Journal of Hydrology (NZ) 43 (1): 59-73, 2004

Determination of regional probability distributions of Canadian flood flows using L-moments

Sheng Yue[1]* and ChunYuan Wang[2]

[1] *US EPA, Mid-continent Ecology Division, 6201 Congdon Blvd., Duluth, 55804, USA. Fax: (218) 529-5003; Email: yue.sheng@epa.gov.*
**Corresponding author*

[2] *Water Resources & Hydropower Section, Two-evaluation Dept., China Development Bank, No.29, Fuchengmenwai Street, Xicheng District, Beijing, 100037, P.R. China*

Abstract

Knowledge of the probability distribution of flood peak flows is needed for the planning, design, and management of flood control and prevention infrastructure. The average weighted distance, based on L-moment ratio diagrams, and the *Z* statistic, based on L-kurtosis, were applied to identify acceptable types of probability distributions for annual flood peak flows (annual maximum daily streamflow) in Canada. The results demonstrated that the generalised extreme value (GEV), 3-parameter lognormal (LN3), and log-Pearson type III (LP3) probability distributions are reasonable for representing flood peak flows in Canada. However, some differences in the distribution types among different climatic regions were also identified, which should also be considered in the selection of the distribution used for flood peak flows.

Keywords

L-moment ratio diagram; probability distribution; flood flow; regional frequency analysis; statistical hydrology; Canada

Introduction

In order to optimally design and operate flood control and prevention structures such as reservoirs, spillways, and river stopbanks, the statistical properties of flood peak flows are required. A better understanding of the probabilistic behavior of floods is a prerequisite for efficient design and management with the desired degree of reliability. Overestimates of flood peaks will lead to increased construction costs, whereas underestimates may result in increased risk to structures, resulting in tremendous damage or even loss of human lives. The purpose of flood frequency analysis is to select a probability distribution type of flood flows for a site or a region that can be used to represent statistical properties of flood flows and to derive critical values for the design return period.

The necessity of regional frequency analysis of flood flows has been addressed by Cunnane (1987), GREHYS (1996), and others. Since Hosking (1990) introduced the method of L-moment diagrams, numerous studies have been conducted to identify regional probability distributions of flood flows using the approach. L-moments are almost unbiased and less sensitive to outliers in comparison to the conventional product moments, e.g., mean, variance, skewness and kurtosis (Hosking and Wallis, 1997) Studies using the method of L-moment ratio diagrams to select the probability distribution types of flood flows include the works of Wallis (1988), Nathan and Weinmann (1991), Gringas and Adamowski (1992), Pilon and Adamowski (1992), Pearson (1991), Karim and Chowdhury (1993), Rao and Hamed (1994), Vogel and Wilson (1996), and Ben-Zvi and Azmon (1997) among others. These studies demonstrated that the generalised extreme value (GEV), Pearson type III (P3), log-Pearson type III (LP3), 3-parameter log-normal (LN3), and generalised Pareto (GPA) distributions might be suitable for representing the statistics of flood flows, depending on specific regions.

In Canada, the 2- and 3-parameter log-normal (LN2, LN3), Gumbel (GUM), generalised extreme value , gamma (GAM), Pearson type III and log-Pearson type III probability distributions are all recommended distribution types for describing the statistics of flood peak flows (Watt, 1989). There have, however, been few studies of the probability distribution of flood peak flows in Canada. Pilon and Adamowski (1992) found that flood flow data at 23 sites in Nova Scotia may best be described by the generalised extreme value distribution. Gringas and Adamowski (1992) also reported that the generalised extreme value distribution best represented flood peak flows at 53 sites in New Brunswick.

The population or true type of distribution of Canadian flood peak flows is not known—the only way to determine a possible distribution is to examine the distribution followed by sample data. The method of L-moments developed by Hosking and Wallis (1997) provides a reasonable way to investigate the type of distribution of a hydrological variable of interest for a region. Regional frequency analysis is especially useful for watersheds that lack observation records, as the identified regional distribution type can be utilised for water resources management.

This work uses the method of L-moments to identify the probability distributions of flood peak flows across Canada, which are represented by annual maximum daily streamflows. It should provide useful information for Canadian hydrological engineers to select a suitable distribution to use in hydrologic engineering.

Method

L-moment ratio diagrams

L-moments are linear combinations of order statistics that are less sensitive to outliers and virtually unbiased for small samples (Hosking and Wallis, 1997). The L-moments (λ_i, i = 1, 2, 3, 4) of any probability distribution are defined by

$$\lambda_1 = \beta_0 \tag{1a}$$

$$\lambda_2 = 2\beta_1 - \beta_0 \tag{1b}$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{1c}$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{1d}$$

where β_r (r = 0, 1, 2, 3) is the probability-weighted moment, which can be defined as

$$\beta_r = E\{X[F(X)]^r\} \tag{2}$$

in which $F(X)$ is the cumulative distribution function of a random variable X. In practice, the probability-weighted moments must be estimated from a finite sample of data. The unbaised probability-weighted moment estimators are given by the following formula

$$\hat{\beta}_r = \frac{1}{n}\binom{n-1}{r}^{-1}\sum_{j=1}^{n-r}\binom{n-j}{r}x_{(j)} \tag{3}$$

where n is the sample size, and $\{x_{(j)}\}$ is the observations in descending order, i.e., $x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(n)}$.

L-moment ratios: L-coefficient of variation (L-cv, τ_2), L-coefficient of skewness (L-skewness, τ_3), and L-coefficient of kurtosis (L-kurtosis, τ_4) are defined by

$$\tau_2 = \lambda_2/\lambda_1 \quad (4a)$$

$$\tau_3 = \lambda_3/\lambda_2 \quad (4b)$$

$$\tau_4 = \lambda_4/\lambda_2 \quad (4c)$$

Hosking and Wallis (1997) documented that if the mean of a probability distribution exists, then all of the L-moments exist, and the L-moments uniquely define the distribution, i.e., no two distributions have the same L-moments.

L-moment ratio diagrams are the plots of L-cv (τ_2) versus L-skewness (τ_3) for 2-parameter distributions, and L-kurtosis (τ_4) versus L-skewness (τ_3) for 3-parameter distributions. They are useful as a guide in selecting an appropriate distribution for describing a set of variables, as a distinct relationship between L-moment ratios exists for various probability distributions. By comparing sample L-moment ratios to theoretical values via the L-moment ratio diagrams, a probability distribution best fitting the sample data may be obtained.

