standardization of the data, similar to the BA method. In the DM method, standardized monthly flow sequences in combination with a scaling parameter Φ yielded effective drought length, which, when multiplied by drought intensity and the average of 12 monthly standard deviations, resulted in the appropriate values of reservoir capacity. The results of this paper affirm that the GPM method offers little special merit in obtaining reservoir capacity in view of the rigor of computational efforts and uncertainty in the correction factors for significantly autocorrelated (dependent) annual flows. The DM method was found to be comparable to the BA method, though it requires standardization of the monthly flow data. The study suggests that all three methods result in comparable estimates of reservoir capacity for nearly independent annual flows with a slight edge to the Behavior analysis (BA) method.



Citation: Sharma, T.C.; Panu, U.S. Reservoir Capacity Estimation by the Gould Probability Matrix, Drought Magnitude, and Behavior Analysis Methods: A Comparative Study Using Canadian Rivers. *Hydrology* **2023**, *10*, 53. <https://doi.org/10.3390/hydrology10020053>

Academic Editor: Pingping Luo

Received: 13 January 2023

Revised: 8 February 2023

Accepted: 15 February 2023

Published: 20 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: behavior analysis; drought magnitude; draft ratio; standardized hydrological index; transitional probability matrix; probability of failure

1. Introduction

The reservoir capacity or volume (C_R) can be estimated using river flows for a given draft ratio (α) and the probability of failure (PF) for supplying the water. The draft ratios are expressed as the ratio to the mean annual flow (MAF), such as 75% ($0.75 \mu_a$), 50% ($0.50 \mu_a$), etc., where μ_a is the mean of the annual flow sequences (MAF) under consideration. The flows from a river can be analyzed using annual, monthly, or weekly scales for assessing the reservoir capacity. The mean flow would turn out to be nearly the same at all scales, though the variance and autocorrelation structure would differ significantly from each other at respective scales. A majority of reservoirs across the globe are designed for α ranging from 40 to 90% [1] with an adopted α of 75% [2–4]. At times, reservoirs with α in the vicinity of 50% also are in existence [1,5]. The monthly scale of flows is deemed adequate in sizing the reservoirs [2–4]. The other important parameter associated with sizing a reservoir is the probability of failure (PF) of the reservoir to supply water to meet the demand. For pragmatic reasons, the PF values of 2.5, 5, and 10% are regarded as satisfactory [3], though theoretically, PF as low as 0% can be used to calculate the reservoir capacity. A large reservoir is not easy to construct because it is likely to face insurmountable

financial, social, and environmental challenges; especially when a big dam is involved. The PF of a reservoir is defined as the ratio of months that a reservoir failed to meet its demand to the total number of months of the data used in the analysis [3,4].

There are several methods used for estimating the capacity of a reservoir, such as the Behavior analysis (BA) [1,3], the Drought magnitude method [5,6], the Sequent peak algorithm (SPA) [7–12], and the Gould probability matrix (GPM) method [13–23]. The authors explored the SPA earlier in the context of the Canadian river flow data [5,6]. In this paper, three methods, viz. the GPM [13–23], the DM [5,6], and the BA [1,3], are analyzed, compared, and discussed.

2. Preliminaries on the Gould Probability Matrix (GPM) Method

Three methods, viz. the GPM, DM, and the BA, are the focus of analysis in this paper. The details of the DM method are well documented by the authors [5,6] and are only succinctly described in Section 3. Likewise, the details on the BA method are well described in [1,3] and thus are only briefly described in Section 3. The authors have used the GPM method for the first time; hence, its salient features are described in detail in the following, though the full computational algorithm is well documented in [3,4,11].

The analysis and/or estimation by the GPM method begins with the application of the following water balance Equation (3).

$$Z_{t+1} = Z_t + X_t - D_t - \Delta E_t - L_t \quad (1)$$

where Z_{t+1} and Z_t are the storage contents (m^3) at the end and beginning of the month, X_t (m^3) is the river inflow during the t^{th} month, D_t (also termed as a draft) is the release (m^3) from the reservoir during the t^{th} month, ΔE_t is the net evaporation loss during the t^{th} month, and L_t is the leakage loss through seepage, minor abstractions, etc. In the present analysis, since three methods are being compared, therefore the entity ΔE_t and L_t shall remain nearly the same for all the methods, therefore, can be disregarded for further consideration. The water balance Equation (1) thus becomes as follows.

$$Z_{t+1} = Z_t + X_t - D_t \quad (2)$$

The draft, D_t , is expressed as ' α MAF', where α is the draft ratio such as 0.75 or 0.50 and MAF (i.e., mean annual flow, μ_a (m^3/year)). Since Equation (2) is being implemented on a monthly basis, and thus μ_a is represented by μ_0 , i.e., m^3/month . The GPM method, with its theoretical base originating from the work of Moran [13,14], has been advanced by Gould (1961) [15,16]. Gould's procedure indulging transitional probability matrix algorithm has successfully been applied worldwide. The work initiated by [15] was further developed by McMahon (1976) [17], Theo and McMahon (1982) [18]. In recent years, it has found applications in African catchments, where the data are patchy and scant. For example, Parks et al. (1989) [19] applied the GPM method to catchments in Botswana, and Otieno and Ndiritu (1997) [20] applied it to the rivers of Kenya. Ragab et al. (2001) [21] applied it to the hilly catchments of Tunisia for the design of reservoirs. Ibn-Abubaker (2008) [22] applied it to assess the performance of an existing reservoir on the Tiga-dam in Kano state, Nigeria and suggested more judicious use of the reservoir by releasing extra water to the downstream users. Recently, Kraus et al. (2022) [23] tested the performance of five reservoir sizing methods, including the GPM, to existing reservoirs in northwestern Ontario, Canada. Further details on this method are elaborated in a recent book by Nagi et al. (2002) [11].

The GPM method essentially treats time and water contents as discrete variables. Thus, the trial reservoir capacity C_R is divided into k zones, with k being a function of the coefficient of variation (cv_a) of annual flows. The volume in each zone, excluding the top and bottom zones, is obtained as follows.

$$\text{Volume (w)} = C_R / (k - 2) \quad (3)$$

The volume in the top (k^{th}) and the bottom zone is taken as zero. These zones would be numbered as 0, 1, 2, 3, ..., $k-1$ (total being k), with the corresponding cumulative volume ($v = \text{sum of } w\text{'s}$) in the relevant zone. The water balance Equation (2) is applied month-by-month by taking each year of data (N years). Starting at month 1 (the beginning of the year) and then progressively proceeding to month 12, which outputs the zone in which the year ends. The tally sheets are prepared to indicate the numbers 0, 1, 2, and $k-1$ according to the storage (Z_{t+1}) obtained through the water balance equation for each year. Such operations are conducted for each possible starting zone, and the element corresponding to the starting and the ending zone in the tally sheet is denoted as Z_t and Z_{t+1} . The $[k \times k]$ matrix is thus prepared to comprise numerals representing the volume which in turn are converted to probabilities upon dividing by N . This matrix is powered multiple times to obtain the steady-state matrix.

Any failures (reservoir emptying in months) in a particular zone during the year are also noted. The total number of failures (i.e., for all years under consideration) in each zone is converted to probabilities of failure (pf) divided by $N \times 12$. The products of 'pf' in each zone to the corresponding element in the steady-state matrix are summed to yield the total failure probability, PF. When the PF is preassigned such as 5% ($=0.05$), several trials of C_R may be required such that the calculated PF equals the preassigned PF. Further details are provided in Section 3, along with a worked example in Appendix A.

The GPM method does not require the standardization of the monthly streamflow data. Essentially, it is designed for the flow sequences in which the annual flows are independent or random. When the annual flow data show significant autocorrelation, the estimates of C_R are corrected using a correction factor, which is a function of lag-1 autocorrelation (ρ_a) (McMahon, 1976) [17] and (Srikanthan and McMahon (1985) [24].

3. Data and Computational Algorithms of Reservoir Volumes

Sixteen rivers from western to Atlantic Canada (Figure 1, Table 1) were involved in the analysis. The monthly and annual flow data for these 16 rivers were extracted from the Canadian hydrological database (Environment Canada, 2020) [25]. The rivers encompassed drainage areas ranging from 187 to 25,900 km², with the data bank spanning from 35 to 106 years. Other pertinent statistics of interest are: the coefficient of variation of annual flows (cv_a) ranged from 0.13 to 0.72, and lag-1 autocorrelation (ρ_a) ranged from 0.03 to 0.62. The higher values of cv_a and ρ_a are associated with rivers in the Canadian Prairies. The rivers in northern Ontario and eastern Canada tend to display a negligible level of persistence (insignificant values of ρ_a) in the annual flow sequences.

The values of statistical parameters, mean (μ), standard deviation (σ), coefficient of variation (cv), and lag-1 autocorrelation (ρ) for these rivers at annual and monthly scales were computed and are summarized in Table 1, where subscript 'a' stands for annual and 'o' for the monthly sequences. Since $cv (= \sigma/\mu)$, therefore instead of σ , the values of cv are stated (Table 1) for brevity and ease of comprehension.

Table 1. Summary of statistics of annual and monthly flows of selected rivers across Canada.

