hydrology

MDPI

Article

Modelling Hydrological Droughts in Canadian Rivers Based on Markov Chains Using the Standardized Hydrological Index as a Platform

Tribeni C. Sharma and Umed S. Panu *

Department of Civil Engineering, Lakehead University, Thunder Bay, ON P7B-5E1, Canada; tcsharma@lakeheadu.ca

* Correspondence: uspanu@lakeheadu.ca

Abstract: The standardized hydrological index (*SHI*) is the standardized but not normalized (normal probability variate) value of the streamflow used to characterize a hydrological drought, akin to the standardized precipitation index (*SPI*, which is both standardized and normalized) in the realm of the meteorological drought. The time series of the *SHI* can be used as a platform for deriving the longest duration, L_T, and the largest magnitude, M_T (in standardized form), of a hydrological drought over a desired return period of T time units (year, month, or week). These parameters are predicted based on the *SHI* series derived from the annual, monthly, and weekly flow sequences of Canadian rivers. An important point to be reckoned with is that the monthly and weekly sequences are non-stationary compared to the annual sequences, which fulfil the conditions of stochastic stationarity. The parameters, such as the mean, standard deviation (or coefficient of variation), lag 1 autocorrelation, and conditional probabilities from *SHI* sequences, when used in Markov chain-based relationships, are able to predict the longest duration, L_T, and the largest magnitude, M_T. The product moment and *L*-moment ratio analyses indicate that the monthly and weekly flows in the Canadian rivers fit the gamma probability distribution function (*pdf*) reasonably well, whereas annual flows can be regarded to follow the normal *pdf*. The threshold level chosen in the analysis is the long-term median of *SHI* sequences for the annual flows. For the monthly and weekly flows, the threshold level represents the median of the respective month or week and hence is time varying. The runs of deficit in the *SHI* sequences are treated as drought episodes and thus the theory of runs formed an essential tool for analysis. This paper indicates that the Markov chain-based methodology works well for predicting L_T on annual, monthly, and weekly *SHI* sequences. Markov chains of zero order (*MC0*), first order (*MC1*), and second order (*MC2*) turned out to be satisfactory on annual, monthly, and weekly scales, respectively. The drought magnitude, M_T, was predicted satisfactorily via the model $M_T = I_d \times L_c$, where I_d stands for drought intensity and L_c is a characteristic drought length related to L_T through a scaling parameter, ϕ (= 0.5). The I_d can be deemed to follow a truncated normal *pdf*, whose mean and variance when combined implicitly with L_c proved prudent for predicting M_T at all time scales in the aforesaid relationship.

Keywords: conditional probability; extreme number theorem; Markov chain; standardized magnitude; theory of runs; truncation level; truncated normal probability distribution

Academic Editor: Jorge Lorenzo-Lacruz

Received: 16 December 2024
Revised: 18 January 2025
Accepted: 28 January 2025
Published: 31 January 2025

Citation: Sharma, T.C.; Panu, U.S. Modelling Hydrological Droughts in Canadian Rivers Based on Markov Chains Using the Standardized Hydrological Index as a Platform. *Hydrology* **2025**, *12*, 23. https://doi.org/10.3390/hydrology12020023

1. Introduction

Various definitions have been proposed for the analysis, prediction, and simulation of hydrological droughts. The first ever objective definition was given by Yevjevich (1967) [1] as "the deficiency in the water supply on Earth, or the deficiency in precipitation, runoff or in accumulated water in various storage capacities". Linsley et al. (1982) [2] considered a hydrological drought as a "period during which stream flows are inadequate to supply established uses under a given water management system". By all accounts, it is understood that a hydrologic drought refers to a deficiency of water in stream flows, surface, or subsurface water bodies as to be insufficient to meet water demands in various sectors of human society, e.g., agriculture, hydropower generation, water supply, industry, recreation, navigation, fish habitats, etc., at a given time point and/or space. When dealing with hydrologic droughts based on stream flows, the term streamflow drought is also commonly used. Since hydrological droughts have been the subject of studies recognizing the shortfall of water supplies (for power generation through dams, municipal water uses, irrigation, recreation, etc.) from rivers, therefore, indices based on some statistics of river flow sequences are used to characterize them. Various approaches have been introduced to model hydrological droughts, such as the extreme number theorem [3,4] and Markov chains [5–8], among others. An excellent review of such modelling approaches has been provided by Mishra and Singh (2011) [9].

1.1. Standardized Hydrological Index (SHI) Versus Standardized Precipitation Index (SPI)

For identifying the severity of meteorological droughts, a precipitation sequence is standardized (denoted by $e_i = (x_i - \mu)/\sigma$; x_i is the precipitation sequence with mean μ and standard deviation σ of x_i sequence) and then normalized (converted to standard normal variate, z_i) which is named as the standardized precipitation index (*SPI*) by McKee et al. (1993) [10] and Guttman (1999) [11]. The negative values of z_i denote drought conditions while the positive values denote non-drought conditions. For example, a monthly *SPI* series obtained from a month-by-month standardization of precipitation values is subsequently transformed into a normalized series. Such a series can be truncated at the median level and all those events (or epochs) below the threshold level represent the drought conditions. Different categories of drought, such as mild, moderate, severe, and extreme, have been identified based on the values of *SPI*.

In the context of hydrological droughts, the standardized series of river flows (e_i) is named here as the standardized hydrological index (*SHI*) series (Sharma and Panu, 2010) [12–14] and is not normalized. As long as streamflow sequences obey the normal *pdf*, the resulting *SHI* series will also remain normal. However, if the streamflow series is non-normal, then the *SHI* series will also be non-normal. It should be noted that in the drought literature, the standardized and normalized value of streamflow has also been named as the standardized drought index (*SDI*) [7] and standardized runoff index (*SRI*) [15]. In tandem with *SPI*, Nalbantis and Tsakiris (2009) [7] have also classified hydrological droughts as mild, moderate, severe, and extreme based on *SDI* values.

1.2. Parameters of Hydrological Drought

The important parameters defining a hydrological drought for useful applications are as follows: (a) duration and (b) magnitude (earlier called severity). At times, the ratio of magnitude to duration, called intensity, is also used. It should be mentioned that in the earlier literature, from the 1960s until recently (say, the early part of the first decade of the 21st century in 2000–2006), the term severity in hydrologic droughts was used to denote the cumulative deficit. The cumulative deficit has been commonly referred to as magnitude in the context of meteorological drought or when the drought variable is precipitation [10,16,17]. The acceptance of the term magnitude over the term severity is evidenced by recent publications [12–14,18–20]. This transition seems to be motivated by the ambiguity in the meaning of the term "severity". In the context of meteorological drought, severity is expressed in the form of indices. For example, currently, a popular index termed as the standardized precipitation index (*SPI*) suggested by [10,11] has been used to denote the severity of meteorological droughts. In the context of hydrological droughts, however, severity denotes a deficit in volumetric or depth units and therefore its usage conflicts with the index connotation associated with meteorological drought. The term magnitude alleviates this anomaly and is in sync with the volume (essentially magnitude) connotation associated with the cumulative deficit. In view of the foregoing discussion, the term magnitude has been used to denote the cumulative deficit in this paper.

The basic element for deriving the above parameters is the truncation level, which divides the time series of a drought variable into "deficit" and "surplus" sections. The truncation level is synonymous with the threshold or cutoff level and these terms are used interchangeably in the ensuing text. The parameters of drought, viz., duration, magnitude, and intensity (=magnitude/duration) are functions of the properties of a deficit section. The drought duration (L) has units of time such as year, month, week, or day depending on the time scale of a variable manifesting the drought. For instance, if one is dealing with monthly stream flows, then drought duration will be expressed in months. The term deficit (D) refers to the cumulative shortage below the chosen demand level (represented by threshold level) and thus it has the unit of volume, i.e., m^3 or likewise. For ease of interpretation and the potential inter-comparison of drought scenarios in varied environments, the analysis for durations and deficits can be carried out in the standardized domain. For a standardized flow sequence, the deficit sum (standardized deficit) is termed as magnitude, denoted by M ($M = D/\sigma$), where σ is the standard deviation of the flow sequence [1], which is a dimensionless entity. However, the value of duration (L) remains unchanged with its unit struck off, i.e., a week, a month, or a year depending on the time scale of the sequence being analyzed. The quantity D/L can be termed as an average deficit and in terms of dimensionless entities, it is termed as intensity (I_d) = M/L. The longest run length for a sample size of T years is denoted as L_T and the corresponding magnitude as M_T (or deficit: $D_T = M_T \times \sigma$).