For the normal, Gumbel, and exponential probability distributions, τ_3 and τ_4 have one value, while τ_2 can have various values. Thus, in a $\tau_3 \sim \tau_2$ diagram, their theoretical $\tau_3 \sim \tau_2$ curves are vertical lines. In a $\tau_3 \sim \tau_4$ plot, they are shown by a point (τ_3, τ_4). For the theoretical relation of $\tau_3 \sim \tau_2$ of 2-parameter probability distributions (gamma, 2-parameter lognormal, and 2-parameter Weibull, the polynomial approximations $\tau_2 = f(\tau_3)$ developed by Vogel and Wilson (1996) are used for analysis. For the $\tau_3 \sim \tau_4$ of 3-parameter distributions—Pearson type III, log-Pearson type III, 3-parameter lognormal, generalised extreme value, 3-parameter Weibull, generalised Pareto and generalised logistic—the polynomial approximations $\tau_4 = f(\tau_3)$ are provided in Hosking and Wallis (1997). The L-skewness and L-kurtosis for the 3-parameter Weibull probability distribution are equal to $-\tau_3$ and τ_4 of the generalised extreme value distribution, respectively. The log-Pearson type III distribution describes a random variable whose logarithms are Pearson type III-distributed. Thus, by comparing the observed $\tau_3 \sim \tau_4$ relations of the logarithms of sample data to the theoretical $\tau_3 \sim \tau_4$ of the Pearson type III probability distribution, the suitability of the log-Pearson type III distribution for representing the sample data can be examined.

Calibration

A visual inspection of the observed and theoretical relationships of L-moment ratios may be used to judge the goodness-of-fit of a probability distribution to the sample data. When the theoretical curve of the L-moment ratio for a probability distribution crosses the center of the cluster formed from the sample L-moment ratios, (i.e., the number of the sample estimates evenly scattered around the theoretical curve), and goes along with the tendency of the cluster, the probability distribution is considered to be suitable for representing the sample data.

Direct visual inspection is somewhat subjective, and it may not be possible to distinguish the differences in two or more probability distribution types when all of them seem to be possible candidates to represent sample data in a L-moment ratio diagram (Peel *et al.*, 2001). Kroll and Vogel (2002) proposed measuring the closeness between sample and theoretical L-moment ratios using the average weighted distance (AWD) method, which is presented by:

$$\text{AWD} = \frac{\sum_{i=1}^{N} n_i \, d_i}{\sum_{i=1}^{N} n_i} \quad (5a)$$

$$d_i = \begin{cases} \left| \tau_2[\tau_3^0(i)] - \tau_2^0(i) \right| & \text{for a 2-parameter} \\ \left| \tau_4[\tau_3^0(i)] - \tau_4^0(i) \right| & \text{for a 3-parameter probability distribution} \end{cases} \quad (5b)$$

where N is the number of sites in analysis; n_i is the record length at site i; $\tau_k^0(i)$ (k =2, 3, and 4) are the observed or sample L-cv, L-skewness, and L-kurtosis, respectively. $\tau_2[\tau_3^0(i)]$ and $\tau_4[\tau_3^0(i)]$ are the theoretical L-cv and L-kurtosis values calculated from a probability distribution corresponding to a given sample L-skewness, respectively. A probability distribution with the smallest average weighted distance value is the best choice among various possible distributions for representing sample data.

A goodness-of-fit measure (*Z* statistic)

A goodness-of-fit measure, *Z* statistic, developed by Hosking and Wallis (1993), can also be used to select a possible regional probability distribution. The *Z* statistic for a distribution is expressed as:

$$Z = (\bar{t}_4^R - \tau_4^{\text{dist}}) / \sigma_4 \quad (6)$$

where $\bar{t}_4^R$ is the regional average L-kurtosis, σ_4 is the standard deviation of $\bar{t}_4^R$, and $\bar{t}_4^{dist}$ is the L-kurtosis of a fitted distribution. A probability distribution with the smallest value of $|Z|$ is considered the best choice among possible distributions. At the significance level $\alpha = 0.10$, the critical value of Z is 1.64, i.e., if a probability distribution whose $|Z| \leq 1.64$, then it is assessed to be an acceptable distribution for representing sample data at $\alpha = 0.10$.

Data

The quality of sample data is crucial for determining a probability distribution type of river flows. Manmade structures and human activities can have significant effects on rivers—upstream containment structures, diversions, and water abstraction for consumptive use may greatly distort records of river flows. Flood frequency analysis using such data would be questionable, so it is desirable to use flow data that are free from the effects of human activities. This study makes use of streamflow data from the gauging stations in the Reference Hydrometric Basin Network (RHBN) database, which is a network of streamflow and water level stations across Canada (Environment Canada, 1999). The basins in the Reference Hydrometric Basin Network are identified as having either pristine or stable hydrological conditions with good quality records. The Reference Network consists of 259 hydrometric stations, including some sites with only seasonal observations. Annual maximum daily streamflow data for representing annual flood peak flows are provided in the Reference Hydrometric Basin Network database CD. The earliest flood peaks used were recorded in 1908, and the latest ones in the data set were recorded in 1998.

Canada is a vast country, with regions with widely varying climate and physiography. Hare and Thomas (1979) and Gullett (1992) divided Canada into 11 major homogenous climatic regions, based on the differences in climatological and physiographical conditions, which are shown in Figure1. Region 1 has luxuriant Pacific rainforest, and a warm, humid Mediterranean climate. Further inland, Region 2 is dominated by sub-alpine