Name, Location, and the Numeric Identifier of the River in Figure 1	Data Size (Year)	Area (km ²)	μ_0 (m ³ /s)	cv_a	cv_o	cv_m	cv_{av}	ρ_a	ρ_{m1}	ρ_{m2}
1	2	3	4	5	6	7	8	9	10	
[1] Bow River at Banff (51°10'30" N, 115°34'10" W)	110 (1911-20)	2210	39.01	0.13	1.05	0.41	0.24	0.06	0.50	0.76
[2] Beaver River at Cold Lake Reserve (54°21'18" N, 110°13'2" W)	65 (1956-20)	4505	18.81	0.72	1.49	1.24	0.98	0.36	0.77	0.90
[3] Churchill River above Otter Bridge (55°38'47" N, 104°44'5" W)	57 (1964-20)	119,000	296.29	0.37	0.50	0.48	0.43	0.59	0.94	0.97
[4] Sturgeon River Weir (54°26'20" N, 103°10'30" W)	35 (1961-95)	14,600	46.91	0.49	0.53	0.66	0.49	0.62	0.91	0.98
[5] Island Lake River near Island Lake (54°03'34" N, 94°39'34" W)	46 (1948-93)	25,900	86.18	0.28	0.54	0.45	0.38	0.27	0.87	0.92
[6] Gods River below Allen Rapids (55°01'35" N, 93°50'10" W)	61 (1934-94)	14,000	154.58	0.28	0.42	0.44	0.38	0.36	0.94	0.96
[7] English River at Umfreville (49°52'30" N, 91°27'30" W)	99 (1922-20)	6230	58.54	0.32	0.74	0.85	0.51	0.20	0.76	0.88
[8] Neebing River at Thunder Bay (48°23'00" N, 89°18'23" W)	66 (1954-19)	187	1.62	0.37	1.48	2.01	0.81	0.20	0.43	0.73
[9] Pic River near Marathon (48°46'26" N, 86°17'49" W)	50 (1971-20)	4270	50.21	0.24	1.03	1.08	0.56	0.13	0.41	0.71
[10] Pagwachau River at Highway11 (49°46'00" N, 85°14'00" W)	53 (1968-20)	2020	23.01	0.25	1.18	1.63	0.62	0.06	0.36	0.69
[11] Goulis River near Searchmont (46°51'37" N, 83°38'18" W)	53 (1968-20)	1160	18.33	0.21	1.05	1.04	0.58	0.08	0.33	0.66
[12] Becancour A Lyster (46°22'08" N, 71°37'21" W)	53 (1923-68)	1410	30.59	0.20	1.08	1.06	0.62	0.03	0.26	0.65
[13] Beaurivage Sainte Entiene (46°39'33" N, 71°17'19" W)	75 (1926-00)	709	14.19	0.26	1.19	1.38	0.62	0.19	0.24	0.64
[14] Lepraue River at Lepraue (45°10'11" N, 66°28'05" W)	101 (1919-19)	239	7.41	0.22	0.81	1.02	0.59	0.10	0.23	0.62
[15] Upper Humber River (49°14'34" N, 57°21'36" W)	68 (1953-20)	2210	80.21	0.13	0.87	0.75	0.44	0.18	0.13	0.56
[16] Torrent River at Bristol pool (50°36'26" N, 57°09'05" W)	61 (1960-20)	624	24.79	0.15	0.88	0.75	0.44	0.18	0.16	0.59

NOTE: The overall coefficient of variation ($cv_o = \sigma_o / \mu_o$) is based on non-standardized monthly flows, cv_{av} ($=\sigma_{av} / \mu_o$) signifies the average monthly coefficient of variation, and cv_m ($=\sigma_{max} / \mu_o$) is the maximum value coefficient of variation, in which σ_{av} is the average of 12 monthly values, and σ_{max} stands for the maximum value of standard deviation among 12 monthly standard deviations. The notation cv_a means annual coefficient of variation, with μ_a referring to the mean annual flow (MAF). The ρ_{m1} and ρ_{m2} , respectively, stand for lag-1 autocorrelation in the MA1 and MA2 sequences that are constructed from the monthly SHI sequences in the context of the DM method.



Figure 1. Location of river gauging stations across Canada (source Environment Canada).

3.1. Computational Procedure for the GPM Method

Following the detailed description of the procedure outlined in [3], the procedure adopted for the present study is summarized below. Sturgeon River ($\rho_a = 0.62$) is used as an illustration with tables and calculations shown in Appendix A.

Step 1: For each river, the monthly flow data (X_t) in the units of million ($10^6 \text{ m}^3/\text{month}$) were collated year-wise, starting from the beginning of a record until 2020 or until the data were available. In the case of the Sturgeon River, data were available until 1995. Based on the monthly flow data, μ_o (in $10^6 \text{ m}^3/\text{month}$) was computed. The draft D_t in Equation (2) was computed as 0.75 (or 0.50) of μ_o . The annual flow data were also used to compute the annual coefficient of variation (cv_a) and lag-1 autocorrelation, ρ_a (Table 1).

Step 2: The initial storage capacity (C_R) in the units of (10^6 m^3) was assumed, which was divided into 15 zones (13 zones of equal volume, first and last zones with volume as 0). The initial guess of C_R was obtained (derived) by using the Behavior Analysis (i.e., through the month-by-month routing through Equation (2) in the lumped fashion), which is simpler in terms of the rigour of calculations. In the present case, it was assumed to be $1500 \times 10^6 \text{ m}^3$ as the Behaviour analysis yielded a value of $1550 \times 10^6 \text{ m}^3$ at the draft ratio of 75%.

Step 3: The water balance Equation (2) was applied on a monthly basis for each year, starting with each of the 15 zones (zone 0 to zone 14) for the period of record. In the process of routing, the initial value of storage was taken as the mid-point between the minimum and the maximum volume (Table A1 in Appendix A) applicable to that zone. The routing was carried out in terms of the volume units and the output (storage volume) converted to zone number, which has the upper and lower limits as earmarked in Table 1. For each of the starting zones, the number of failures (months, showing zone number 0) was counted. A total of 12N months were routed with initial zones from 0 to 14, and the values of failures (nf) corresponding to each starting zone were noted. These numbers nf were converted to a probability of failure (pf) by dividing each entry by 12N. Table A2 (Appendix A) shows the values of nf and pf corresponding to each starting zone.

Step 4: For each year, the routing was carried out by taking the initial zone as 0, 1, 2, ..., 14 in the month of January, while the resulting zone occurred during the month of December (obtained from routing such as 0, 1, 2, ..., 14) was recorded (or noted). The number of occasions (frequency), which resulted as pairs 0-0 (the January storage occurred in zone 0 while the December storage also occurred in zone 0 through routing), 0-1, 0-2, ..., 0-14, were noted in column 2 (Table A3). Likewise, the number of other occurrences was

noted in column 3 (Table A3) for pairs 1-0 (the January storage in zone 1, and the December storage in zone 0), 1-1, 1-2, ..., 1-14 were obtained through routing. Such operations continued until the number of occasions 14-0 (the January storage is in zone 14 and the December storage is in zone 0), 14-1, ..., 14-14 were recorded in the last column (Table A3). Table A3 contains a number of such occasions and can be designated as the transitional matrix of frequency of occurrences. The numbers in each column should add to the total number of years (in the present case, 35). The numbers in each cell of Table A3 are divided by N, which results in the probabilities as shown in Table A4 (Appendix A). The sum of the numbers in each column should add up to 1. In the present case, since the rounding is carried out to three decimal places, the sum may not add exactly to 1 but can be rounded to 1. Thus, a 15×15 matrix of probabilities, termed a transitional probability matrix (TPM), was obtained.

Step 5: A steady-state probability matrix was obtained by powering the transitional matrix 6–8 times. This can be ascertained by a review of values (up to three decimal places) in the resulting matrix at each stage of powering. That is, in each row, the values of probability would become the same. For Canadian rivers, powering a transitional probability matrix six times was found to be adequate in obtaining a steady-state probability matrix, as is shown in Table A5 (Appendix A).

Step 6: The steady-state probability of the reservoir in a particular zone was multiplied by the value of the probability of failures ‘pf’ for that zone (Table A2). The resulting sum from all zones is the overall PF of the reservoir (Table A6). Such a computed PF is expected to correspond to the desired PF of 2.5, 5 or 10%. Should the computed value of PF not correspond to the desired value of PF, then the assumed value of C_R should be modified until the desired value of PF is met. On average, such an operation required 2–5 iterations. For this purpose, generally, three iterations were found to be sufficient, which could easily and efficiently be accomplished by modern-day desk/laptop computers.

For rivers displaying significant annual lag-1 autocorrelation, a correction factor (CF), as proposed by Srikanthan and McMahon (1985) [24], was applied to yield the correct value of reservoir capacity, C_R . The CF values due to [24] were interpolated from graphs reported in the book [3]. Since the values of CF are based on interpolation, the values of C_R should be regarded as approximate.