The prediction of hydrologic drought length (L_T) and magnitude (M_T) are crucial for the design and management of drought amelioration measures. The Markov chain (MC) methodology provides a powerful method for such prediction. This paper deals with the prediction of L_T and M_T on annual, monthly, and weekly scales, indulging Markov chains while applying them to Canadian rivers as a case study. It should be borne in mind that L_T and M_T in the ensuing text refer to the expected values in the process of prediction or estimation.

2. Model Preliminaries

2.1. A Note on the Choice of Truncation or Threshold Level

The truncation or threshold level is the most crucial element in analyzing the hydrological drought parameters as discussed above; therefore, the choice of a threshold level is of paramount importance. It is usually expressed in the form of the time-invariant or time-variant statistics of a drought variable. Several investigators have considered the threshold level as the long-term mean or median flow [4,14,21–24], while others as some percentile level in the flow duration curve ranging from Q50 (flows exceeding 50% of the time) to Q95 [25–29]. A flow duration curve could be constructed based on annual, monthly, weekly, or daily flow sequences. For a near-normal probability distribution function (*pdf*) of a drought variable, the mean serves as a desired threshold level, whereas for a skewed distribution, the median is construed as a better measure [12–14]. From the point of view of the design of reservoirs, an important element is the draft ratio (a ratio of yield from the reservoir to the mean annual flow, MAF). The draft ratio may vary from 30 to 90%, but McMahon and Adeloye (2005) [30] adopted a value of 75% when using monthly flows for the design of reservoirs. Looking at the recorded flows from the Canadian rivers, it can be ascertained that the median generally falls around 75% of the mean. This observation, therefore, supports the contention that the median is a better threshold or truncation level for the analysis and prediction of hydrologic droughts. Pragmatically, for a regional drought frequency analysis, a value of threshold levels such as Q70 (flows equaling or exceeding 70% of the time) or Q80 would portray more tangible (severe) drought impacts over the region. When drought impacts are vividly tangible, for example, on a short-term contingency planning basis for the amelioration of droughts, one could even conduct drought investigations at the Q90 level to allow for the mobilization of resources within the constraints of time and cost. In such situations, the daily hydrographs (which are non-stationary in a stochastic sense) are chopped at a uniform level, say Q90. A threshold level can be assigned a probability quantile, such as $q = P(x \leq x_0)$, where q is the probability of a drought corresponding to the threshold level x_0 and P(.) stands for the notation of the cumulative probability. The drought probability q is not only dependent on the threshold level but also on the *pdf* of the drought variable. Sharma (2000) [4] suggested a simple analytical method for determining the drought probability quantile, with q corresponding to a threshold level for the normal, lognormal, and gamma *pdfs* of a drought variable.

2.2. A Note on the Theory of Runs Used in the Probability-Based Modelling of Hydrological Drought

The statistical theory of runs has been a major tool in analyzing hydrological droughts in the probabilistic approach since the early 1960s, with the pioneering work of Yevjevich (1967) [1] and advancement by Sen (1980) [3] and Dracup et al. (1980) [21], among others. In short, a time series of *SHI* sequences, in the theory of runs, is truncated at the desired threshold level, resulting in segments of uninterrupted wet (surplus) or deficit (drought) spells (or runs). The length of a deficit run is equivalent to the duration of a drought event, and a run sum (i.e., a sum of the deficit epochs) is equivalent to a drought magnitude. The sequential occurrences of deficit epochs could be random or may follow a dependence structure (~the simplest dependence structure being the Markovian). In simple terms in such cases, the number of runs (or the number of drought events) can be modelled using the Poisson probability law, the run length (or drought duration) can be modelled using the geometric probability law, and the drought intensity can be modelled by using the truncated normal *pdf*. Thus, the theory of runs provides a suitable tool for analyzing the parameters of hydrological droughts.

One major requirement for applying the theory of runs is that a time series of a drought variable must be statistically stationary. The requirement of stationarity is generally met for annual precipitation or annual streamflow sequences as drought variables. The sequences of monthly or weekly stream flows are non-stationary and must be transformed into stationary time series by standardization month-by-month or week-by-week (known as stationarization), as the case may be for hydrological drought investigations. Once a suitable probability distribution describing a monthly or a weekly flow sequence has been identified, the underlying dependence structure of the flow sequence can be investigated. Succinctly, the *SHI* series (month-by-month or week-by-week standardized series of river flows) is a stationary time series, which can be used for the analysis and modelling of a hydrologic drought.

2.3. Estimation of Drought Probability (q) Based on a Threshold Level

For a time series, x_i truncated at a threshold x_0, the standardized threshold level is expressed as $e_0 = (x_0 - \mu)/\sigma$. Further, if x is normally distributed, so would be e. A normally distributed or normalized sequence 'e' can be written as a standard normal deviate (z) and for such a sequence, the probability $q = P(x \leq x_0) = P(z \leq z_0)$ is evaluated as follows.

$$q = P(z \leq z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} exp(-0.5\, z^2) dz = F(z_0) \quad (1)$$

For example, the value of q at a threshold level equivalent to the mean flow for a normal *pdf* of a streamflow sequence is 0.5, which can be obtained by the integration of the above standard normal probability function from −∞ to 0 ($z_0 = 0$). Likewise, the truncation level z_0 for the drought probability $q = 0.3$ can determined from the normal probability table, or the following polynomial function (Equation (2)), as discussed in [31], can be used iteratively. The advantage of Equation (2) is that it can be easily programmed on an Excel Macro platform.

$$F(z_0) = B = 0.5\,[1 + 0.196854|z_0| + 0.115194|z_0|2 + 0.000344|z_0|3 + 0.019527|z_0|4]^{-4} \quad (2)$$

where B is the dummy variable as a short notation for the value of the right-hand side expression, i.e., $q = B$ for $z_0 < 0$ and $q = 1 - B$ for $z_0 \geq 0$. The notation $|z_0|$ represents the absolute value. The error evaluated by this formula is less than 0.00025.

For the gamma distribution, a value of q is obtained by computing normalized standard deviate z_0 corresponding to the standardized threshold level e_0 using Wilson–Hilferty transformation [32] in combination with Equation (2) or the standard normal probability tables.

$$z_0 = 3cv^{-1}([cv\, e_0 + 1]^{0.333} - 1) + 0.333cv \quad (3)$$

For example, in the case of the gamma *pdf*, with $cv = 0.5$ and $e_0 = 0$, q can be estimated firstly by computing z_0 (Equation (3)), which, when plugged in Equation (2), will yield the value of q. The calculations will output $z_0 = 0.167$ and $q = 0.57$. Likewise, at the threshold level equal to the median ($q = 0.5$) and by plugging in the value of $z_0 = 0$, the value of e_0 can be computed equal to −0.055 from Equation (3). Put another way, at $q = 0.5$, the threshold level would be 0 for the normal *pdf*, whereas for the gamma *pdf*, it would be −0.055. Further, it should be noted that e_0 in Equation (3) is equivalent to the truncation level of the *SHI* sequence following the gamma *pdf* and is denoted as SHI_0. At other truncation levels with different values of q, z_0 should first be computed from Equation (1) and then converted to SHI_0 using Equation (3).

Another method of determining SHI_o is by chopping the *SHI* series at different cutoff levels and then computing q by a counting method. The cutoff level, which renders $q = 0.5$, is the appropriate value of SHI_o. This can easily be done by writing a program on a Microsoft Excel-based Macro and is quite accurate. Succinctly, this is a shortcut method in which data are not normalized, but the threshold or truncation level is normalized so as to correspond with the desired probability, q. The method also alleviates the need to calculate the cv, as there is an element of uncertainty as to which estimate of cv should be used in Equation (3). There are three estimates of cv: (a) the ratio of overall σ to overall μ of the flow sequence (denoted as cv_o), (b) the average of 12 monthly cvs (denoted as cv_{av1}), and (c) the ratio of the average of 12 monthly σs to overall mean μ (denoted as cv_{av}). They all differ from each other. The best estimate of cv, based on the shortcut method discussed above, was generally found near to cv_o.

2.4. Evaluation of the Longest Drought Duration, L_T, and Largest Magnitude, M_T

The probabilistic relationships for L_T can be obtained by applying the Markov chain-based relationship, discussed as follows.