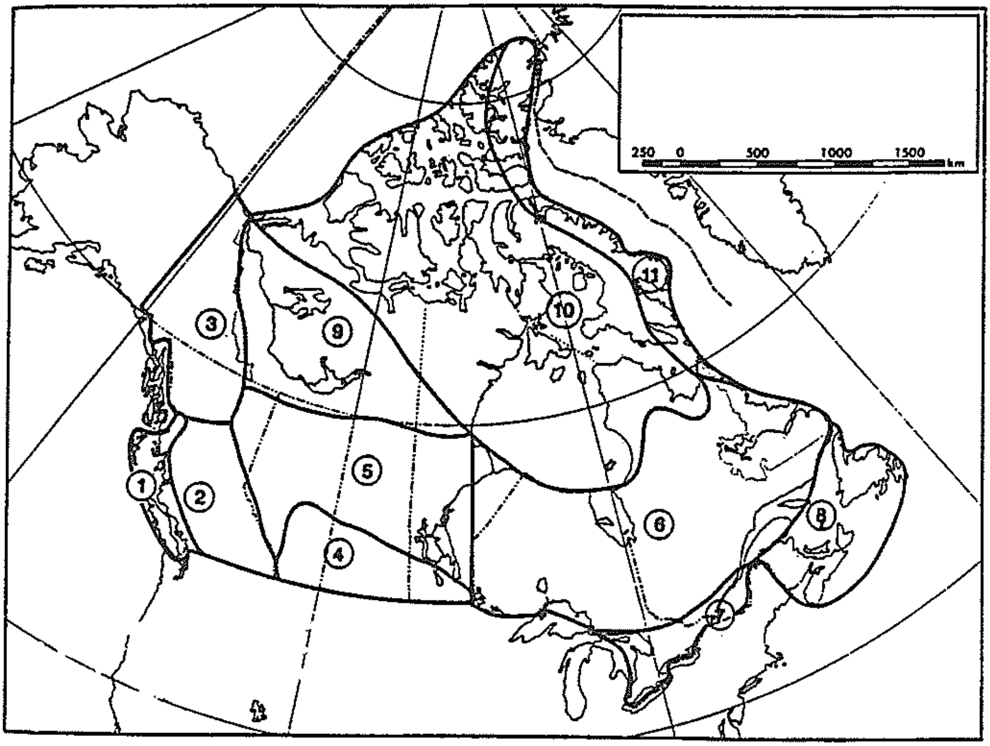

Figure 1 – Major climatic regions of Canada – 1. Pacific; 2. South British Columbia mountains; 3. Yukon and northern British Columbia mountains; 4. Prairie; 5. Northwest forest; 6. Northeastern forest; 7. Great Lakes & St Lawrence; 8. Atlantic; 9. Mackenzie; 10. Arctic Tundra; 11. Arctic mountains.

and Montane forest; Region 3, also known as the Cordillera, has mixed parched grassland with moist forest. Region 4 or the Prairies, are sun-drenched grain fields and grasslands; Region 5 contains a mixture of parkland and boreal forest, which stretches across to Region 6 (Eastern Canada). By the moderating effects of the Great Lakes, Region 7 has predominantly deciduous forest, in contrast to the relatively unpredictable Maritime climate of the Atlantic (Region 8). Region 9 is a mixture of Boreal forest and tundra, while regions 10 and 11 are Canada's frozen, windswept, treeless polar deserts. There is no observation station available in climate region 11 in this study. The possible probability distribution types of flood flows in each of the 10 regions and across the entire country will be analyzed using the two approaches mentioned in Section 2.

Identification results

Discordancy test for site data

One test of discordance is to calculate the D-statistic, defined in terms of L-moments (Hosking and Wallis, 1993, 1997), which can be used to determine whether a site should be removed from a region. Let $\mathbf{u}_i = [\, t^{(i)} \; t_3^{(i)} \; t_4^{(i)} \,]$ be a vector containing the t, t_3, and t_4 values at site i, which are the sample L-moment ratios corresponding to τ_2, τ_3, and τ_4, respectively. The group averages

$\bar{\mathbf{u}}$ and sample covariance matrix $\mathbf{S}$ are defined as:

$$\bar{\mathbf{u}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{u}_i \tag{7a}$$

$$\mathbf{S} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{u}_i - \bar{\mathbf{u}})(\mathbf{u}_i - \bar{\mathbf{u}})^{\mathrm{T}} \tag{7b}$$

and

$$D_i = \frac{1}{3}(\mathbf{u}_i - \bar{\mathbf{u}})^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{u}_i - \bar{\mathbf{u}}) \tag{7c}$$

The average of D_i over all sites is 1.0. If $D_i > 3$ at site i, then the site is considered to be discordant from the rest of a group or region, and two possibilities should be investigated—either there may be an error in the data, or the station may belong to another region. It is also possible that a site with $D_i > 3$ is due to sampling errors.

The D_i value for each site within each of the ten regions was computed and only one site each in regions 3, 6, 8, and 9 were identified with values of $D > 3$: D = 4.84, 3.67, 4.70, and 3.22, respectively. Without further information regarding these sites with D > 3, we do not intend to remove them from the regions. The results indicate that within each of the regions delineated based on their physiography and climatology, the nature of flood flows tends to be similar.

Identification results based on the L-moment diagrams

To avoid errors due to short record length, only sites with observation records of longer than 20 years are used for analysis. Figures 2 through 4 compare the fitness of the theoretical probability distributions to the observations in each of the regions. Figure 2 illustrates that none of the 2-parameter distributions approximates the observations well, since their theoretical $\tau_3 \sim \tau_2$ lines do not pass across the center of the cluster of the estimates of $\tau_3 \sim \tau_2$. For the 3-parameter distributions, it is hard to judge which one best fits the observations (Figs. 3 and 4) in each of the regions. It is evident that it is almost impossible to select the best-fit distribution type based solely on the L-moment ratio diagrams.

To discern the distribution type that best fits the observations in each of the regions, the average weighted distance values for these distributions were computed using Eq. (5), and are listed in Table 1. The rank orders of distribution types to fit the observations according to the average weighted distance values are also presented in Table 1. The average weighted distance criterion provides a more accurate assessment among these distribution types. It can be seen that the generalised extreme value distribution best corresponds to observations in regions 1, 2, 3, 6, and 8. The best-fit distribution and the corresponding average weighted distance value are shaded in gray in Table 1. With consideration of sampling errors, the distributions with the average weighted distance values that are close to that of the generalised extreme value are also considered as potential candidates, which are marked in bold italic format in Table 1. The 3-parameter lognormal or log-Pearson type III distribution provides an alternative in regions 1, 2, 3, 6, and 8 depending on the regions. The log-Pearson type III corresponds to the observations best in regions 4 and 10, with the Pearson type III and generalised extreme value as potential candidates, respectively. In regions 5, 7, and 9, the 3-parameter Weibull, 3-parameter lognormal, and generalised logistic distributions fit the observations best, respectively, with the Pearson type III, log-Pearson type III and generalised extreme value distributions as potential candidates, respectively. The results confirm the generalised extreme value distribution recommended by Pilon and Adamowski (1992) and Gringas and Adamowski (1992) for flood flows in Nova Scotia and New Brunswick, which are located in region 8.