An illustrative example of the Sturgeon River was initiated with $C_R = 1500 \times 10^6 \text{ m}^3$, which resulted in a value of PF equal to 0.0028, which is too low. The next value assumed of C_R was $1000 \times 10^6 \text{ m}^3$, which also yielded a low value of PF equal to 0.01. A revised value of C_R equal to $500 \times 10^6 \text{ m}^3$ yielded a value of PF equal to 0.049, which is very close to 0.05. A further refinement of C_R equal to $496 \times 10^6 \text{ m}^3$ yielded a value of PF exactly equal to 0.05 (or 5%). All tables in Appendix A, thus, refer to the Sturgeon River for the case of C_R equal to $496 \times 10^6 \text{ m}^3$. Since annual lag-1 autocorrelation ($\rho_a = 0.62$) for the Sturgeon River is quite high, the above C_R was corrected by multiplying a correction factor (CF = 2.01), which was obtained through interpolations from [3]. The corrected value of C_R thus became $997 \times 10^6 \text{ m}^3$ ($= 2.01 \times 496 \times 10^6 \text{ m}^3$). This value is at variance from the estimates obtained from the BA method ($1550 \times 10^6 \text{ m}^3$) and the DM method ($1530 \times 10^6 \text{ m}^3$). Since estimates by the BA and the DM methods are nearly equal, the corrected value by the GPM method still needs revision so as to match the values by the BA and the DM methods. This can be undertaken by improving the value of the correction factor (CF), and a procedure is discussed in Section 4 under Results and Discussion.

All such necessary computations reported in this study were accomplished by writing a Macro in Visual Basic by coupling the input data with the MS Excel framework. The Macro had the capability to accommodate the number of zones (k) from 15 to 20. When 20 zones were considered, the computational efforts also proportionately increased as the matrix size became 20×20 , but such efforts were competently accomplished by modern-day high-capacity desktop/laptop computers.

3.2. Computational Procedure for the DM Method

An important requirement for the DM method is the month-by-month standardization of the flow sequences, which renders the data to be stationary. In the case of monthly flow sequences, there are 12 values of means (i.e., $\mu_1, \mu_2, \dots, \mu_{12}$) and standard deviations (i.e., $\sigma_1, \sigma_2, \dots, \sigma_{12}$) that were used in respective standardization resulting in the standardized hydrological index (SHI) sequence, which is also termed as the moving average order 1 (MA1) sequence. For brevity, these values of monthly statistics are not shown in Table 1. The autocorrelation at lag-1 for monthly SHI sequences is designated as ρ_{m1} . When two consecutive values in an SHI sequence are averaged out in a moving average fashion, the resultant sequence is termed a moving average order 2 (MA2) sequence. The autocorrelation from MA2 sequences is denoted as ρ_{m2} (Table 1). The algorithms for computing the aforesaid statistics were adopted as documented in [3,4].

The other step in the analysis was to compute the overall mean and standard deviation viz., μ_o and σ_o (with $cv_o = \sigma_o/\mu_o$) of the non-standardized monthly flow sequences. That is, the monthly flows are taken in a sequential format to compute the above statistics [5,6]. The threshold values of 'SHI_o' were computed at cutoff levels ($0.75 \mu_o$ and $0.50 \mu_o$) using the overall $\mu (= \mu_o)$ and $\sigma (= \sigma_o)$ of the non-standardized monthly flow sequences. The threshold value of SHI_o at $0.75 \mu_o$ is equal to $((0.75 \mu_o - \mu_o)/\sigma_o = -0.25/cv_o)$ and likewise at $0.50 \mu_o$ is equal to $(-0.50/cv_o)$. The other cutoffs required at $0.75 \mu_o$ are SHI_{av} ($= -0.25 \mu_o/\sigma_{av}$) and SHI_m ($= -0.25 \mu_o/\sigma_{max}$), in which σ_{av} is the average of the 12 monthly values of the standard deviations and σ_{max} is the maximum value among the 12 values. These lines of SHI_o, SHI_{av} and SHI_m are used for truncating the SHI sequences (MA1 or MA2 derivatives) for estimating the C_R in the DM method.

A detailed description of the DM method for estimating the reservoir capacity is well documented in [5,6]. Essentially, the method works on the hypothesis that drought magnitude ($= \text{drought intensity} \times \text{drought length}$), as advanced by Dracup et al. (1980) [26] and implemented by Sharma and Panu (2013, 2014) [27,28]. In summary, the SHI sequences in this method are truncated by an SHI_o, SHI_{av}, or SHI_m line. The drought length (L_T , in which T is the return period of the drought event $\approx$ sample size N) is evaluated using a Markov chain order 1 (MC1) or Markov chain order 0 (MC0) model. The drought length (L_T) thus obtained when multiplied by drought intensity (I_d) yields the drought magnitude (M_T). A crucial parameter in this transformation is Φ , which transforms L_T into effective drought length, $L_c [= \Phi L_m + (1 - \Phi) L_T]$, where L_m is the mean drought length. The value of L_m , respectively, for MC0 and MC1 considerations, are ($L_m = 1/(1 - q)$), and ($L_m = 1/(1 - q_q)$), where q and q_q are simple and conditional drought probabilities at the respective truncation levels. The drought magnitude (M_T) is obtained through a multiplicative relationship linking the mean (or the mean and variance) of the drought intensity with the drought length. The M_T based on the MC0 with a mean of (I_d) is named MT0, and the one with both the mean and variance of (I_d) is named MT0V. Similar notations stand for the MC1 scenario, i.e., MT1 and MT1V. The drought magnitude (M_T) is transformed into drought volume ($D_T = \sigma_{av} \times M_T$) equivalent to reservoir capacity (C_R). M_T could take the form of MT0, MT0V, MT1, or MT1V. All these values are shown in columns 6 and 10 in relevant tables. As noted earlier and reiterated here, in the process of evaluating the drought magnitude, M_T , the sequences of SHI (i.e., standardized monthly sequences) form the computational platforms. The MA1 derivative of SHI sequences was adequate for all rivers except the Beaver River, for which the MA2 derivative was required only at the draft ratio of 0.75.

3.3. Computational Procedure for the BA Method

The Behavior Analysis also employs Equation (2), and a value of reservoir capacity C_R is assumed. In the process of routing, the initial value Z_{t-1} is assumed to be equal to C_R . All the negative and zero values of Z_t are set to zero. The sum of all failure months is denoted as 'nf', and the ratio of 'nf' to total months ($12N$) in the analysis represents the probability of failure PF. The values of $Z_t \geq C_R$ are set equal to C_R in the process of routing. When the

preassigned value of PF (say 0.05 or 5%) does not match the calculated value shown above, another value of C_R is assumed until the computed PF corresponds to the desired PF. In general, the BA method is simple in terms of computational rigour. The monthly sequences can be routed without standardization, unlike the DM method.

4. Results and Discussion

4.1. Identifying the Number of Zones (Nz) for Use in the GPM Method

An important point is to establish the number of zones to be adopted in the GPM method in view of the diverse values of cv_a and ρ_a . It should be noted that the cv_a ranged from 0.13 to 0.72, and ρ_a ranged from 0.06 to 0.63. Therefore, the analysis was carried out with two sets of zones, i.e., Nz equal to 15 or 20. The values of PF equal to 2.5, 5, and 10% were chosen with draft ratios of 75% and 50%. The C_R values so computed are summarized in Table 2. A comparison was made between the pairs of columns 2, 5; 3, 6; and 4, 7; at the draft ratio of 75% and likewise between pairs of columns 8, 11; 9, 12; and 10, 13; at the draft ratio of 0.50 (Table 2). It is apparent from this table that these values within the pairs are almost indistinguishable, with little change between them. Some of these differences have arisen out of rounding off errors such as those associated with the Beaver River (shown in bold letters). This observation suggests that Nz equal to 15 is adequate for the analysis. Therefore, all the analyses reported in this paper on the GPM method were accomplished using Nz equal to 15. The computation time and rigour are much less compared to the use of Nz equal to 15 than Nz equal to 20. This finding is in sync with the recommended values in the literature [3,4] in accordance with the range of cv_a .

Table 2. Comparison of the estimates of reservoir capacity using 15 and 20 zones.

River and Relevant Parameters	C_R (10^6 m ³) at Draft Ratio = 0.75						C_R (10^6 m ³) at Draft Ratio = 0.50					
	Nz = 20			Nz = 15			Nz = 20			Nz = 15		
	PF Level (%)			PF Level (%)			PF Level (%)			PF Level (%)		
	10	5	2.5	10	5	2.5	10	5	2.5	10	5	2.5
1	2	3	4	5	6	7	8	9	10	11	12	13
Bow River, $cv_a = 0.13$, $\rho_a = 0.06$	290.0	321.0	337.2	290.0	321.0	339.0	133.0	150.0	160.0	133.0	149.0	160.5
Beaver River, $cv_a = 0.72$, $\rho_a = 0.36$	442.0	690.0	935.0	450.0	677.0	936.0	115.0	175.5	246.0	116.0	175.0	244.0
Sturgeon River, $cv_a = 0.43$, $\rho_a = 0.63$	287.0	500.0	712.0	286.0	497.0	712.0	11.5	38.0	74.8	11.5	38.0	74.0
Islands lake River, $cv_a = 0.28$, $\rho_a = 0.27$	280.0	493.0	680.0	282.0	490.0	682.0	10.0	40.0	74.0	10.0	40.2	74.0
English River $cv_a = 0.32$, $\rho_a = 0.20$	322.0	500.0	712.0	320.0	505.0	712.0	50.0	107.0	158.0	49.0	107.0	157.0
Beconcour River $cv_a = 0.20$, $\rho_a = 0.03$	176.5	235.0	281.0	177.0	234.0	281.0	60.3	86.7	112.8	60.3	86.7	112.8
U. Humber River $cv_a = 0.13$, $\rho_a = 0.18$	277.0	358.0	430.0	277.0	358.0	429.0	103.0	142.0	185.0	103.0	142.0	185.0

NOTE: The bold numerals show the discrepancy in the estimates of C_R but they are within 2% of each other. Otherwise, all other values are much less than 2%.