When the SHI_i series is cut off at a threshold level SHI_0, the values above the threshold level are positive or in the wet (w) state and those below are negative or in the drought (d) state. So, the SHI_i series can be transformed into a sequence of discrete states in terms of w and d (example: *wwwddddwwwwdd)*. One can define the following notations for probabilities: $P(d)$ (simple probability), i.e., the probability of any time period (week, month, or year) being a drought period at a given truncation level = q; $P(d|d)$ (conditional probability), i.e., the probability of any period being drought period given that past period was also a drought period (which is a first-order persistence) = q_q; $P(d|d,d)$ (conditional probability), i.e., the probability of any period being a drought period given that past and past-to-past periods were also drought periods (second order persistence) = q_{qq}. The same connotations apply to the wet state, i.e., $p = P(w)$; $p_p = P(w|w)$; and $p_{pp} = P(w|w,w)$. Likewise, $p_{qq} = P(w|d,d)$; $q_p = P(d|w)$; and $q_{qp} = P(d|d,w)$. The simple and conditional probabilities sum to 1, as expressed below.

$$p + q = 1;\ p_p + q_p = 1;\quad p_q + q_q = 1;\ p_{pp} + q_{pp} = 1;\ p_{qq} + q_{qq} = 1;\ p_{qp} + q_{qp} = 1 \tag{4}$$

The probability distribution of the lengths of drought and wet spells in sequential occurrences (such as *wwwwdwwddwdddww---d*) can be described by well-known geometric distribution [3,33].

For a second-order Markov chain (*MC*2), the following equations can be derived (Chin 1977 [5], Sharma and Panu 2010 [12]).

$$L_T = 2 - [\log\{T(1-q)q_p q_{qp}\}/\log(q_{qq})] \tag{5}$$

Likewise, for a first-order Markov chain (*MC1*), Equation (5) can be reduced to:

$$L_T = 1 - [\log\{T(1-q)q_p\}/\log(q_q)] \tag{6}$$

For a zero-order Markov chain (*MC0*), Equation (6) reduces to

$$L_T = -[\log\{T(1-q)\}/\log(q)] \tag{7}$$

The parameters q_q, q_p, q_{qp}, and q_{qq} can be estimated by a counting procedure. However, the parameters q_q and q_p can also be estimated from the information of cv, ρ (lag 1 autocorrelation in *SHI* sequences) using the theoretical relationship shown in Equation (8), due to Crammer and Leadbetter (1967) [34]. However, no such equation exists for the estimation of second-order parameters, viz. q_{qp} and q_{qq}.

$$q_q = q + \frac{1}{2\pi q}\int_0^{\rho}[exp\{-0.5e_0^2/(1+\vartheta)\}](1-\vartheta^2)^{-0.5}d\vartheta \tag{8}$$

where v is a dummy variable of integration. The integral in Equation (8) can be evaluated by a numerical procedure and values of q_q for a given ρ and e_0 (SHI_o) can be computed. It should be noted that parameters q_p (= $1 - p_p$) and p_p can be estimated using Equation (8), in which q is replaced by p. The parameters q, q_p, q_{qq}, and q_{qp} can also be estimated by counting the letters '*w*' and '*d*' obtained by cutting off the SHI_i series at the desired threshold level SHI_0 (such that the q value is equal to 0.5) [6,35]. The number of *d*s (say n_1), *dd*s (say n_2, two *d*s occurring consecutively), and *ddd*s (say n_3) are counted. For a sample size of n, one obtains $q = n_1/n$, $q_q = n_2/n_1$ and $q_{qq} = n_3/n_2$. Since there are only two letters, namely w and d, the number of ws = $n - n_1$. The number of pairs successively occurring as *ww* (w is preceded by w) are counted (say n_4). Therefore, $p_p = n_4/(n - n_1)$ and hence, $q_p = (1 - p_p)$. Similarly, the number of pairs occurring as *dw* (d precedes w, say n_5) and appearing as *ddw* (d precedes d which precedes w) are counted (say n_6), which can be used to estimate the value of q_{qp} as n_6/n_5. Such calculations were effectively accomplished by writing a program in Visual Basic on the Microsoft Excel frame. Following the above procedure, the second-order parameters (q_{qq} and q_{pq}) can also be estimated from the non-standardized flow series by trimming at the desired truncation level.

The drought magnitude, M_T, can be estimated from the following relationships involving the *pdf* of I_d and a characteristic drought length, L_c. Beginning with the basic axiom of probability, i.e., $E(M_T) = \int M_T\, P(M_T = Y)$, one can determine the expected value of M_T. Since Y (a value ranging from 0 to, say, a maximum of 200) is a continuous variable, P ($M_T = Y$) can be evaluated as $=P$ ($M_T \le Y$) − P ($M_T \le Y$-Δ), where Δ (delta) is a small increment as used in numerical integration. Employing the extreme number theorem [36], the above probabilities can be evaluated through the following relationship (Sen, 1980) [3].

$$P(M_T \le Y) = exp[-Tq(1-q_q)(1-P(M \le Y))] \quad (9)$$

$P(M \le Y)$ can be determined by employing the normal probability function with mean (μ_M) and standard deviation ($=\sigma_M$) shown in Equations (10) and (11) as follows.

$$\mu_M = L_c\, \mu_d \quad (10)$$

$$\sigma_M^2 = L_c\, \sigma_d^2 \left(\frac{1+\rho}{1-\rho} - \frac{2\rho(1-\rho^{Lc})}{Lc(1-\rho)^2}\right) \quad (11)$$

where ρ is the lag 1 autocorrelation in a series of the drought variable (stream flows) and L_c is a characteristic drought length. The parameters μ_d and σ_d pertain to the *pdf* of drought intensity, I_d, and are assumed to follow a truncated normal distribution. The expressions for μ_d and σ_d can be derived as follows ([3,4]).

$$\mu_d = -\frac{exp(-0.5z_0^2)}{q\sqrt{2\pi}} - z_0 \quad (12)$$

$$\sigma_d^2 = 1 - \frac{z_0\, exp(-0.5z_0^2)}{q\sqrt{2\pi}} - \frac{exp(-z_0^2)}{q^2 2\pi} \quad (13)$$

In the foregoing relationships, z_0 is the standard normal deviate corresponding to the cutoff level SHI_0 and q is the corresponding probability.

Likewise, an estimate of L_c can be obtained from the following relationship (Sharma and Panu, 2013) [13].

$$L_c = \phi L_m + (1-\phi)L_T \quad (14)$$

where ϕ is a scaling parameter in the range from 0 to 1 and L_m is the mean drought length, which, for the first-order dependence, can be expressed as follows.

$$L_m = \frac{1}{1-q_q} \quad (15)$$

In the above formulations, the value of I_d turns out to be negative (since drought epochs are below the threshold level and hence negative in terms of sign); therefore, the absolute value should be used in the calculation of M_T. Further, for the zero-order Markovian structure $q_q = q$. It should be noted that when the analysis is implemented in the standardized domain, L_T, I_d, and M_T are all dimensionless (i.e., without units). A simplistic version of M_T can be represented through Equation (10) in which L_c is replaced by L_T (or $\phi = 0$). This version only considers the mean of I_d and variance is disregarded. In some situations, particularly on an annual scale, it may yield satisfactory results [14]. The deficiency in volumetric units, D_T, can be computed as D_T ($=\sigma \times M_T$) in the respective unit of σ (m^3/year, m^3/month, or m^3/week).

3. Markov Chain-Based Drought Relationships Applied to Canadian Rivers

3.1. Flow Data Acquisition and Preliminary Analysis

Data for the analysis constituted the natural (i.e., unregulated) and uninterrupted flow records of 27 rivers across Canada (Figure 1) and are listed in Table 1. The major considerations for the selection of stations and the period of data were the existence of natural flow regimes with continuous records unprovoked by human intervention and with minimal need for data infilling. Thus, the daily flow data for these 26 rivers were extracted from the Canadian hydrological database (Environment Canada, 2020) [37]. The selected rivers were representative of a wide range of drainage basins (37 km^2 to 32,400 km^2) and a historical database period (year CE 1911 to 2020), which required virtually no data infilling. Daily flows were transformed to weekly flows such that each of the first 51 weeks would be composed of 7 days, while the 52nd week would contain the remainder of the days. That is, the last week of the year would comprise 8 or 9 (during a leap year) days. Monthly and annual flows were already listed in the aforesaid Canadian hydrologic database.

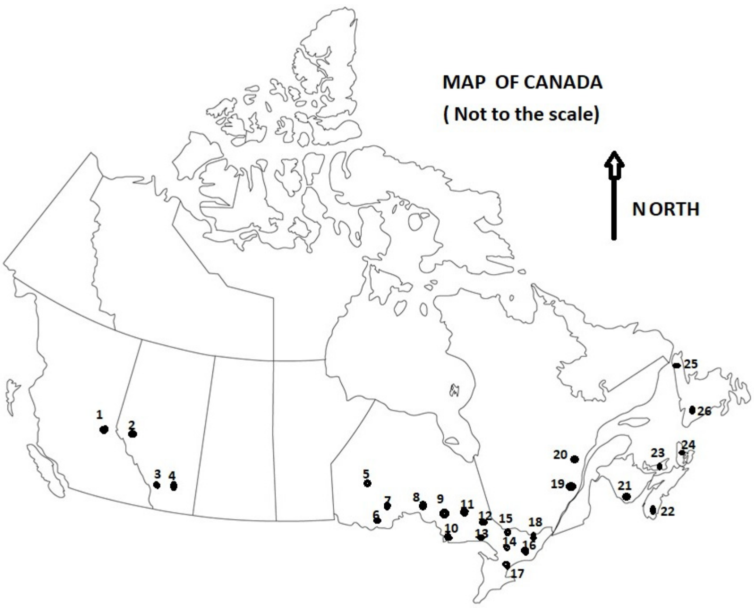

Figure 1. Spatial location of hydrometric stations across Canada used in the analysis (Source: Environment Canada).