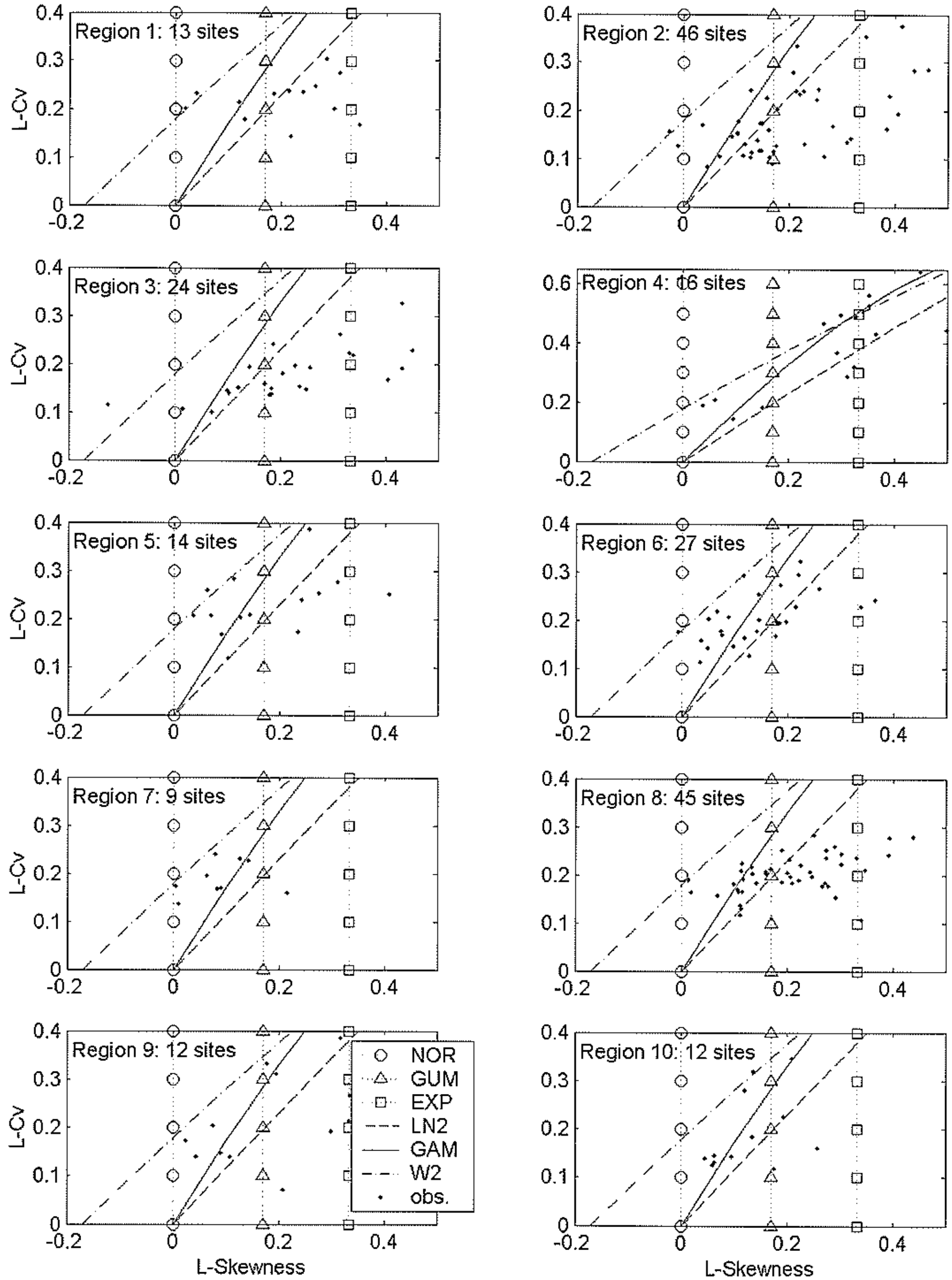

Figure 2 – L-moment ratio diagrams of L-cv versus L-skewness for annual flood flows for each of the climatic regions.