4.2. Inter-Comparison of Reservoir Capacity (C_R) for Rivers with Independent Flows

The rivers under study can be classified into two groups: group A, with a negligible level of autocorrelation or the annual flows that can be regarded as independent in the statistical sense ($\rho_a \leq 1.65/N^{0.5}$, N being the sample size at the 90% level of confidence), and group B, with a significant level of autocorrelation ($\rho_a > 1.65/N^{0.5}$). Using the above criterion, rivers # 1, 9, 10, 11, 12, 14, 15, and 16 (Table 1) fall in group A, and rivers # 2, 3, 4, 5, 6, 7, 8, and 13 fall in group B.

The values of C_R were computed for rivers in group A, and the results are summarized in Table 3. It is apparent from Table 3 that the best comparable C_R values were found between the BA and DM methods. The divergence was found for the estimates based on the GPM, as shown by the bold numerals in Table 3. The major discrepancies were discovered at the PF level of 10%. The minimal discrepancies were observed at the PF of 5%, meaning that at this level, all the methods yield comparable estimates of C_R . The BA method is a simple water balance analysis through Equation (2); the estimates based on this method can be deemed to be the most accurate as there are no parameters involved in applications of the water balance equation. Therefore, the BA-based values of the reservoir capacity can be taken as reference entities for comparison purposes.

Table 3. Comparison of estimates of reservoir capacity using GPM, BA, and DM methods with insignificant autocorrelation (ρ_a , i.e., independence) in the annual flows.

River # as Shown in Table 1, Figure 1	Reservoir Capacity (C_R , 10^6 m ³) at Draft Ratio = 0.75					Reservoir Capacity (C_R , 10^6 m ³) at Draft Ratio = 0.50			
	PF (%)	GPM	BA	DM	Parameters on DM Method	GPM	BA	DM	Parameters on DM Method
1	2	3	4	5	6	7	8	9	10
[1] Bow River, $cv_a = 0.13$, $\rho_a = 0.06$, N = 110	10.0	290.0	286.0	283.0	MT1V, $\Phi = 0.64(1)$	133.0	133.0	128.0	MT1, $\Phi = 0.35(1)^*$
	5.0	321.0	323.0	317.0	MT1V, $\Phi = 0.54(1)$	149.0	151.0	147.0	MT1, $\Phi = 0.21(1)$
	2.5	339.0	343.0	336.0	MT1V, $\Phi = 0.49(1)$	160.5	164.8	160.0	MT1, $\Phi = 0.12(1)$
[9] Pic River, $cv_a = 0.24$, $\rho_a = 0.13$, N = 50	10.0	340.0	325.0	317.0	MT1, $\Phi = 0.48(3)$	134.0	133.0	126.0	MT1, $\Phi = 0.78(3)$
	5.0	480.0	490.0	483.0	MC1V, $\Phi = 0.04(3)$	193.0	195.0	192.0	MT1, $\Phi = 0.45(3)$
	2.5	579.0	588.0	578.0	MC1V, $\Phi = 0.61(3)$	255.0	270.0	255.0	MT1, $\Phi = 0.13(3)$
[10] Pagwa-chaun, $cv_a = 0.25$, $\rho_a = 0.06$, N = 53	10.0	208.0	182.0	177.0	MT1, $\Phi = 0.43(3)$	108.0	75.0	73.3	MT1, $\Phi = 0.69(3)$
	5.0	237.0	240.0	239.0	MT1, $\Phi = 0.12(3)$	127.0	108.0	106.0	MT1, $\Phi = 0.38(3)$
	2.5	250.0	300.0	296.0	MT1V, $\Phi = 0.62(3)$	130.0	152.0	150.0	MT0V, $\Phi = 0.20(3)$
[11] Goulis, $cv_a = 0.24$, $\rho_a = 0.13$, N = 53	10.0	100.0	99.0	98.0	MT1, $\Phi = 0.62(3)$	31.5	31.0	30.2	MT1, $\Phi = 0.93(3)$
	5.0	135.0	138.5	136.0	MT1, $\Phi = 0.35(3)$	44.5	45.0	45.4	MT1, $\Phi = 0.69(3)$
	2.5	166.0	169.0	167.0	MT1, $\Phi = 0.13(3)$	65.0	66.0	64.4	MT1, $\Phi = 0.39(3)$
[12] Becancour, $cv_a = 0.20$, $\rho_a = 0.03$, N = 53	10.0	177.0	175.0	172.0	MT1, $\Phi = 0.57(3)$	60.3	61.2	61.1	MT1, $\Phi = 0.88(3)$
	5.0	234.0	230.0	231.0	MT1, $\Phi = 0.32(3)$	86.7	86.7	86.03	MT1, $\Phi = 0.64(3)$
	2.5	281.0	275.0	281.0	MT1, $\Phi = 0.12(3)$	112.8	114.0	112.1	MT1, $\Phi = 0.40(3)$
[14] Lepraeue, $cv_a = 0.22$, $\rho_a = 0.10$, N = 101	10.0	39.9	38.0	35.5	MT1, $\Phi = 0.69(3)$	8.0	13.1	13.4	MT0, $\Phi = 0.67(3)$
	5.0	45.0	48.0	47.2	MT1, $\Phi = 0.50(3)$	16.0	20.0	20.4	MT1, $\Phi = 0.64(3)$
	2.5	61.5	61.7	60.8	MT1, $\Phi = 0.28(3)$	24.2	27.5	27.7	MT1, $\Phi = 0.40(3)$
[15] U. Humber, $cv_a = 0.13$, $\rho_a = 0.18$, N = 68	10.0	277.0	277.0	275.0	MT0, $\Phi = 0.23(3)$	103.0	104.0	98.9	MT1, $\Phi = 0.32(3)$
	5.0	358.0	355.0	351.0	MT1, $\Phi = 0.20(3)$	142.0	136.0	134.0	MT1, $\Phi = 0.52(3)$
	2.5	429.0	412.0	402.0	MT0V, $\Phi = 0.63(3)$	185.0	184.5	179.0	MT1, $\Phi = 0.20(3)$
[16] Torrent, $cv_a = 0.15$, $\rho_a = 0.18$, N = 61	10.0	97.0	97.0	95.3	MT1, $\Phi = 0.62(1)$	35.2	34.5	32.5	MTV0, $\Phi = 0.90(1)$
	5.0	125.0	125.0	122.0	MT1, $\Phi = 0.43(1)$	49.0	49.0	48.1	MT1, $\Phi = 0.80(1)$
	2.5	146.0	147.0	146.0	MT1, $\Phi = 0.75(1)$	63.0	63.0	59.7	MT1, $\Phi = 0.46(1)$

Asterisk (*) number within parentheses indicates truncation level on SHI sequences such as 1 means SHI_{10} , 3 means SHI_{30} . MT1 and MT0 mean the drought magnitude is based on Markov chain (MC) order 1 or 0 multiplied by the mean of drought intensity (I_d). MC1V and MC0V are based on both the mean and variance of the drought intensity (I_d).

The GPM-based values were plotted against the BA estimates as pairs, and likewise, the DM-based values against the BA based are shown in Figure 2A,B. The correspondence between the estimates by the GPM and DM methods can be judged by using Nash–Sutcliffe efficiency (NSE) and the mean relative error (MRE), as used in [29] and shown in the respective figures. These statistics at the draft ratio of 75%, turned out to be as NSE

(=99.12%) and MRE (=−0.15%) for the pair of the GPM and BA methods, whereas, for the GPM and BA pair, the values were NSE (=99.94%) and MRE (=−1.37%). Similar behavior was displayed at the draft ratio of 50%, in which NSE = 99.73%, and MRE is low (=−2.77%) by the DM method. However, the GPM method yielded NSE = 97.65% (2% smaller than the DM method) with MRE=−0.07%. Although all methods can be deemed to yield comparable values of C_R , the values by the DM method on the 1:1 line against the BA method fit better (square points in Figure 2A,B) compared to the GPM-based values (diamond points). It can also be observed that in a few instances, the GPM-based estimates are larger than the estimates by the BA and DM methods. This anomaly can partly be explained by rounding off errors to arrive at the exact value of the PF. For example, if the required PF is 2.5% in the iteration process, the closest PF reached could be 2.48% or 2.54%. Under such a circumstance, the C_R value, say, corresponding to PF (=2.54%), can be taken as equivalent to PF (=2.50%). It should also be noted that the values of NSE are quite high (around 99%), which is vindicated by the plots in which all the points are falling almost on a 1:1 line, meaning that the variance of residuals is very small compared to the variance of the BA based (abscissa) values. Nonetheless, the MRE has values of −0.07 (at draft ratio = 75%) and −2.77 % (draft ratio = 50%), further suggesting that the deviations are very small, alluding to the fact that all methods, viz. the BA, the DM, and the GPM yield nearly same estimates of C_R for the case of independent annual flows. Under such a situation, one can use the simplest BA method as a preliminary design method, which should be verified by calculations with the DM and the GPM methods.