Table 1. Statistical properties of the annual, monthly, and weekly flows of the rivers across Canada.

The Numeric Identifier of the River in Figure 1 with Name and Gauging Station Identity	Data Size (Years)	Area (km^2)	μ_o, CV_a, CV_m, CV_w, ρ_a, ρ_m, ρ_w
[1] Fraser at Shelley, BC08KB001 **	70 (1951-20)	32,400	812.72, 0.13, 0.85, 0.90, −0.01, 0.69, 0.75
[2] Athabasca at Athabasca, AB07BE001	69 (1952-20)	74,600	426.30, 0.24, 0.91, 0.35, 0.23, 0.62, 0.81
[3] Bow at Banff, AB05BB001	110 (1911-20)	2210	39.24, 0.13, 1.05, 0.24, 0.06, 0.50, 0.72
[4] South Saskatchewan at Medicine Hat, AB05AJ001	69 (1952-20)	56,369	167.08, 0.32, 1.03, 0.52, 0.06, 0.50, 0.78
[5] Pipestone at Karl Lake, ON04DA001	54 (1967-20)	5960	58.69, 0.31, 0.94, 0.55, 0.08, 0.58, 0.89
[6] Neebing at Thunder Bay, ON02AB008	66 (1954-19)	187	1.62, 0.37, 1.48, 0.81, 0.18, 0.43, 0.63
[7] Pic near Marathon, ON02BB003	50 (1971-20)	4270	50.14, 0.23, 1.03, 0.56, 0.12, 0.41, 0.74
[8] Pagwachaun at Highway#11, ON04JD005	53 (1968-20)	2020	23.01, 0.25, 1.18, 0.62, 0.06, 0.36, 0.74
[9] Nagagami at Highway#11, ON04JC002	70 (1951-20)	2410	24.51, 0.22, 1.01, 0.47, 0.08, 0.49, 0.87
[10] Batchawana at Batchawana, ON02BF001	50 (1971-20)	1190	22.33, 0.20, 1.05, 0.55, 0.04, 0.28, 0.62

[**11**] Goulis near Searchmont, ON02FB002	53 (1968-20)	1160	18.37, 0.21, 1.05, 0.58, 0.10, 0.33, 0.69
[**12**] Whitson at Chelmsford, ON02CF007	60 (1961-20)	243	3.04, 0.24, 1.19, 0.54, 0.11, 0.39, 0.67
[**13**] North French near Mouth, ON04MF001	54 (1967-20)	1190	95.72, 0.21, 1.06, 0.55, -0.06, 0.35, 0.72
[**14**] Labase at North Bay, ON02DD013	56 (1975-20)	70.4	0.91, 0.21, 1.10, 0.61, -0.02, 0.19, 0.43
[**15**] Chippewa at North Bay, ON02DD014	56 (1975-20)	37.3	0.62, 0.19, 0.84, 0.49, 0.01, 0.25, 0.43
[**16**] Commanda at Commanda, ON02DD015	46 (1975-20)	106	1.76, 0.23,0.96, 0.53, -0.15, 0.30, 0.58
[**17**] N. Magnetwan at Pickerel Lake, ON02EA010	52 (1969-20)	149	2.86, 0.23, 0.94, 0.53, 0.08, 0.51, 0.51
[**18**] Shekak at Highway#11, ON04JC003	36 (1951-86)	3290	36.10, 0.18, 1.07, 0.93, -0.10, 0.45, 0.83
[**19**] Becancour A Lyster, QC02PL001	46 (1923-68)	1410	30.60, 0.20, 1.08, 0.62, 0.03, 0.26, 0.62
[20] Beaurivage A. Sainte Entiene, QC02PJ007	75 (1926-00)	709	14.20, 0.26, 1.19, 0.62, 0.19, 0.24, 0.49
[21] Lepreau at Lepreau, NB01AQ001	101 (1919-19)	239	7.43, 0.22, 0.81, 0.59, 0.11, 0.23, 0.49
[22] Carruther at Saint Anthony, PE01CA003	59 (1962-20)	46.8	0.97, 0.23, 1.04, 0.57, 0.07, 0.22, 0.48
[23] Bevearbank at Kinsac, NS01DG003	88 (1922-19)	97	3.04, 0.19, 0.80, 0.59, -0.19,0.13, 0.43
[24] N. Margaree at Margaree Valley, NS01FB001	90 (1929-20)	368	17.01, 0.14, 0.76, 0.47, 0.15, 0.18, 0.46
[25] Upper Humber at Reidville, NF02YL001	68 (1953-20)	2210	80.21, 0.13, 0.87, 0.44, 0.18, 0.13, 0.48
[26] Torrent River at Bristol Pool, NF02YC001	61 (1960-20)	624	24.86, 0.15, 0.88, 0.44, 0.21, 0.16, 0.57

Asterisk (**) indicates the location of the station on the map. In column 1, the first two characters are provincial identifiers to recognize the province in which a particular river is located, viz. BC means British Columbia, etc.

3.2. Identification of the pdf and Dependence Structure of Flow Sequences

The analysis of drought parameters using an analytical approach generally begins with the identification of the *pdf* of the drought variable and its dependence structure. Therefore, the *pdfs* and dependence structures of the annual, monthly, and weekly flow sequences were identified as follows.

3.2.1. The *pdf* and Dependence Structure of Annual Flow Sequences

The statistical parameters μ, *cv*, and lag 1 autocorrelation, ρ (denoted as ρ_a) of annual flows were computed and are summarized in Table 1. The values of the coefficient of skewness (*cs*) were also computed and ranged from 0 to 0.78 with a mean value of 0.35 from a sample size $N = 67$ (average of all gauging stations). Based on the standard statistical test for *cs* [i.e., 95% confidence limits ($0 \pm 1.96 \times \sqrt{(6/N)}$; [38] $N = 67$ for the normal *pdf*], the limits of *cs* were bounded within −0.59 and 0.59. It can be noted that the *cs* values of the annual flows in the majority of stations (Table 1) lay within the above bounds and thus qualified the requirement of a norma1 *pdf*. Further, to affirm the hypothesis of the normal *pdf*, *L*-skewness (*L-cs*) and L-kurtosis (*L-ck*) analyses (Hosking, 1990) [39] were performed on annual flow data. The graph between *L-ck* (ordinate, values from 0.08 to 0.19 with a mean of 0.12) and *L-cs* (abscissa, values from 0 to 0.15 with a mean of 0.07) resulted in a wild scatter around the mean value of *L-ck* of 0.12. It is shown [39] that *L-ck* for the normal *pdf* was independent of *L-cs* and had a constant value of 0.1225. The mean value (=0.12) of *L-ck* from the samples matched the theoretical value of 0.1225 very closely. Thus, these calculations of moments provide strong evidence that the normal *pdf* can be deemed satisfactory to model the annual flow or *SHI* sequences in the rivers under consideration.

For investigating the dependence structure in the annual *SHI* sequences, autocorrelations up to the first 10 lags were computed and these tended to be small (ρ_a in the range of −0.15 to 0.23, Table 1). These values lay within the 95% confidence limits for ρ [=−0.24 to 0.24; ($0 \pm 1.96 \times \sqrt{(1/N)}$ with $N = 67$] [38], suggesting that annual flows could be deemed independent in terms of their consecutive occurrences. Succinctly, for the analysis of hydrologic droughts, the annual stream flows (or *SHI* sequences) of the rivers under question can be regarded as independent normal sequences.

3.2.2. The *pdf* and Dependence Structure of Monthly Flow Sequences

For identifying the *pdf* of the monthly flows, the product moment and *L*-moment relationships were used [39,40]. For each river, the values of the statistics μ, σ, or *cv* and ρ for monthly flows were computed (Table 1) and necessary plots were drawn in terms of product moments and *L*-moments. The scatter plot (*cs* against *cv*) depicts points in the product moment ratio diagram (Figure 2a), in which points are plotted around the theoretical line *cs* = 2 *cv* for the gamma *pdf*. Thus, the plot is a good indicator that the underlying probability distribution of monthly flows follows the gamma *pdf*. To affirm the hypothesis of the gamma distribution, the *L*-moments were computed for the gamma *pdf* and the plot of (*L-ck*) versus (*L-cs*) was drawn [40]. The *L*-moment plot showed a good correspondence between the observed and gamma-distributed points (Figure 2b), thus affirming the hypothesis of the gamma *pdf* as a descriptor of the probabilistic structure of the monthly flows of rivers under consideration.