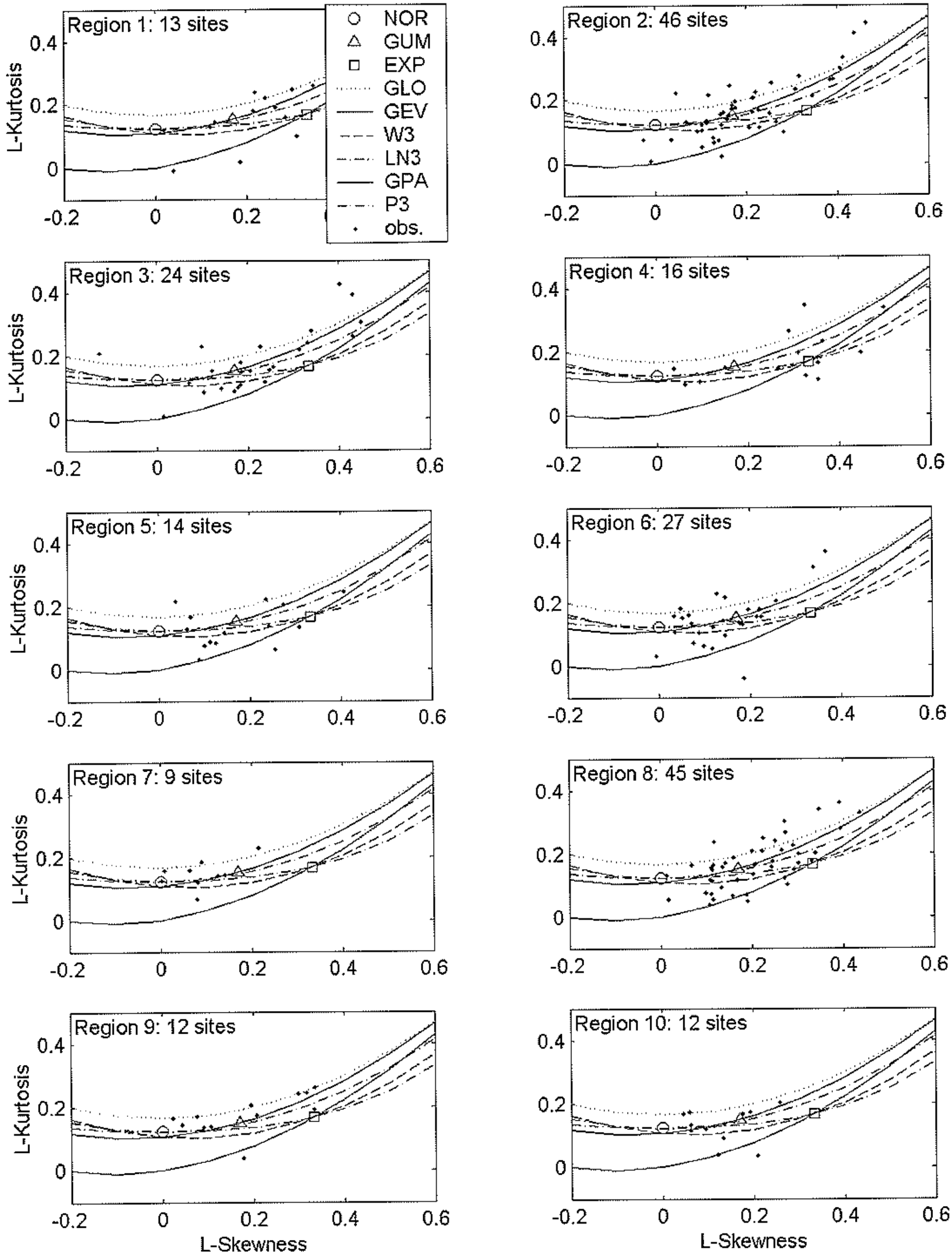

Figure 3 – L-moment ratio diagrams of L-kurtosis versus L-skewness for annual flood flows for each of the climatic regions.

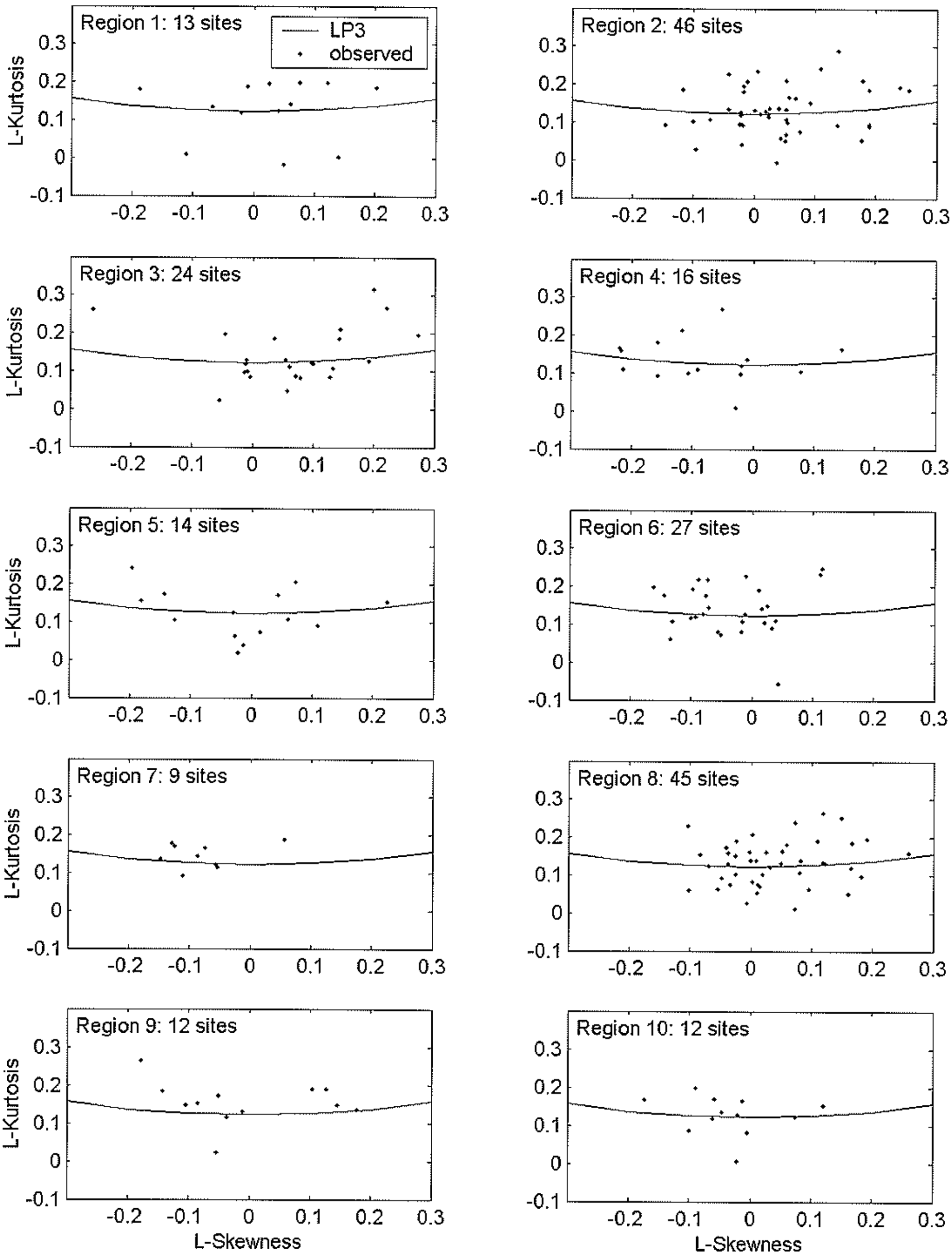

Figure 4 – L-moment ratio diagrams of L-kurtosis versus L-skewness for the log Pearson type III distribution for annual flood flows for each of the climatic regions.

Table 1 Average weighted distance values for various probability distributions

SUB-CLIMATIC REGIONS

Region	No. of sites	AWD values LN2	GAM	W2	GLO	GEV	W3	LN3	GPA	P3	LP3
1	13	0.0839	0.1386	0.1528	0.0586	0.0387	0.0825	*0.0427*	0.0857	0.0502	0.0513
2	46	0.0862	0.1525	0.1967	0.0578	0.0416	0.0957	0.0473	0.0897	0.0601	*0.0464*
3	24	0.1111	0.1935	0.2156	0.0617	0.0462	0.1003	*0.0477*	0.0820	0.0611	0.0481
4	16	0.0837	0.0686	0.0894	0.0740	0.0497	0.1061	0.0421	0.0726	*0.0390*	0.0380
5	14	0.1037	0.1274	0.1207	0.0733	0.0554	0.0463	0.0492	0.0728	*0.0464*	0.0473
6	27	0.0689	0.0855	0.1150	0.0595	0.0382	0.0525	*0.0386*	0.0950	0.0421	0.0418
7	9	0.1115	0.0891	0.0698	0.0347	0.0328	0.0386	0.0284	0.1203	0.0321	*0.0317*
8	45	0.0807	0.1486	0.1794	0.0523	0.0411	0.0681	*0.0445*	0.0867	0.0558	0.0481
9	12	0.1025	0.1272	0.1457	0.0281	*0.0347*	0.0480	0.0363	0.1055	0.0498	0.0470
10	12	0.0795	0.0789	0.1221	0.0510	*0.0362*	0.0454	0.0369	0.0952	0.0402	0.0358