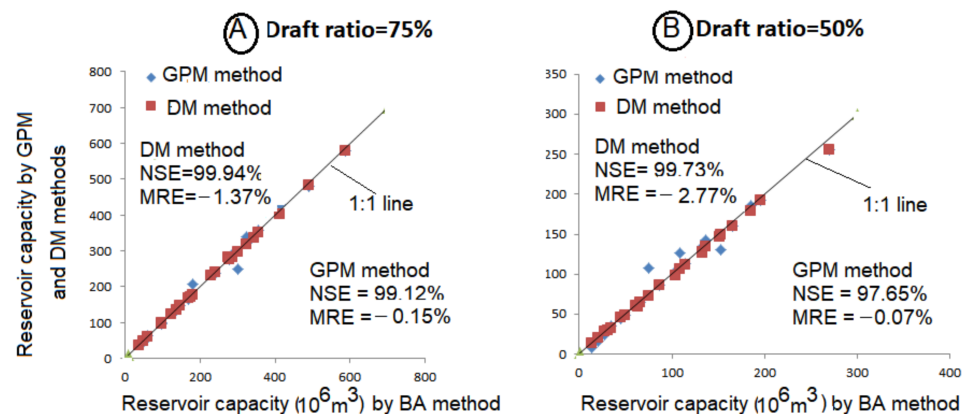


Figure 2. Plots of C_R based on the GPM and DM methods versus the BA method for independent annual flows. (A) Draft ratio = 75%, (B) Draft ratio = 50%.

The DM method requires the standardization of monthly flows to transform them into SHI sequences. In the DM method, it was observed that the MA1 derivative of SHI sequences is adequate for evaluating the C_R values, and the need for an MA2 derivative did not arise. Likewise, in the majority of cases, the truncation level SHI_{av} was the best-fitting line, followed by SHI_o . The MC1 formulation adequately simulated drought lengths, which when multiplied by the mean of the I_d , yielded drought magnitude (denoted as MT1 in Table 3) for transformation into C_R . The DM method has several intermediate steps which espouse the elements of uncertainty. For example, (a) there is an element of uncertainty in the optimal choice of a correct mode of the drought intensity (I_d), i.e., the use of the mean or the variance-based equation to be coupled with Markov chain (MC) based relationship for drought length, (b) the other element of uncertainty arises from the choice of a truncation level (SHI_o , SHI_m or SHI_{av}), and (c) the third important element is the selection of a correct value of the drought length scaling parameter, Φ . The method essentially computes the drought magnitude (M_T), which is converted into C_R using σ_{av} as a multiplier. The σ_{av} is taken to be uniformly applicable to all situations though it may vary slightly with respect to the truncation level, the PF, and the draft ratio. Despite the above limitations, the choice

of σ_{av} as a multiplier seems to be robust because the DM-based estimates compare closely and consistently with the BA estimates (Table 3).

4.3. Inter-Comparison of Reservoir Capacity (C_R) for Rivers with Dependent Flows

The computations of C_R in rivers of group B were carried out similarly to the rivers in group A and are shown in Table 4. An important point to be noted in Table 4 is that for rivers with high ρ_a , the GPM-based values are substantially smaller compared to those based on the BA estimates. It is interesting to note that even for dependent annual flows, the BA and DM-based methods are comparable (columns 4, 5; 8, 9). The divergence between the GPM and the BA estimates is portrayed (diamond points) in Figure 3A,B, whereas the DM estimates against the BA estimates (square points) match much better in the graphical plots. The relevant statistics for NSE and MRE are also displayed in these figures.

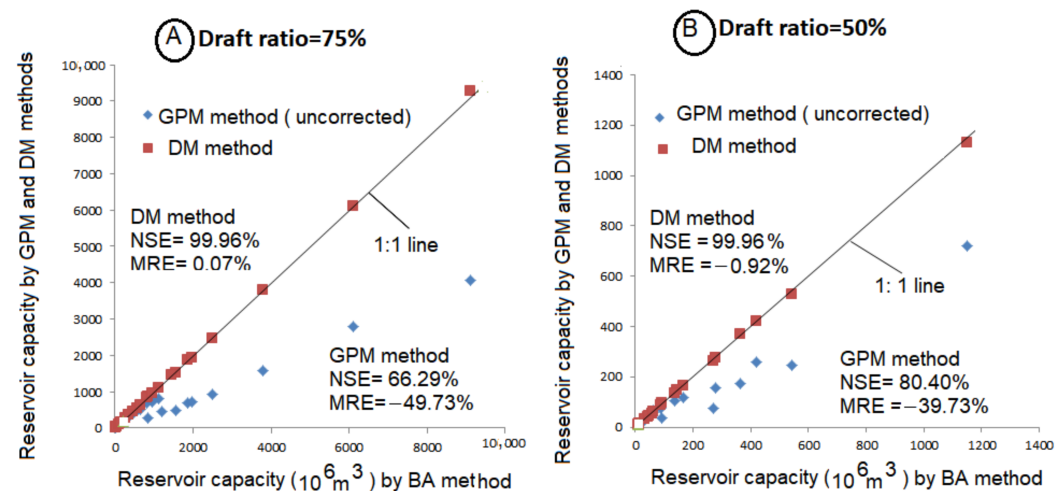


Figure 3. Plots of C_R based on GPM (uncorrected) and DM methods against BA method for annual flows. (A) Draft ratio = 75%, (B) Draft ratio = 50%.

For the dependent flows, it can be noted that NSE between the BA and the DM estimates are quite high ($>99\%$), as portrayed in Figure 3A,B. Likewise, the MER values are low ($= -0.92\%$). On the contrary, for the auto-correlated annual flows, these statistics between the GPM and the BA methods are very discouraging, respectively, with $NSE \approx 66\%$ and 80% and $MER \approx -50\%$ and -40% . Such large values of MER are well reflected in the remarkable departure of diamond points in Figure 3A,B from the 1:1 line. It should be borne in mind that the BA and the DM methods (in monthly sequences) account for the seasonality and persistence effects in their algorithms, regardless of the persistence in the annual sequences. In the GPM method, the annual flow sequences are assumed to be independent, even though the monthly sequences within a year could be dependent. That is why, for the autocorrelated annual flows, the GPM method results in low estimates of the reservoir capacity (C_R). These estimates must be corrected because it is well-known that autocorrelation increases the storage requirement of the reservoirs [1,2]. To this end, McMahon (1976) [17] presented correction factors (CFs) graphically based on the correlation using the reservoir capacity of the BA method as a reference value. The work was further extended by Srikanthan and McMahon (1985) [24], who carried out extensive simulation studies and developed curves displaying the correction factors (CFs) as a function of ρ_a , draft ratio (α), and storage ratio (ratio of storage with zero ρ_a to MAF) and are documented in [3].

Table 4. Comparison of estimates of reservoir capacity using GPM, BA, and DM methods at significant autocorrelation (ρ_a , signifying dependence) in the annual flows.