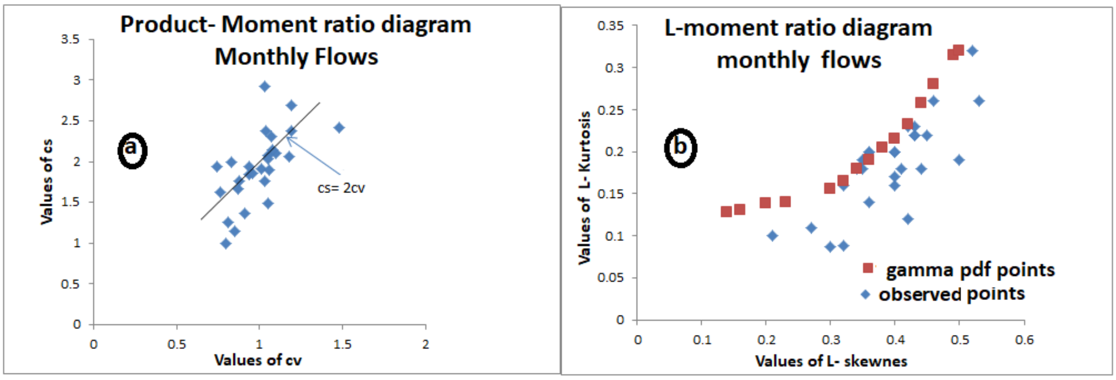

Figure 2. (**a**) Product moment ratio diagram and (**b**) L-moment ratio diagram for monthly flow sequences.

Once the underlying *pdf* of monthly flows is ascertained, the next step is to identify the dependence structure in the *SHI* sequences. Therefore, autocorrelations in the *SHI* sequences from lags 1 to 10 were computed. These autocorrelations tended to decay exponentially, with the lag 1 value (denoted as ρ_m) being significant (confidence limit 0 ± 1.96 × $\sqrt{(1/n)}$, n = sample size in months) [38]. In view of the exponential decay of *SHI*, the monthly *SHI* sequences in the Canadian rivers under question can be regarded as Markovian (autoregressive order-1, *AR-1*) with the gamma *pdf*. The *AR-1* behaviour of *SHI* sequences can be deemed as a precursor to reflect in the first-order Markov chain (*MC1*) structure of drought length and magnitude.

3.2.3. The *pdf* and Dependence Structure of Weekly Flow Sequences

For the identification of a *pdf* and dependence structure of weekly flows, the procedure used for monthly flows was adopted. For each river, the values of the weekly statistics μ, *cv* (Table 1), and *cs* were computed and the necessary plots were drawn in terms of product moments and *L*-moments. The values of *cs* are not tabulated in Table 1 for the paucity of space but are portrayed in Figure 3a.

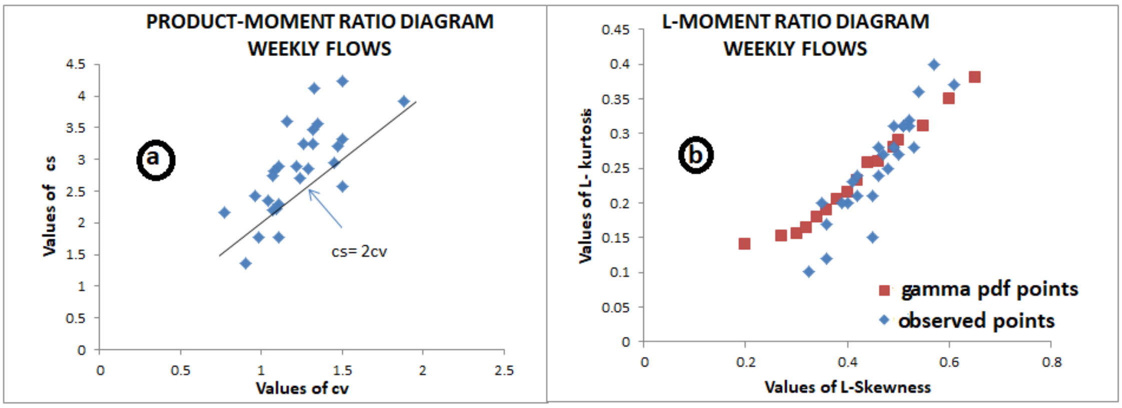

Figure 3. (**a**) Product moment ratio diagram and (**b**) L-moment ratio diagram for weekly flow sequences.

Similar to the case of the monthly flows, the product moments for the weekly flows tend to plot around the theoretical line $cs = 2\ cv$, applicable for the gamma *pdf* (Figure 3a). Likewise, the points for the L-moments are also plotted close to the theoretical values for the gamma *pdf* (Figure 3b). Based on the product moment and *L*-moment diagrams, it can be stated that the weekly flow sequences of the Canadian rivers obey the two-parameter gamma *pdf*. After ascertaining the *pdf* fitting of the weekly flow, the underlying dependence structure of these flow sequences was investigated. For all rivers, the SHI_i sequences were subjected to autocorrelation analysis to uncover the presence of Markovian or other higher order persistence. Very high values of autocorrelations (denoted by ρ_w in Table 1) tend to signify that SHI_i sequences can be regarded to contain memory contents beyond *AR-1*. Under such a scenario, a hypothesis can be formed to model L_T and the associated M_T through *MC2*.

4. Results and Discussion

4.1. Comparison of Predicted and Observed Drought Lengths, L_T

Beginning with the notion that annual flow (or *SHI*) sequences are normal and independent ($\rho = 0$), the value of z_0 is 0.0 for the threshold level equal to the median flow and the corresponding value of q is computed to be 0.5 from Equation (1). Since the ρ value can be assumed to be virtually zero, therefore $q_q = q$ from Equation (8). This leads to the condition in which the structure of drought lengths falls in the regime of *MC0*. Equation (8) thus can be used to predict L_T by plugging in $q = 0.5$; $T = F\ N$ (N = sample size shown in Table 1). The notation F is the correction factor in the plotting position formula to compute the return period, T, using N as the sample size. Adamowski (1981) [41] has suggested $F = (1 + 0.5/N)/0.75$ for flood analysis of the Canadian rivers, that is, for the sample size of $N = 100$, the return period $T = 1.34N$; and for $N = 1000$, $F = 1.33$. It should be noted that in the Weibull formula, $F = (N + 1)/N$. It was noted that the predicted values of L_T (denoted as L_T-p) compared well with the observed values (denoted as L_T-o), as can be seen in the lower portion of the graph in Figure 4a. The average length (duration) of the drought was observed (L_T-o) to be 5 years for an average sampling period of $N = 67$, and the predicted value was found to be 6 years ($T = 90$ years using the Adamowski formula in Equation (8)). As a general rule, one can expect a 100-year drought to persist for 6 years, whereas a 50-year drought can be expected to last for 5 years (Equation (8)). The recent widespread hydrologic drought of 1998–2002 is a testimony to the above finding.

A similar analysis, as used for annual *SHI* sequences, was extended to the monthly *SHI* sequences, in which at the median truncation level ($q = 0.5$), z_0 was computed using

Equation (3). Since monthly *SHI* sequences tend to obey the *AR-1* process [14], therefore, L_T is modelled by *MC1*, represented by Equation (6). An important point to be noted in monthly drought analysis is that N is converted in months and z_0 is based on the gamma *pdf*, whose value would be less than zero at the median threshold level. Also, the T in the calculation was taken as 1.33 N (because N was greater than 360, i.e., a minimum sample size of recorded flows), as discussed above.

All the computations needed were completed through a Macro designed to operate in a Microsoft Excel spreadsheet. The parameters q_q and q_p in the relevant equations were estimated by the analytical Equation (8). The value of ρ (=ρ_m) was taken as the lag 1 autocorrelation from the monthly *SHI* sequences. The cutoff level SHI_0 was obtained by an iterative procedure described in Section 2.3. The average L_T-*o* was observed as 12 months for a sampling period of N = 67 years (804 months) and L_T-*p* was also found to be 12 months. The points of L_T-*o* versus L_T-*p* are depicted in Figure 3a in the middle region where a reasonable correspondence is evident.