Ranks from best-fit (left) to worst-fit (right)

Region										
1	GEV	*LN3*	P3	LP3	GLO	W3	LN2	GPA	GAM	W2
2	GEV	*LP3*	LN3	GLO	P3	LN2	GPA	W3	GAM	W2
3	GEV	*LN3*	LP3	P3	GLO	GPA	W3	LN2	GAM	W2
4	LP3	*P3*	LN3	GEV	GAM	GPA	GLO	LN2	W2	W3
5	W3	*P3*	LP3	LN3	GEV	GPA	GLO	LN2	W2	GAM
6	GEV	*LN3*	LP3	P3	W3	GLO	LN2	GAM	GPA	W2
7	LN3	*LP3*	P3	GEV	GLO	W3	W2	GAM	LN2	GPA
8	GEV	*LN3*	LP3	GLO	P3	W3	LN2	GPA	GAM	W2
9	GLO	*GEV*	LN3	LP3	W3	P3	LN2	GPA	GAM	W2
10	LP3	*GEV*	LN3	P3	W3	GLO	GAM	LN2	GPA	W2

ENTIRE COUNTRY

Record Length	No. of sites	AWD values LN2	GAM	W2	GLO	GEV	W3	LN3	GPA	P3	LP3
>=20	223	0.0884	0.1310	0.1555	0.0567	0.0420	0.0717	*0.0432*	0.0895	0.0511	*0.0449*
>=30	160	0.0887	0.1308	0.1567	0.0542	0.0396	0.0734	*0.0412*	0.0909	0.0494	*0.0436*
>=40	77	0.0841	0.1220	0.1519	0.0507	0.0340	0.0764	*0.0360*	0.0965	0.0456	*0.0380*
>=50	46	0.0786	0.1094	0.1423	0.0503	0.0301	0.0615	*0.0329*	0.0974	0.0410	*0.0352*

Ranks from best-fit (left) to worst-fit (right)

Record Length										
>=20	GEV	*LN3*	*LP3*	P3	GLO	W3	LN2	GPA	GAM	W2
>=30	GEV	*LN3*	*LP3*	P3	GLO	W3	LN2	GPA	GAM	W2
>=40	GEV	*LN3*	*LP3*	P3	GLO	W3	LN2	GPA	GAM	W2
>=50	GEV	*LN3*	*LP3*	P3	GLO	W3	LN2	GPA	GAM	W2

AWD - average weighted distance, NOR – normal, EXP – exponential, GAM – gamma, P3 - Pearson type III, LP3 - log-Pearson type III, LN2 - 2-parameter lognormal, LN3 - 3-parameter lognormal, GEV - generalized extreme value, GUM – Gumbel, W2 - 2-parameter Weibull, W3 - 3-parameter Weibull, GPA - generalized Pareto, GLO - generalized logistic

It is evident that the statistical properties of annual flood flows in a large portion of Canada may be represented by the generalised extreme value distribution. However, some differences were found in the distribution types of annual flood flows across the country, which may stem from the major differences in climate and physiography among the regions. Some differences may simply come from sampling errors due to the sparse number of sites in some regions, but which is not clear here.

Canada lies north of latitude 45°N, with long winters generally lasting about four or more months. About one-third to one-half of its precipitation is snow, which accumulates within a basin during the winter and is released from the basin as snowmelt runoff during spring. Spring often is the peak flow season because of this contribution of spring snowmelt. The snowmelt alone or the snowmelt plus rain-on-snow generates the greatest annual flood, both in its flood peak and volume (Watt, 1989). Because of flows generated by snowmelt, it is possible that flood peak flows may largely have the same type of distribution across the country. To investigate this, we pooled the sample data across the country for analysis. Sample data with record lengths of $n \geq 20$, 30, 40, and 50 are used, to take into account the effect of the length of record on the choice of a probability distribution. The results are plotted in Figures 5 through 7 and the average weighted distance values are listed in Table 1. Consistent results were obtained among the four sets of samples of varying record length, i.e., the generalised extreme value distribution provides a best-fit to the observations across the entire country, with the 3-parameter lognormal and log-Pearson Type III as potential candidates.

Identification results based on the *Z* statistic

The Z statistics for six 3-parameter probability distributions: generalised logistic, generalised extreme value, 3-parameter lognormal, Pearson type III, generalised Pareto, and log-Pearson type III were

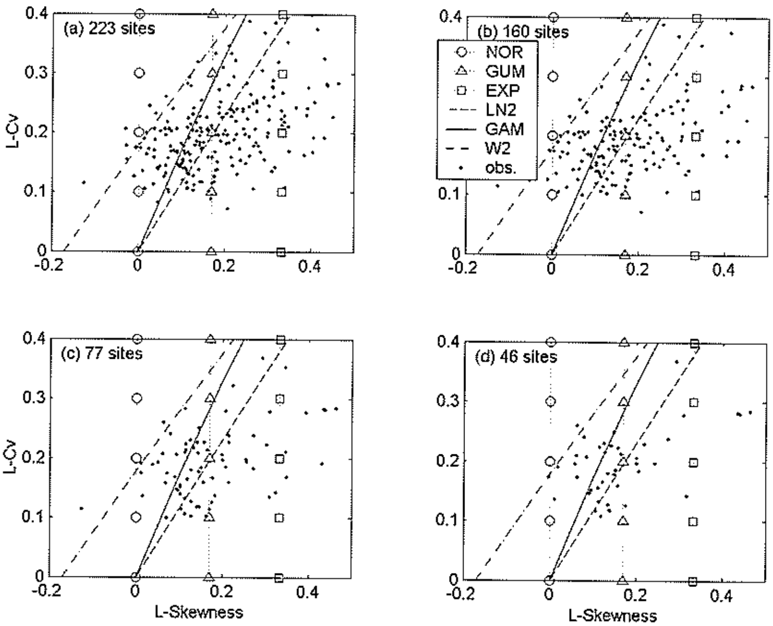

Figure 5 – L-moment ratio diagrams of L-CV versus L-skewness for annual peak flows for the entire country: (a) $n \geq 20$; (b) $n \geq 30$; (c) $n \geq 40$; (d) $n \geq 50$.