River # as Shown in Table 1, Figure 1	Reservoir Capacity (C_R , 10^6 m ³) at Draft Ratio = 0.75					Reservoir Capacity (C_R , 10^6 m ³) at Draft Ratio = 0.50			
	PF (%)	GPM	BA	DM	Parameters of DM Method	GPM	BA	DM	Parameters of DM Method
1	2	3	4	5	6	7	8	9	10
[2] Beaver $cv_a = 0.72$, $\rho_a = 0.36$, $N = 65$	10.0	450.0, 810.0	1200.0	1180.0	MT1, $\Phi = 0.24(2)$	116.0, 118.3	165.0	166.0	MT0, $\Phi = 0.70(2)$ *
	5.0	677.0, 1360.8	1860.0	1880.0	MT1V, $\Phi = 0.55(2)$	175.0, 180.3	363.0	368.0	MT0V, $\Phi = 0.61(2)$
	2.5	936.0, 2059.2	2510.0	2460.0	MT1V, $\Phi = 0.25(2)$	244.0, 253.8	541.0	530.0	MT1, $\Phi = 0.66(2)$
[3] Churchill $cv_a = 0.59$, $N = 57$	10.0	1580.0, 3175.8	3800.0	3790.0	MT1, $\Phi = 0.71(3)$	65.0, 71.5	64.0	na	na
	5.0	2800.0, 5684.0	6100.0	6100.0	MT1, $\Phi = 0.27(3)$	260.0, 397.8	420.0	423.0	MT0, $\Phi = 0.45(1)$
	2.5	4080.0, 8445.6	9100.0	9276.0	MT1, $\Phi = 0.66(3)$	720.0, 1188.0	1150.0	1130.0	MT0V, $\Phi = 0.25(3)$
[4] Sturgeon $cv_a = 0.43$, $\rho_a = 0.62$, $N = 35$	10.0	286.0, 606.3	850.0	863.0	MT1, $\Phi = 0.38(3)$	11.5, 21.9	12.0	na	na
	5.0	497.0, 804.3	1550.0	1530.0	MT1V, $\Phi = 0.58(3)$	38.0, 76.4	93.0	97.5	MT0V, $\Phi = 1.0(3)$
	2.5	712.0, 1780.0	1970.0	1940.0	MT0, $\Phi = 0.30(3)$	74.0, 162.8	270.0	265.0	MT1, $\Phi = 0.83(3)$
[5] Islands, $cv_a = 0.28$, $\rho_a = 0.27$, $N = 46$	10.0	282.0, 513.0	280.0	273.0	MT0, $\Phi = 0.78(3)$	10.0, 18.0	10.0	na	MT1, $\Phi = na(3)$
	5.0	490.0, 950.6	572.0	547.0	MT1, $\Phi = 0.67(3)$	40.2, 73.2	44.0	42.2	MT0, $\Phi = 0.1(3)$
	2.5	682.0, 1091.2	830.0	826.0	MT1, $\Phi = 0.31(3)$	74.0, 117.0	88.0	87.3	MT0, $\Phi = 0.20(3)$
[6] Gods, $cv_a = 0.28$, $\rho_a = 0.36$, $N = 46$	10.0	390.0, 393.9	470.0	444.0	MT0V, $\Phi = 0.77(3)$	na	na	na	na
	5.0	816.0, 1003.7	1123.0	1110.0	MT1, $\Phi = 0.84(3)$	14.5, 23.0	20.0	na	na
	2.5	1480.0, 1998.0	1440.0	1470.0	MT1, $\Phi = 0.61(3)$	147.0, 188.2	145.0	146.0	MT0, $\Phi = 0.20(3)$
[7] English, $cv_a = 0.32$, $\rho_a = 0.20$, $N = 99$	10.0	320.0, 329.6	354.0	351.0	MT1, $\Phi = 0.92(3)$	49.0, 56.4	57.0	55.0	MT0, $\Phi = 0.90(3)$
	5.0	505.0, 681.8	645.0	643.0	MT1, $\Phi = 0.61(3)$	107.0, 123.0	136.0	134.0	MT0, $\Phi = 0.05(3)$
	2.5	712.0, 996.9	960.0	960.0	MT1, $\Phi = 0.28(3)$	157.0, 175.8	280.0	274.0	MT1, $\Phi = 0.66(3)$
[8] Neebing, $cv_a = 0.37$, $\rho_a = 0.20$, $N = 66$, $N = 66$	10.0	21.0, 29.4	22.0	22.0	MT1V, $\Phi = 0.95(3)$	10.4, 12.0	9.3	9.1	MT1, $\Phi = 0.61(3)$
	5.0	28.0, 40.6	30.0	30.0	MT1, $\Phi = 0.60(3)$	10.3, 11.9	12.2	12.0	MT1, $\Phi = 0.38(3)$
	2.5	33.8, 52.40	39.0	38.0	MT1, $\Phi = 0.28(3)$	11.4, 12.8	15.0	15.0	MT0V, $\Phi = 0.42(3)$
[13] Beaurivage, $cv_a = 0.26$, $\rho_a = 0.19$, $N = 75$	10.0	92.0, 103.0	92.0	91.1	MT0V, $\Phi = 0.24(3)$	33.0, 35.3	33.0	33.0	MT0, $\Phi = 0.58(3)$
	5.0	118.0, 139.2	117.5	116.0	MT1, $\Phi = 0.25(3)$	44.0, 48.0	47.0	46.0	MT1, $\Phi = 0.49(3)$
	2.5	140.0, 182.0	149.0	147.0	MT0V, $\Phi = 0.46(3)$	58.5, 64.3	61.5	60.7	MT1, $\Phi = 0.21(3)$

Asterisk (*) number within parentheses indicates truncation level on SHI sequences such as 2 means SHI_m , 3 means SHI_{av} . MT1 MT0, MT1V and MT0V are explained in Table 4. “na” means not applicable as the method under this column did not yield the C_R Values. For example, for the Gods River, no method resulted in CR values as they were too low, and the method could not compute the needed values.

These correction factors (CFs) were interpolated from the graphs documented in [3] and are shown in Table 5 (values titled as *lit* which means literature). It can be seen that the correction factors range from 2.18 to 3.12 for rivers with high ρ_a at the draft ratio equal to 0.75 and are smaller in the range from 1.0 to 1.35 at the draft ratio equal to 0.5. It may further be noted that the CFs are not far from 1 when the draft ratio is equal to 0.5, meaning that C_R estimates are closer to the case of independent flows regardless of high ρ_a . In other words, at low draft ratios (<0.5), C_R values based on the GPM method may not require any corrections, even for correlated annual flows.

It may be borne in mind that CFs values are a function of storage ratio, i.e., (the ratio of the GPM storage/ μ_a). The storage ratio can also be computed for the BA-based storage values, i.e., (BA storage/ μ_a). The storage ratios so computed are shown in bold numerals (columns 5 and 9, Table 5). Using these values of CFs (columns 6 and 10, Table 5) as multipliers, the new C_R values were computed and are summarized in bold numerals (Table 4) along with the uncorrected values of C_R .

Table 5. Correction factors (CF) for reservoir capacity (C_R in 10^6 m³) using the GPM method at significant lag-1 autocorrelation (ρ_a , i.e., dependence) in the annual flows.

River # as Shown in Table 1, Figure 1	PF (%)	Reservoir Capacity (C_R , 10^6 m ³) Draft Ratio = 0.75					Reservoir Capacity (C_R , 10^6 m ³) Draft Ratio = 0.50				
		GPM	BA	Ratio C_R/μ_a	CF (lit) *	CF (ob)	GPM	BA	Ratio C_R/μ_a	CF (lit) *	CF (ob)
1	2	3	4	5	6	7	8	9	10	11	12
[2] Beaver $\mu_a = 592.5 \times 10^6$ $\rho_a = 0.36$, $cv_a = 0.72$	10.0	450.0	1200.0	0.76, 2.03 †	1.80 *	2.67	116.0	165.0	0.20, 0.28 †	1.02	1.42
	5.0	677.0	1860.0	1.14, 3.14	2.01	2.75	175.0	363.0	0.30, 0.61	1.03	2.07
	2.5	936.0	2510.0	1.58, 4.23	2.20	2.68	244.0	541.0	0.41, 0.58	1.04	2.21
[3] Churchill $\mu_a = 9345.0 \times 10^6$, $\rho_a = 0.59$, $cv_a = 0.37$	10.0	1580.0	3800.0	0.17, 0.41	2.01	2.41	65.0	64.0	0.007, 0.01	1.10	0.98
	5.0	2800.0	6100.0	0.30, 0.65	2.03	2.18	260.0	420.0	0.028, 0.06	1.53	1.62
	2.5	4080.0	9100.0	0.44, 0.97	2.07	2.23	720.0	1150.0	0.078, 0.12	1.65	1.60
[4] Sturgeon $\mu_a = 1479.5 \times 10^6$ $\rho_a = 0.62$, $cv_a = 0.43$	10.0	286.0	850.0	0.19, 0.57	2.12	2.97	11.5	12.0	0.008, 0.008	1.90	1.04
	5.0	497.0	1550.0	0.34, 1.05	2.30	3.12	38.0	93.0	0.026, 0.062	2.01	2.45
	2.5	712.0	1970.0	0.48, 1.33	2.50	2.77	74.0	270.0	0.050, 0.18	2.20	3.65
[5] Islands lake $\mu_a = 2718.1 \times 10^6$, $\rho_a = 0.27$, $cv_a = 0.28$	10.0	282.0	280.0	0.10, 0.10	1.82	0.99	10.0	10.0	0.004, 0.004	1.80	1.00
	5.0	490.0	572.0	0.18, 0.21	1.94	1.17	40.2	44.0	0.015, 0.016	1.82	1.09
	2.5	682.0	830.0	0.25, 0.31	1.60	1.22	74.0	88.0	0.027, 0.032	1.58	1.19
[6] Gods $\mu_a = 4875.5 \times 10^6$, $\rho_a = 0.36$, $cv_a = 0.28$	10.0	390.0	470.0	0.08, 0.10	1.01	1.21	na	na	na	na **	na
	5.0	816.0	1123.0	0.17, 0.23	1.23	1.38	14.5	20.0	0.003, 0.004	1.21	1.38
	2.5	1480.0	1440.0	0.30, 0.30	1.35	0.97	147.0	145.0	0.03, 0.03	1.28	0.99

Table 5. Cont.

River # as Shown in Table 1, Figure 1	PF (%)	Reservoir Capacity (C_R , 10^6 m ³) Draft Ratio = 0.75					Reservoir Capacity (C_R , 10^6 m ³) Draft Ratio = 0.50				
		GPM	BA	Ratio C_R/μ_a	CF (lit) *	CF (ob)	GPM	BA	Ratio C_R/μ_a	CF (lit) *	CF (ob)
1	2	3	4	5	6	7	8	9	10	11	12
[7] English $\mu_a = 1846.4 \times 10^6$ $\rho_a = 0.20$, $cv_a = 0.32$	10.0	320.0	354.0	0.17, 0.19	1.03	1.11	49.0	57.0	0.026, 0.031	1.01	1.16
	5.0	505.0	645.0	0.27, 0.35	1.35	1.28	107.0	136.0	0.056, 0.074	1.05	1.27
	2.5	712.0	960.0	0.38, 0.52	1.40	1.35	157.0	280.0	0.082, 0.15	1.10	1.78
[8] Neebing $\mu_a = 51.1 \times 10^6$ $\rho_a = 0.20$, $cv_a = 0.37$	10.0	21.0	22.0	0.41, 0.43	1.40	1.05	10.4	9.3	0.20, 0.18	1.15	0.89
	5.0	28.0	30.0	0.55, 0.59	1.45	1.07	10.3	12.2	0.20, 0.24	1.15	1.18
	2.5	33.8	39.0	0.66, 0.76	1.55	1.15	11.4	15.0	0.22, 0.29	1.12	1.32
[13] Beaurivage $\mu_a = 447.6 \times 10^6$ $\rho_a = 0.19$, $cv_a = 0.27$	10.0	92.0	92.0	0.21, 0.21	1.12	1.00	33.0	33.0	0.074, 0.74	1.07	1.00
	5.0	118.0	117.5	0.27, 0.27	1.18	1.00	44.0	47.0	0.10, 0.11	1.09	1.07
	2.5	140.0	149.0	0.31, 0.33	1.30	1.06	58.5	61.5	0.13, 0.14	1.10	1.05

Asterisk (*) refers to the values of CF as shown in the graphical plots in the literature, i.e., in [13], † refers to the CF values computed as the ratio of C_R (BA)/ C_R (GPM), i.e., column 4/column 3. Asterisk (**) denotes “na” (not applicable), which means C_R values were not yielded by the respective methods and neither were CFs.