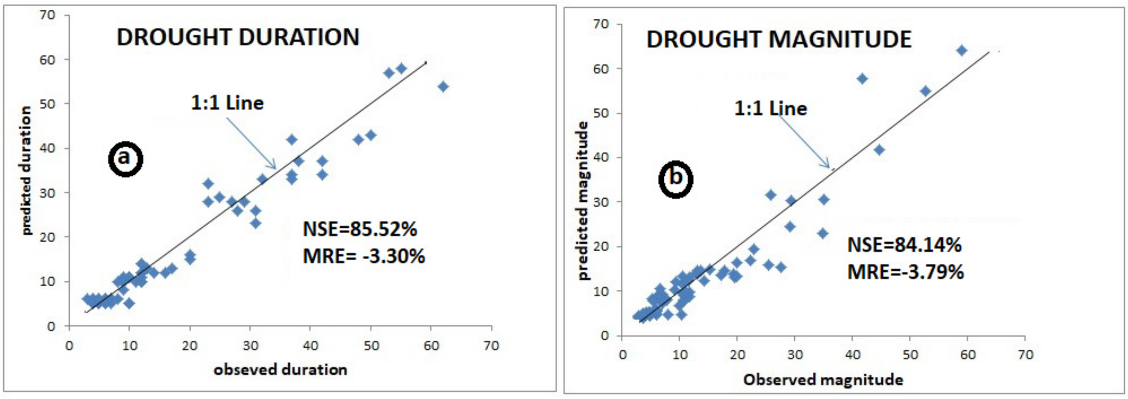

Figure 4. Comparison of predicted and observed (**a**) drought durations, L_T, and (**b**) magnitudes, M_T.

On a weekly scale, the flow sequences of the rivers under consideration tend to display a memory structure exceeding *AR-1*, as is apparent by the very high values of ρ_w (Table 1). Consequently, the *MC*2 structure was considered to model the drought lengths on a weekly scale. Therefore, Equation (5) representing *MC*2 was used to predict L_T-*p*. For each river, the threshold level SHI_0 corresponding to the probability q = 0.5 was computed following the procedure described in Section 2.3. Plugging ρ_w into Equation (8), q_q and q_p were computed. Since there is no formula to compute q_{qq} and q_{qp}, these were estimated by a counting procedure. Thus, the values of L_T-*p* were predicted from Equation (5) by plugging in the parameters q_q, q_p (analytically derived) and q_{qq}, q_{qp} (counting-based). It was observed that the use of ρ_w (based on the *SHI* series) resulted in an under-prediction of L_T-*p*. To ameliorate the under-prediction, ρ_w was replaced by ρ_o (based on the original non-standardized series), and the parameters q_q and q_p were re-evaluated (Equation (8)). This new set of parameters along with q_{qq} and q_{qp} only marginally improved the estimates of L_T-*p*, but under-prediction still persisted. To further improve upon the under-prediction of L_T-*p*, the parameters q_{qq} and q_{qp} were adjusted. A point to be noted is that the second-order parameter q_{qq} was the most sensitive in influencing the estimate of L_T-*p*, whereas the parameter q_{pq} was less sensitive. Further, it was found that q_{qq} was greater than q_q by nearly 2% to 10% with a mean value of 5%. In the absence of any other estimate of q_{qq}, its value was used as 1.05 × q_q and the parameter q_{pq} was left untouched [14]. Using this revised set of parameters, L_T-*p* values were predicted that showed a reasonable correspondence with L_T-*o*. The average length of L_T on a weekly basis was found to be 36 weeks, which was also predicted by the analytical equation based on first-order parameters in combination with the second-order parameters computed using the counting method and suitably adjusted. Accurate estimates by the counting method require large sample sizes (say, N exceeding 10,000) as reported in the literature [5,42]. Garappu et al. (2022) [42] synthesized 500 years of weekly data to arrive at the true values of historical drought parameters over the last millennium.

The L_T-p values estimated through Equations (5) and (6) are plotted against the counterpart values of L_T-o at all three time scales in Figure 4a. The adequacy of fit was judged using NSE (Nash–Sutcliffe efficiency) and MRE (mean relative error) [13,14]. The scatter of the points (predicted and observed values) of $E(L_T)$ on the 1:1 line in Figure 4a appears to be acceptable, as suggested by $NSE \approx 85\%$ with a mean relative error (MRE) of −3.3%. This plot therefore alludes to the reasonable ability of the second-order Markov chain model to simulate and predict the T-week drought durations at the median threshold level in the SHI_i sequences.

4.2. Comparison of Predicted and Observed Drought Magnitudes, M_T

The estimation of the procedure for M_T began with the application of the basic axiom of probability, i.e., $E(M_T) = \int M_T \, P(M_T = Y)$ as discussed in the latter part of Section 2.4. The evaluation of $P(M_T = Y)$ was a bit involved, as it required the estimates of L_T-p, a parameter ϕ, and estimates of drought intensity (I_d) in the form of its mean (Equation (12)) and variance (Equation (13)). At the annual scale, since ρ is assumed to be zero, L_T-p values were computed using $MC0$ (i.e., $q_q = q_p = q$) and the parameter $\phi = 0.5$ was used in the required equations (Equations (9)–(15)). All the calculations were performed by writing a Macro in the Microsoft Excel frame and necessary probability values in the normal probability function were evaluated by the numerical integration of the required equations. The prediction results are depicted in Figure 4b in the lower portion of the graph and are closely well spread around the 1:1 line of M_T-p versus M_T-o, suggesting a good fit. An average value of M_T-p was found to be 4.95 against an M_T-o of 4.55, an over-prediction of ≈9%. This over-prediction was expected in view of the simplification of the model structure, particularly when $\phi = 0.5$ was assumed uniformly for all rivers and SHI sequences were taken as independent normal.

On a monthly scale, the same set of equations was used, and the L_T structure was deemed to follow $MC1$. Therefore, the parameters q_q and q_p were computed using Equation (8) and the values of L_T-p were estimated. The L_T-p values were computed and in combination with $\phi = 0.5$, then predictions of M_T-p were made while considering the mean and variance of the drought intensity function (I_d). The predicted values, i.e., M_T-p, compared well with the observed counterpart, L_T-o, as can be seen in the middle portion (values from 5 to 25) of the plot in Figure 4b. An average value of M_T-p was found to be 11.38 against its counterpart value of M_T-o equal to 11.52, which are very close to each other, suggesting that $\phi = 0.5$ is a judicious choice for prediction purposes.

On a weekly scale, the prediction of M_T-p faces some challenges. The methodology for estimating M_T-p is confined to the $MC1$ structure of drought lengths, whereas they were found to fall in the domain of $MC2$. In a strict sense, Equations (9)–(15) are applicable to $MC1$ conditions. Therefore, the prediction of M_T-p was conducted assuming the weekly SHI sequences to obey $MC1$. That is, the parameters q_q and q_p were estimated using Equation (8), with the ρ_w from the non-standardized weekly flow sequences and L_T (denoted L_{T1}) values computed. These L_{T1} values were plugged into the relevant equations. The predictions turned out to be satisfactory for $\rho_w < 0.75$ as M_T-p matched the observed counterpart (M_T-o) reasonably well. For $\rho_w > 0.75$, the matching was not satisfactory, as the M_T-p values turned out to be far lower than the counterpart M_T-o values. Under this situation, therefore, all equations were used as such, except that the values of L_{T1} (based on $MC1$) were replaced by L_{T2} (based on $MC2$). The computation of parameters for L_{T2} has been well discussed in the aforesaid section. With the use of ϕ = 0.5, the predictions of M_T-p were made, which compared satisfactorily with M_T-o as depicted in the upper portion of Figure 4b. The average value of M_T-p was found to be 37.26, compared to its counterpart value of M_T-o which was equal to 36.22, which are close to each other. The aforesaid simplified modelling strategy for predicting M_T-p seemed to work satisfactorily for Canadian weekly *SHI* sequences that follow the gamma *pdf*.

The M_T-p values estimated through Equations (9)–(15) were plotted against the counterpart values M_T-o at all three time scales in Figure 4b. The adequacy of fit was adjusted using *NSE* and *MRE*, similar to the case of drought lengths, L_T, discussed above. The scatter of points (M_T-p versus M_T-o) on the 1:1 line in Figure 4b appeared satisfactory, as suggested by the *NSE* ≈ 83% with the mean relative error (*MRE*) of −3.8%. This plot, therefore, alludes to a reasonable ability to predict M_T-p by mixing the *MC1* and *MC2* models for computing L_T-p, and that are to be used according to the ρ_w value associated with the flow conditions of a river.