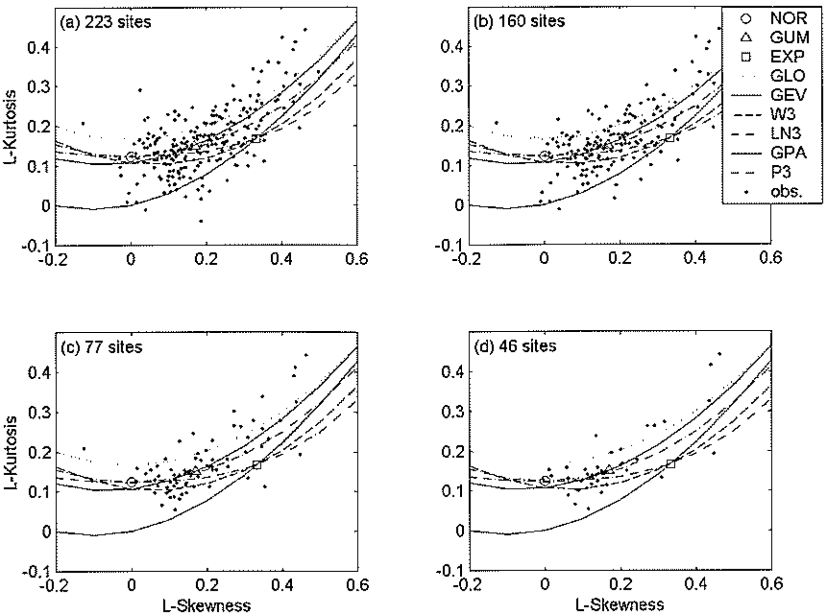

Figure 6 – L-moment ratio diagrams of L-kurtosis versus L-skewness for annual peak flows for the entire country: (a) $n \geq 20$; (b) $n \geq 30$; (c) $n \geq 40$; (d) $n \geq 50$

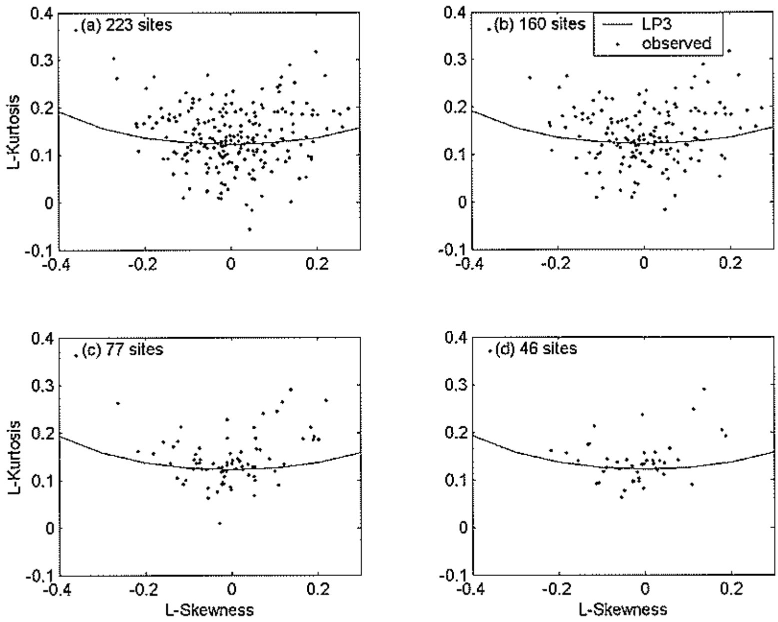

Figure 7 – L-moment ratio diagrams of L-kurtosis versus L-skewness for the suitability of the log Pearson type III distribution for annual peak flows for the entire country: (a) $n \geq 20$; (b) $n \geq 30$; (c) $n \geq 40$; (d) $n \geq 50$

computed using the Fortran Package by Hosking (2000). For the log-Pearson type III distribution, the Z statistic value was obtained by fitting the logarithm of the original sample data to the Pearson type III distribution. The absolute values of Z statistics for each of the different climatic regions and for the entire country are presented in Table 2. At the

Table 2 Probability distributions of annual maximum 1-day flow

Region	**Number of sites**	**Z Statistic**					
		GLO	**GEV**	**LN3**	**GPA**	**P3**	**LP3**
1	13	2.45	*0.31* *	0.10 *	4.51	1.03 *	0.59 *
2	46	2.09	1.78	2.66	10.61	4.52	*1.78*
3	24	1.02 *	*1.15* *	1.73	6.17	2.89	1.35 *
4	16	2.84	0.74 *	*0.06* *	4.30	1.54 *	0.04 *
5	14	3.07	0.78 *	0.47 *	4.25	*0.35* *	0.17 *
6	27	3.62	0.36 *	*0.43* *	8.70	1.27 *	1.23 *
7	9	1.64 *	1.60 *	1.17 *	7.91	1.38 *	*1.52* *
8	45	2.5	1.23 *	*2.17*	9.81	4.11	2.85
9	12	0.55 *	*1.19* *	1.42 *	5.01	2.02	1.82
10	12	2.12	0.02 *	*0.04* *	4.34	0.34 *	0.41 *
		Ranks from best-fit (left) to worst-fit (right)					
1		LN3*	*GEV**	LP3*	P3*	GLO	GPA
2		GEV	*LP3*	GLO	LN3	P3	GPA
3		GLO*	*GEV**	LP3*	LN3	P3	GPA
4		LP3*	*LN3**	GEV*	P3*	GLO	GPA
5		LP3*	*P3**	LN3*	GEV*	GLO	GPA
6		GEV*	*LN3**	LP3*	P3*	GLO	GPA
7		LN3*	*P3**	LP3*	GEV*	GLO*	GPA
8		GEV*	*LN3*	GLO	LP3	P3	GPA
9		GLO*	*GEV**	LN3*	LP3	P3	GPA
10		GEV*	*LN3**	P3*	LP3*	GLO	GPA