The graphs were plotted between new C_R (corrected values) against the BA-based values and are depicted in Figure 4A,B. It can be noted that by involving the correction factor, the C_R values improved, but there is still a significant departure from the BA-based estimates. Though the NSE increased (91% and 97.60%), but the MER was still high (−16.15% and −7.85%). This was also noted in the illustrative example of the Sturgeon River, as presented in Appendix A. Taking the BA estimates as the reference values, the corrected GPM values display considerable divergence, which is well corroborated by Figure 4A,B. This alludes to the fact that there is a need to rework these correction factors at low storage ratios at a smaller interval of α and ρ_a through a simulation exercise. As an illustration, these values of CFs were computed as the ratio of C_R by the BA and by the GPM methods and are shown in *italics* (columns 7 and 11, Table 5). These values of the CFs named “ob” (observed) are larger than cited in the literature and thus warrant a need to rework them. Further work on this aspect was conducted by Phatarfod (1986) [30], who derived mathematical expressions which are demanding in comprehension sense and need further simplification by practicing users. The observed CF values can be linked to the results based on the work of Phatarfod [30]. This aspect of the investigation can form a subject of a separate study and is therefore not further explored in this paper.

From the analysis presented in this paper, it is clear that the GPM method is satisfactory for the rivers with little annual autocorrelation, particularly when data are scant and, at times, discontinuous. The method seems less promising for the rivers, which are ridden with significant annual autocorrelations, such as in Canadian prairies. The rivers in Canadian prairies tend to have significant annual lag-1 autocorrelation, ρ_a (>0.20), and the C_R estimates need correction through multiplication by accurate values of CFs. Presently, such values are interpolated from the published literature and therefore are only rough estimates. Evolving the accurate values of CFs using simulation is fairly involved and not easily obtained. On the contrary, the BA and the DM methods can be used without any hassles as long as the monthly streamflow data of 30 years or so are available in a

continuous format and without a break. The algorithms of these methods are designed to account for the first-order persistence in the monthly sequences, and no CFs are involved in the estimation of C_R . The DM method requires the standardization of the data resulting in SHI sequences. In the Canadian rivers, the SHI sequences tend to follow the gamma pdf, which should be normalized using the Wilson–Hilferty transformation as documented in [27]. All the calculations should be conducted in the normalized domain. However, Sharma and Panu [28] noted that the transformation in the normalized probability domain made insignificant improvements in the estimates of C_R compared to the estimation using the SHI sequences in the gamma probability domain itself. In other words, statistically standardized monthly flows designated as SHI can be used for Markov Chain modelling of drought lengths and correspondingly the drought magnitudes as demonstrated in the present analysis. This simple shortcut is sufficiently accurate for the Canadian rivers in which the monthly flow sequences follow the Gamma probability distribution. In other words, the monthly streamflow sequences can be treated as normally distributed in the DM method, though they are mildly skewed, falling in the region of the Gamma pdf.

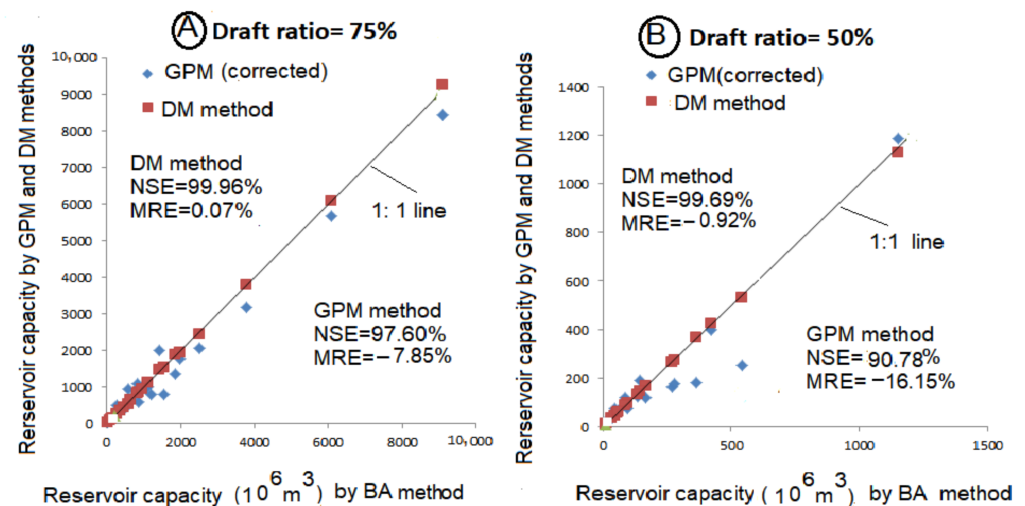


Figure 4. Plots of C_R based on the GPM (corrected) and the DM methods against the BA method for dependent annual flows. (A) Draft ratio = 75%, (B) Draft ratio = 50%.

In the present analysis, two draft ratios, viz. 75% and 50%, were used. There are situations when a draft ratio as low as 30% may require consideration for environmental and/or economic reasons. Under such situations, all the above methods tend to fail, particularly at the PF levels of 10% and 5%, which yield unreliable estimates of the C_R . While confronted with such circumstances, the values of C_R should be estimated at the PF level of 2.5% or less (say 0%), and the C_R values so estimated can be used as design values at all PF levels. In other words, recourse should be taken to the sequent peak algorithm (SPA), which provides the conservative values of C_R [5] with no consideration of PF.

5. Conclusions

Based on the study using the Canadian rivers with wide diversity in terms of annual lag-1 autocorrelation (ρ_a) and coefficient of variation (cv_a), the following conclusions can be derived.

1. For the analysis using the GPM method, 15 zones in a reservoir are sufficient to yield reliable estimates of the reservoir capacity.
2. The estimates of the reservoir capacity (C_R) by the BA and the DM methods in this study were found to be nearly equal to each other for all values ρ_a and cv_a , which implies that the DM method is a competent substitute for the BA method.
3. In the DM method, which requires the standardized monthly data, i.e., the SHI sequences, the Markov chain order 1 (MC1) or MC0 model yielded the drought lengths,

which when multiplied by drought intensity, resulted in the drought magnitudes. The drought magnitude (M_T) multiplied by the average standard deviation of 12 months (σ_{av}) resulted in satisfactory estimates of reservoir capacity (C_R). Both the BA and the GPM methods can be implemented on the non-standardized monthly flow sequences.

4. For annual flows with $\rho_a < 0.20$, the estimates of reservoir capacity (C_R) by the GPM method were found in parity with the estimates of the BA or DM methods. The BA method, being simplest in terms of calculation rigor, enjoys a slight edge over the DM and the GPM methods.
5. The estimates of reservoir capacity (C_R) tend to become smaller in the GPM method when the values of $\rho_a > 0.20$ or become remarkably low with high values of ρ_a , say 0.5 or greater. These low values of C_R can be improved by invoking the correction factors as a multiplier while using the available graphical values from the literature. However, there is a need to further refinement of the correction factors (CFs).

Author Contributions: T.C.S. and U.S.P. collaboratively accomplished various tasks related to this manuscript, including draft preparation and finalization of the manuscript for journal publication, such as conceptualization and development of the methodology, and data collection and data analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The partial financial support of the Natural Sciences and Engineering Research Council of Canada for this paper is gratefully acknowledged.

Conflicts of Interest: No potential conflict was reported by the authors.

Appendix A

An Illustration of the Computational Procedure for the GPM on the Sturgeon River

The case of the Sturgeon River with an assumed value of the reservoir capacity (C_R) = 496×10^6 m³ for PF = 0.05. The lag-1 autocorrelation at the annual scale, $\rho_a = 0.62$. Number of zones in the reservoir, $k = 15$. Volume in each zone = $496/(15-2) = 38.15 \times 10^6$ m³, with first and last zones having volume = 0 m³. The cumulative volume in zones 0 through 14 can be computed as follows.

Table A1. Volume of water in each zone (10^6 m³).

Zone	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Max vol.	0	38.15	76.30	114.46	152.62	190.77	228.92	267.08	305.23	343.38	381.54	419.67	457.85	496.0	496.0
Min. vol.	0	0	38.15	76.30	114.46	152.62	190.77	228.92	267.10	305.23	343.38	381.49	419.69	457.85	496.0

Table A2. Monthly failure tracking.