4.3. A Discussion on the Use of MC-Based Drought Models for Drought Amelioration

A Markov chain-based methodology was advanced to predict L_T and M_T on annual, monthly, and weekly scales for Canadian rivers using the *SHI* sequences as the platform. On an annual scale, the methodology is trivial as the annual stream flows tend to be independent and obey the normal *pdf*, and, therefore, L_T and M_T can be easily predicted using Equations (7)–(15) applied on the annual *SHI* sequences. The same set of equations with the gamma *pdf* of flows can be applied to estimate the above parameters on a monthly basis. In both situations, estimates of μ, *cv*, and ρ can provide reliable predictions of L_T and M_T at the median flow level (q = 0.5) and the return period, T (=*1.33 N*, N being the sample size to be used during the model fitting and parameter estimation through the enumeration method). The analysis on a weekly basis takes a tedious turn and thus a recourse is taken to the *MC2* model, in which there is a paucity of closed-form equations for estimating second-order conditional probabilities, viz. q_{qq} and q_{qp}. Therefore, historical data are used to estimate these parameters using a counting (or enumeration) method involving both the non-standardized flow and *SHI* sequences. Potentially, there are three values of M_T for a T-year drought and consequently, three values of D_T. A logical question arises as to which one of them should be used for planning the drought mitigation measures. To elucidate this point, one can consider the case of the Neebing River (station ON02AB008), Canada, with the following statistics: mean flow = 1.61 m^3/s; and σ = 0.60 m^3/s (annual), 1.30 m^3/s (month-by-month averaged out value), and 1.79 m^3/s (week-by-week averaged out value) with ρ = 0 (annual), 0.43 (monthly), and 0.63 (weekly). It should be borne in mind that σ_{av} (average of 12 monthly σs in the monthly analysis, and, accordingly, 52 weekly σs) was found to be the best estimator of σ for use in the conversion equation $D = \sigma \times M$ on the monthly and weekly scales for drought analysis (Sharma and Panu, 2010) [12]. These statistics are based on the sample of 67 years (N = 67, meaning T = 1.34 × N = 90 ≈ 100 years). Using the above statistics, it can be estimated that a 100-year drought is likely to continue for 6 years, 13 months, or 31 weeks, respectively, when analyzed on annual, monthly, and weekly *SHI* sequences. The corresponding M_T values were predicted to be 4.8, 10.4, and 18.94. Based on these statistics, calculations can be conducted to assess the deficit volumes of water in the rivers that are necessary to meet the water demands for various applications. For example, the drought duration when expressed in weeks was found to be 312 (6-year drought = 6 × 52) based on the annual analysis; 56 weeks (13-month drought = 13 × 4.33) based on the monthly analysis; and 31 weeks based on the weekly analysis itself. The advantage in analyzing droughts at a shorter time scale is vividly clear, as the T-year drought would last for 31 weeks only when the conditions are expected to be precarious during the protracted drought. The corresponding values of the deficit volumes D_T (=$\sigma \times M_T$) can be estimated as the following: $0.60 \times 4.8 \times c_1 \approx 90.0$ million m^3, $1.30 \times 10.4 \times c_2 \approx 40.0$ million m^3, or $1.79 \times 18.94 \times c_3 = 20.5$ m^3 on annual, monthly, and weekly scales, respectively. Note that c_1 (=31.5×10^6), c_2 (=2.95×10^6), and c_3 (=0.605×10^6) are conversion constants to convert the annual, monthly, and weekly flow rates into volumes.

Based on the above statistics, one can infer that a 100-year protracted drought can last for 31 weeks, which can occur in any one year or can span in parts from one year to the subsequent year. Most likely, the drier months would fall in the grip of this severe drought, which is apparent from the analysis on a monthly basis (10 months). The most conservative value for the design of a water storage system to make up for the shortfall could be taken as the maximum of the above three values, which is 90 million m^3. In other words, analyses based on three time scales are complementary to each other in providing information for planning drought mitigation measures. Annual analysis, being trivial, is a rapid way to seek information on the proneness of a region to face the severity of drought episodes in terms of the protracted duration and accompanying water shortages. It can be perceived to be a useful tool for the regional mapping of droughts. The weekly analysis, being data-intensive and computationally rigorous, provides finer details on drought scenarios in terms of its persistence time and associated water shortages. Therefore, the week-based analysis could probably be more useful for site-specific drought studies directed to the design of reservoirs, irrigation planning, drought forecasting, rationing of water, or short-term drought management strategies. On the other hand, the monthly analysis is a compromise but would be more complementary to the annual-based analysis, where finer details on the drought frequency, duration, and magnitude are sought for a particular region.

It is to be noted that, in terms of reservoir sizing, the monthly flow sequences are regarded as more desirable [43,44]. Similar findings have been observed [45] in which analysis based on monthly flow sequences resulted in the optimal size of reservoirs. Since the draft at 75% of the mean annual flow is considered adequate (McMahon and Adeloye, 2005) [30], therefore, the cutoff level at median flow is a judicious choice. It should also be noted that drought magnitude estimates as yielded by the present methodology are one method of assessing the reservoir storage among other methods, such as the Sequent peak algorithm, the Behavior analysis, the Gould gamma method, and the Gould probability matrix method [45], among others.

In a strict sense, the aforesaid analysis applies to the *SHI* sequences following the gamma or normal *pdf*. For more non-normal *pdfs* of flows such as lognormal, the above analysis should be carried out in the logarithmic domain. The analysis presented in this paper is accurate enough on annual and monthly scales, as the drought sequences fall in the regime of *MC0* or *MC1*. Under the scenarios of *MC0* and *MC1*, the estimates of parameters can be obtained by theoretical equations, and the small sample size of gauged river flows also yields accurate estimates of parameters. On a weekly scale, the structure of drought lengths and magnitudes falls in the domain of *MC2*, which renders the analysis less accurate in view of the difficulty in estimating the second-order probabilities.

It was noted that the mean, μ_d, of drought intensity, I_d, ranged from 0.60 (week) to 0.80 (annual) and a trivial model $M_T = \mu_d L_T$ could be used to predict drought magnitude as a quick prediction model. For example, on the annual scale, a 100-year drought for the Neebing River was expected to run for a duration of 6 years. The μ_d values for the river were found to be 0.80, 0.72, and 0.70 on the annual, monthly, and weekly scales. On the annual scale, the corresponding $M_T = 0.80 \times 6 = 4.80$. The deficiency of the water in the river due to this extended drought can be computed as follows: $D_T = 0.60 \times 31.5 \times 10^6 \times 4.80 = 90.72$ million m^3, which is virtually the same as was computed by the rigorous calculations involving the variance of I_d in combination with the scaling parameter $\phi = 0.5$. At the monthly scale, L_T = 10 months, so the M_T can be computed as 0.72 × 10 = 7.2. The corresponding D_T = 1.30 × 2.95 × 10^6 × 7.2= 27.61 million m^3, which is less than that found (=40 million m^3) from the rigorous calculations involving variance-based equations but is not too far off. Likewise, the similar estimate on the weekly scale is as follows: M_T = 21.7(=0.70 × 31) and the corresponding D_T = 0.605 × 10^6 × 21.7 = 13.1 million m^3, which is less than the rigorously computed value = 20.5 million m^3. These values using the simple model can be taken as a guide for arriving at the more refined estimates from the variance of I_d-based estimates, which are usually larger.

5. Conclusions

The hydrological droughts based on the annual, monthly, and weekly flow sequences of the 26 rivers analyzed in this paper refer to the *T*-year durations and magnitudes using the sample size (*N*) ranging from 36 to 110 years. Based on three time scales, the flow sequences were standardized and trimmed at the median flow level. For the monthly and weekly droughts, analyses were conducted by the month-by-month or week-by-week standardization of all monthly or weekly flow sequences. The standardization procedure also rendered the non-stationary stochastic monthly or weekly flow sequences into stationary stochastic sequences, thus making the analysis tractable. A standardized flow sequence thus obtained is named a standardized hydrological index (*SHI*), in tandem with a standardized precipitation index (*SPI*), which is commonly used in the context of meteorological droughts. The product moment and *L*-moment analyses revealed that the monthly and weekly streamflow sequences obey the gamma *pdf*, whereas the annual flow sequences can be deemed to follow the normal *pdf*.

The analysis in this paper demonstrated that L_T on an annual and monthly basis can be predicted using the *MC0* or *MC1* models. For the modelling of weekly droughts, the *MC2* model was found to be more adequate. The parameters in the *MC1* model could be evaluated using the closed-form equation with the information of *cv* and lag 1 autocorrelation (ρ) in the *SHI* sequences. The *MC2* model required information on the simple, first-, and second-order conditional probabilities. The simple drought probability q and the first-order conditional probabilities q_q and q_p were obtained based on the closed-form equations using the information of *cv* and ρ of the weekly flow sequences (referred to as non-standardized or original flows). The second-order conditional probabilities, viz., q_{qq} and q_{qp}, were obtained from the non-standardized flow sequences and *SHI* sequences with some adjustments involving the first-order probabilities.

The linkage relationship $M_T = I_d \times L_c$ (in which I_d represents drought intensity and obeys a truncated normal *pdf*, and L_c a characteristic drought length connected to L_T through a scaling parameter ϕ) proved satisfactory for predicting the drought magnitude. The value of $\phi = 0.5$ was found to be uniformly applicable at all time scales. The performance of the above linkage relationship was robust and reliable at the annual and monthly scales, whereas at the weekly scale, it needed adjustment in the values of the parameters.