ENTIRE COUNTRY

Region	**Number of sites**	**Z Statistic**					
		GLO	**GEV**	**LN3**	**GPA**	**P3**	**LP3**
>=20	223	6.59	2.41	*3.90*	22.46	7.45	4.34
>=30	160	5.45	2.51	*3.78*	20.21	6.85	4.75
>=40	77	*3.57*	2.92	3.82	17.23	6.16	4.38
>=50	46	3.52	2.22	*2.76*	14.65	4.50	3.48
		Ranks from best-fit (left) to worst-fit (right)					
>=20		GEV	*LN3*	*LP3*	GLO	P3	GPA
>=30		GEV	*LN3*	*LP3*	GLO	P3	GPA
>=40		GEV	*GLO*	*LN3*	LP3	P3	GPA
>=50		GEV	*LN3*	*LP3*	GLO	P3	GPA

NOR – normal, EXP – exponential, GAM – gamma, P3 - Pearson type III, LP3 - log-Pearson type III, LN2 - 2-parameter lognormal, LN3 - 3-parameter lognormal, GEV - generalized extreme value, GUM – Gumbel, W2 - 2-parameter Weibull, W3 - 3-parameter Weibull, GPA - generalized Pareto, GLO - generalized logistic significance level of 0.10, the acceptable probability distribution types, i.e., $Z \leq 1.64$, are marked by an asterisk. By comparing these results with those in Table 1 based on the average weighted distance, it can be seen that the possible distribution types for each of the 10 regions and the entire country identified by the two approaches are almost identical. This further confirms the suitability of the selected distributions for representing peak flows in Canada.

Overall, on the basis of the results using the above two methods, the generalised extreme value distribution is considered to be an acceptable distribution type for representing Canadian flood peak flows, with the 3-parameter lognormal and log-Pearson type III as potential candidates. However, differences in climate and physiography within different regions may also be considered in the selection of the distribution type of flood peak flows, as presented in Tables 1 and 2.

Conclusions

The L-moment ratio diagrams, the average weighted distance criterion, and the Z statistic were applied to identify the distribution type of annual flood flows (annual maximum daily streamflows) in Canada. The different approaches produced almost the same results, which demonstrated that the generalised extreme value, 3-parameter lognormal, and log-Pearson type III probability distributions are all reasonable distribution types for representing the statistical properties of flood peak flows in Canada. Nevertheless, the climatological and physiographical differences in different climatic regions should also be taken into account in the selection of distribution types of flood peak flows, as evidenced in Tables 1 and 2. The identified distribution types might be considered as a baseline for Canadian hydrological engineers to determine the probability distribution type for describing the statistics of flood peak flows in their engineering practices for flood prevention , especially in watersheds without gauging stations.

Acknowledgements

This work was completed while the first author worked at Environment Canada. The authors thanks Paul Pilon, Paul Ford, and Paul Campbell for their kind support. The authors would like to thank the anonymous reviewers for their constructive comments, improving the quality of the paper.

References

Ben-Zvi, A.; Azmon, B. 1997: Joint use of L-moment diagram and goodness-of-fit test: a case study of diverse series. *Journal of Hydrology 198*: 245-259.

Cunnane, C. 1987: Review of statistical models for flood frequency estimation. In: V.P. Singh (ed.), *Hydrologic frequency modelling*, Reidel, Dordrecht, The Netherlands, pp. 49-95.

Environment Canada 1999: Establishment of the Reference Hydrometric Basin Network (RHBN) for Canada (RHBN database in CD).

Gringas, D.; Adamowski, K. 1992: Coupling of nonparametric frequency and L-moment analysis for mixed distribution identification. *Water Resources Bulletin 28(2)*: 263-272.

GREHYS (Groupe de recherche en hydrologie statistique) 1996: Presentation and review of some methods for regional frequency analysis. *Journal of Hydrology 186*: 63-84.

Gullett, D.W. 1992: Development of an historical Canadian climate database for temperature. In: G.W. Kite and K.D. Harvey (eds.), *Using Hydrometric Data to Detect and Monitor Climatic Change*, Proceedings of NHRI Workshop No. 8., National Hydrology Research Institute, Saskatoon, Saskatchewan, pp. 27-31.

Hare, F.K.; Thomas, M.K. 1979: *Climate Canada* (2nd edition). John Wiley and Sons Canada Limited.

Hosking, J.R.M. 1990: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B, 52*: 105-124.

Hosking, J.R.M.; Wallis, J.R. 1993: Some statistics useful in regional frequency analysis. *Water Resources Research 29(2)*: 271-281.

Hosking, J.R.M., Wallis, J.R. 1997: *Regional Frequency Analysis: An Approach Based on L-Moments.* Cambridge University Press, Cambridge, U.K.

Hosking, J.R.M. 2000: *Research Report: Fortran routines for use with the method of L-moments,* Version 3.03. RC 20525(90933), IBM Research Division, NY.

Karim, M.A.: Chowdhury, J.U. 1993: A comparison of five distributions used in flood frequency analysis in Bangladesh. In: Proceedings of the International Conference on Hydrology & Water Resources.

Kroll, C.N.; Vogel, R.M. 2002: Probability distribution of low streamflow series in the United States. *ASCE Journal of Hydrologic Engineering 7(2)*: 137-146.

Nathan, R.J.; Weinmann, P.E. 1991: Application of at-site and regional flood frequency analysis. Proceedings of the International Hydrological & Water Resources Symposium, 769-774.

Pearson, C.P. 1991: New Zealand regional flood frequency analysis using L-moments. *Journal of Hydrology (NZ) 30(2)*: 53-64.

Peel, M.C.; Wang, Q.J.; Vogel, R.; McMahon, T.A. 2001: The utility of L-moment ratio diagrams for selecting a regional probability distribution. *Hydrological Science Journal 46(1)*: 147-155.

Pilon, P.J.; Adamowski, K. 1992: The value of regional information to flood frequency analysis using the method of L-moments. *Canadian Journal of Civil Engineering 19*: 137-147.

Rao, A.R.; Hamed, K.H. 1994: Frequency analysis of upper Cauvery flood data by L-moments. *Water Resources Management 8*: 183-201.

Vogel, R.M.; Wilson, I. 1996: Probability distribution of annual maximum, mean, and minimum streamflows in the United States. *ASCE Journal of Hydrologic Engineering 1(2)*: 69-76.

Wallis, J.R. 1988: Catastrophes, computing and containment: living with our restless habitat. *Speculation in Science and Technology 11(4)*: 295-324.

Watt, W.E. 1989: *Hydrology of Floods in Canada: A Guide to Planning and Design.* National Research Council of Canada.

Manuscript received 21 July2003; revised; accepted for publication 5 March 2004