Zone	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
nf	117	104	79	64	52	41	33	23	19	14	9	6	3	2	1
pf	0.279	0.248	0.188	0.152	0.124	0.100	0.077	0.056	0.045	0.033	0.021	0.014	0.007	0.005	0.002

Table A3. Construction of transitional matrix of the frequency of common occurrences.

Zone	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	9	9	8	7	7	5	5	5	5	5	3	3	2	1	1
1	1	1	2	2	1	3	0	0	0	0	2	0	1	1	0
2	2	2	0	1	1	0	3	0	0	0	0	2	0	1	2

Table A3. *Cont.*[illegible]

Note: The above table shows starting zone (January, z_t) on the X-axis and the finishing zone (December, z_{t+1}) on the Y-axis. For example, the intersection of the abscissa 5 and the ordinate 6 is 2, which means there are 2 instances or occasions when the starting zone(January) was 5 and the finishing zone(December) was 6. Likewise, the intersection of abscissa 9 and ordinate 14 is 16, meaning that there are 16 occasions in which January storage was in zone 9, and it ended in zone 14 in December.

Table A4. Construction of transitional probability matrix (TPM).

[illegible]

Table A5. Construction of steady state probability matrix (squared 6 times i.e (TPM)⁶).

[illegible]

Table A5. Cont.

Zone	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042
3	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
4	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
5	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042
6	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017
7	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
8	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
9	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042
10	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
11	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
12	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.045
13	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068
14	0.516	0.516	0.516	0.516	0.516	0.516	0.516	0.516	0.516	0.516	0.516	0.516	0.516	0.516	0.516
total	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table A6. Evaluation of failure probability.

Zone	Steady-State Probability (from Table A5)	Failure Probability in Zones (from Table 2)	Product of Probabilities Column(2) × Column(3)
(1)	(2)	(3)	(4)
0	0.278571	0.084124	0.023435
1	0.247619	0.014667	0.003632
2	0.188095	0.042007	0.007901
3	0.152381	0.014852	0.002263
4	0.123810	0.007977	0.000988
5	0.097619	0.041791	0.00408
6	0.078571	0.017031	0.001338
7	0.054762	0.019094	0.001046
8	0.045238	0.010877	0.000492
9	0.033333	0.042222	0.001407
10	0.021429	0.027646	0.000592
11	0.014286	0.049574	0.000708
12	0.007143	0.044497	0.000318
13	0.004762	0.067693	0.000322
14	0.002381	0.515947	0.001228
PF =		sum of column (4)	0.050

The above-computed value of PF = 0.05 tallies exactly with the desired value of PF = 0.05 using the trial value of $C_R = 496 \times 10^6 \text{ m}^3$, therefore, the GPM based $C_R = 496 \times 10^6 \text{ m}^3$. However, this C_R must be corrected for the annual autocorrelation ($\rho_a = 0.62$) for this river. The correction factor from the graph by Srikanthan (1985) is =2.01; and thus, the actual $C_R = 2.01 \times 496 \times 10^6 \text{ m}^3 = 997.0 \times 10^6 \text{ m}^3$.

References

- McMahon, T.A.; Pegram, G.G.S.; Vogel, R.M.; Peel, M.C. Revisiting reservoir storage—Yield relationships using a global streamflow database. *Adv. Water Resour.* **2007**, *30*, 1858–1872. [\[CrossRef\]](#)
- McMahon, T.A.; Vogel, R.M.; Pegram, G.G.S.; Peel, M.C.; Etkin, D. Global streamflows—Part 2: Reservoir storage yield performance. *J. Hydrol.* **2007**, *347*, 260–261. [\[CrossRef\]](#)
- McMahon, T.A.; Mein, R.G. *River and Reservoir Yield*; Water Resources Publications: Littleton, CO, USA, 1986; p. 149.
- McMahon, T.A.; Adedoye, A.J. *Water Resources Yield*; Water Resources Publications: Littleton, CO, USA, 2005; p. 88.
- Sharma, T.C.; Panu, U.S. A drought magnitude based method for reservoir sizing: A case of annual and monthly flows from Canadian rivers. *J. Hydrol. Reg. Stud.* **2021**, *36*, 100829. [\[CrossRef\]](#)
- Sharma, T.C.; Panu, U.S. Reservoir sizing at the draft Level of 75% of mean annual flow using drought magnitude-based method on Canadian rivers. *Hydrology* **2021**, *8*, 79. [\[CrossRef\]](#)
- Thomas, H.A.; Burden, R.P. *Operation Research in Water Quality Management*; Division of Engineering and Applied Physics, Harvard University: Cambridge, MA, USA, 1963.
- Loucks, D.P.; Stedinger, J.R.; Haith, D.A. *Water Resources Systems Planning*; Prentice-Hall: Englewood, NJ, USA, 1981; p. 33.
- Linsley, R.K.; Franzini, J.B.; Freyburg, D.L.; Tchobanoglous, G. *Water Resources Engineering*, 4th ed.; Irwin McGraw-Hill: New York, NY, USA, 1992; p. 192.
- Lele, S.M. Improved algorithm for reservoir capacity calculation incorporating storage-dependent and reliability norm. *Water Resour. Res.* **1987**, *23*, 1819–1823. [\[CrossRef\]](#)
- Nagy, T.V.; Asante-Dua, K.; Zsuffa, I. *Hydrological Dimensioning and Operation of Reservoirs. Practical Design Concepts and Principles*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2002; p. 192.
- Alrayess, H.; Zeybekoglu, U.; Ulke, A. Different design techniques in determining reservoir capacity. *Eur. Water* **2017**, *60*, 107–115.
- Moran, P.A.P. Probability theory of dams. *Aust. J. Appl. Sci.* **1954**, *5*, 116.
- Moran, P.A.P. *Theory of Storage*; Methuen: London, UK, 1959.
- Gould, B.W. Statistical methods for estimating the design capacity of dams. *J. Inst. Eng. Aust.* **1961**, *33*, 405–416.
- Gould, B.W. Discussion of paper by Alexander. In *Water Resources Use and Management*; Melbourne University Press: Melbourne, Australia, 1964; pp. 161–164.
- McMahon, T.A. Preliminary estimation of reservoir storage for Australian streams. *Eng. Australia CE* **1976**, *18*, 55–59.
- Theo, C.H.; McMahon, T.A. Evaluation of rapid reservoir yield procedure. *Adv. Water Resour.* **1982**, *5*, 202–215.
- Parks, Y.P.; Farquharson, F.A.K.; Plinston, D.T. Use of the Gould probability matrix method of reservoir design in arid and semi-arid regions. In Proceedings of the Sahel Forums on the State-of-the-Art of Hydrology and Hydrogeology of the Arid and Semi-Arid Areas of Africa, Ouagadougou, Burkina Faso, 7–12 November 1988; Demioissie, M., Stout, E., Eds.; IWRA: Urbana, IL, USA, 1989; pp. 129–136.
- Otieno, F.A.O.; Ndiritu, J.G. The effect of serial correlation on reservoir capacity using the modified Gould probability matrix method. *Water S. Afr.* **1997**, *23*, 63–70.
- Ragab, R.; Austin, B.; Modinis, D. The HYDROMED model and its application to semi-arid and Mediterranean catchments with hilly reservoirs 3: Reservoir storage capacity and probability of failure model. *Hydrol. Earth Syst. Sci.* **2001**, *5*, 563–568. [\[CrossRef\]](#)
- Ibn Abubakar, B.S.U. Assessment of reservoir storage in a semi-arid environment using the Gould probability matrix method. *Afr. Res. Rev.* **2008**, *2*, 35–45.
- Kraus, A.; Rice, M.; Watson, G. *Comparative Analysis of Reservoir Sizing Inclusive of the Gould Probability Matrix Method*; An unpublished undergraduate technical report of the civil engineering; Faculty of Engineering, Lakehead University: Thunder Bay, ON, Canada, May 2022.
- Srikanthan, R.; McMahon, T.A. Gould's Probability matrix method the annual autocorrelation problem. *J. Hydrol.* **1985**, *77*, 135–139. [\[CrossRef\]](#)
- Environment Canada. HYDAT CD-ROM Version 96-1.04 and HYDAT-ROM User's Manual. In *Surface Water and Sediment Data*; Water Survey of Canada: Gatineau, QC, Canada, 2020.
- Dracup, J.A.; Lee, K.S.; Paulson, E.G. On the statistical characteristics of drought events. *Water Resour. Res.* **1980**, *16*, 289–296. [\[CrossRef\]](#)
- Sharma, T.C.; Panu, U.S. A semi-empirical method for predicting hydrological drought magnitudes in the Canadian prairies. *Hydrol. Sci. J.* **2013**, *58*, 549–569. [\[CrossRef\]](#)
- Sharma, T.C.; Panu, U.S. Modelling of hydrological drought durations and magnitudes: Experiences on Canadian streamflows. *J. Hydrol. Reg. Stud.* **2014**, *1*, 92–106. [\[CrossRef\]](#)
- Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models: Part 1—A discussion of Principles. *J. Hydrol.* **1970**, *10*, 282–290. [\[CrossRef\]](#)
- Phatarfod, R.M. The Effect of serial correlation on reservoir size. *Water Resour. Res.* **1986**, *22*, 927–934. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.