Author Contributions: T.C.S. and U.S.P. collaboratively accomplished various tasks related to this manuscript, including the draft preparation and finalization of the manuscript for journal publication, such as the conceptualization and development of the methodology, 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: Data is contained within the article.

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

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yevjevich, V. *An Objective Approach to Definitions and Investigations of Continental Hydrologic Drought*; Hydrology Paper 23; Colorado State University: Fort Collins, CO, USA, 1967.
2. Linsley, R.K., Jr.; Kohler, M.A.; Paulhus, J. *Hydrology for* Engineers, 2nd ed.; McGraw Hill: NewYork, NY, USA, 1982.
3. Sen, Z. Statistical analysis of hydrological critical droughts. *ASCE J. Hydraul. Eng.* **1980**, *106*, 99–115.
4. Sharma, T.C. Drought parameters in relation to truncation level. *Hydrol. Process* **2000**, *14*, 1279–1288.
5. Chin, E.H. Modeling daily precipitation occurrence process with Markov chain. *Wat. Resour. Res.* **1977**, *13*, 949–956.
6. Sen, Z. Critical drought analysis by second-order Markov Chain. *J. Hydrol.* **1990**, *120*, 183–202.
7. Nalbantis, I.; Tsakiris, G. Assessment of hydrological drought revisited. *Water Resour. Manag.* **2009**, *23*, 881–897.
8. Akyuz, D.E.; Bayazit, M; Onoz, B. Markov chain models for hydrological drought characteristics. *J. Hydrometeorol.* **2012**, *13*, 298–309. https://doi.org/10.1175/JHM-D-11-019.127.
9. Mishra, A.K.; Singh, V.P. Drought modelling a review. *J. Hydrol*. **2011**, *403*, 147–157. https://doi.org/10.1016/jhydrol2011.03.049.
10. McKee, T.B.; Doesen, N.J.; Kleist, J. The relationship of drought frequency and duration to time scales. In Proceedings of the 8th Conference on Applied Climatology, Anaheim, CA, USA, 17–22 January 1993, pp. 179–184.
11. Guttman, N.B. Accepting the standardized precipitation index: A calculation algorithm. *J. Am. Wat. Resor. Asso.***1999**, *30*, 311–322.
12. Sharma, T.C.; Panu, U.S. Analytical procedures for weekly hydrological droughts: A case study of Canadian rivers. *Hydrol. Sci. J.* **2010**, *54*, 79–82.
13. Sharma, T.C.; Panu, U.S. A semi-empirical method for predicting hydrological drought magnitudes in Canadian prairies. *Hydrol. Sci. J.* **2013**, *58*, 549–569. https://doi.org/10.1080/02626667.2013.772688.
14. Sharma T.C.; Panu, U.S. Modelling of hydrological droughts: Experiences on Canadian streamflows. *J. Hydrol. Reg. Stud*. **2014**, *1*, 92–106.
15. Shukla, S.; Wood, A.W. Use of a standardized runoff index for characterizing hydrologic drought. *Geophys. Res. Lett.* **2008**, *35*, L02405. https://doi.org/10.1029/2007GL032487.
16. Hayes, M. Drought Indices. 2002. Available online: http//www.drought.unl.edu/whatis/indices.pdf. (accessed on 12 July 2024).
17. Sirdas, S.; Sen, Z. Spatio-temporal drought analysis in the Trakya region, Turkey. *Hydrol. Sci. J.* **2003**, *48*, 809–820.
18. Chung, C.; Salas, J.D. Drought occurrence probabilities and risk of dependent hydrologic processes. *ASCE J. Hydrol. Eng.* **2000**, *5*, 259–268.
19. Timilsena, J.; Piechota, T.C.; Hidalgo, H.; Tootle, G. Five hundred years of hydrological drought in the Upper Colorado River basin. *J. Am. Wat. Resor. Asso.* **2007**, *43*, 798–812.
20. Lopez-Moreno, J.I.; Vicente-Serrano, S.M.; Beguria, S.; Garcia-Ruiz; Prortela, M.M.; Almeida, A.B. Dam effects on ght magnitude and duration in a transboundary basin: The lower River Tagus, Spain and Portugal, *Wat. Resour. Res.* **2009**, *45*, W02405. https://doi.org/10.1029/2008WR007198.
21. Dracup, J.A.; Lee, K.S.; Paulson, E.G., Jr. On the definition of droughts. *Wat. Resour. Res.* **1980**, *16*, 297–302.
22. Guven, O. A simplified semi-empirical approach to the probability of extreme hydrological droughts. *Wat. Resour. Res.* **1983**, *19*, 441–453.
23. Rossi, G.; Benedini, M.; Tsakiris, G.; Giakoumakis, S. On regional drought estimation and analysis. *Wat. Resour. Manag.* **1992**, *6*, 249–277.
24. Bonacci, O. Hydrological identification of drought. *Hydrol. Processes* **1993**, *7*, 249–262.
25. Zelenhasic, E.; Salvai, A. A method of streamflow drought analysis. *Wat. Resour. Res.* **1987**, *23*, 156–168.
26. Tallaksen, L.M.; Madsen, S.; Clausen B. On definition and modelling of streamflow deficits and volume. *Hydrol. Sci. J.* **1997**, *42*, 15–33.
27. Demuth, S.; Stahl, K. (Eds) *ARIDE-Assessment of the Regional Impact of Drought in Europe. A Final Report 2001*; Institute of Hydrology, University of Freiburg: Freiburg, Germany, 2021.
28. Hisdal, H.; Tallaksen, L.M. Estimation of regional meteorological and hydrological drought characteristics: A case study of Denmark. *J. Hydrol.* **2003**, *281*, 230–247.
29. Tallaksen, L; van Lanen, H. Hydrological droughts. In *Processes and Estimation Methods for Streamflow and Groundwater*; Elsevier, Amsterdam, The Netherlands, 2004.
30. McMahon, T.A.; Adeloye, A.J. *Water Resources Yield*; Water Resources Publications: Littleton, CO, USA, 2005.

31. Chow, V.T.; Maidment, D.R.; Mays, L.W. *Applied Hydrology*; McGraw Hill Company: New York, NY, USA, 1988.
32. Viessman, W.; Lewis, G.L. *Introduction to Hydrology*, 5th ed.; Prentice Hall: Englewood Cliffs, NJ, USA, 2003.
33. Mathier, L.; Perreault, L.; Bobee, B. The use of geometric and gamma-related distributions for frequency analysis of water deficit. *Stochastic. Hydrol. Hydraul.* **1992**, *6*, 239–254.
34. Cramer, H.; Leadbetter, M.R. *Stationary and Related Stochastic Processes*; John Wiley and Sons, Inc.: New York, NY, USA, 1967; p. 384.
35. Sen, Z. *Applied Drought Modelling, Prediction and Mitigation*; Elsevier Inc.: Amsterdam, The Netherlands, 2015.
36. Todorovic, P.; Woolhiser, D.A. A stochastic model of n-day precipitation. *J. Appl. Meteorol.* **1975**, *14*, 225–137.
37. Environment Canada. 2020. *HYDAT CD-ROM Version 96-1.04 and HYDAT CD-ROM.* User's Manual. Surface Water and Sediment Data Water Survey of Canada .Available online: https://www.canada.ca/en/environment-climate-change/services/water-overview/quantity/monitoring/survey/data-products-services.html (accessed on 12 July 2024).
38. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Jung, G.M. *Time Series Analysis, Forecasting and Control*, 5th ed.; Wiley: Hoboken, NJ, USA, 2015.
39. Hosking, J.R.M. L-moments: Analysis and estimation of distributions using a linear combination of order statistics. *J. R. Stat. Soc. B* **1990**, *52*, 105–124.
40. Vogel, R.M.; Fennessey, N.M. L-moment diagrams should replace product moment diagrams. *Wat. Resour. Res.* **1993**, *29*, 1745–1752.
41. Adamowski, K. Plotting formula for flood frequency. *Water Resour. Bull.* **1981**, *17*, 197–202.
42. Gurrapu, S.; Sauchyan, D.J.; Hodder, K.R. Assessment of hydrological drought risk in Calgary, Canada using weekly river flows of the past millennium. *J. Water Clim. Chang.* **2022**, *13*, 1920–1935.
43. 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.
44. Parks, Y.P.; Gustard, A. Reservoir storage yield analysis for arid and semiarid climate. *IAHS Optim. Alloc. Water Resour.* **1982**, *135*, 49–57.
45. 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. https://doi.org/10.1016/j.ejrh. 2021.100829.