

# Application of the Box-Cox Power Transformation in Extreme Value Analysis of Wind Speed

H. P. Hong

*Professor, Dept. of Civil & Environmental Engineering, University of Western Ontario, London, Canada*

**ABSTRACT:** The Gumbel distribution is one of the commonly used models for the extreme wind speed or the squared wind speed. The best choice between fitting the wind speed or squared wind speed is not apparent. To objectively avoid the need for the choice, we apply the Box-Cox power transformation with the Gumbel distribution and the generalized extreme value distribution in this study. The application of the transformation improves the rate of convergence in extreme value analysis. For the numerical example analysis, we consider the wind data recorded at 14 meteorological stations in Canada. Results indicate that the proposed application with the Gumbel distribution gives consistent estimates of the load effects whether the wind speed or squared wind speed are used for the distribution fitting; the application with the generalized extreme value distribution with Box-Cox transform is ineffective for the extreme data considered.

## 1. INTRODUCTION

The Gumbel distribution is often used to fit the extreme wind speed or squared wind speed. Often the choice of whether one should fit the wind speed or squared wind speed is not straight forward. In developing the wind load in the National Building Code of Canada, the annual maximum wind speed is modeled using the Gumbel distribution (NBCC 2010). The Gumbel distribution is also considered by others to model extreme wind speed (e.g., Peterka and Shahid 1998; Frank 2001; Hong et al. 2013, 2014). The use of the Gumbel distribution to fit the annual maximum wind speed is not universally preferred. Cook (1982) (see also Harris 2004; Cook and Harris 2004) suggested the use of the Gumbel distribution to the squared wind speed which is directly proportional to the wind velocity pressure. This consideration is based on partly that the wind speed is approximately a Rayleigh variate and its squared value is an exponential variate. The maxima of Rayleigh variate is Gumbel distributed with a slow convergence; the maxima of an exponential variate is also Gumbel distributed but with a good convergence rate. However, the probability distribution fitting carried out by Simiu et al.

(2001) indicates that the Rayleigh distribution is not necessarily the best fit distribution to the hourly-mean wind speed. There is no agreed guideline on when one should use the Gumbel distribution to fit the annual maximum wind speed or its squared value.

The application of other distributions such as the generalized extreme value distribution (GEVD) to model annual maximum wind speed was also discussed in the literature (Holmes and Moriarty 1999; Kasperski 2002); a review of earlier work on the extreme value analysis of wind speed is presented by Palutikof et al. (1999). One of the problems in the extreme value analysis is the rate of convergence (Coles 2001; De Haan and Ferreira 2006). The slow convergence is of practical important for extreme wind, especially if the hourly-mean wind speed is Weibull distributed, (Rayleigh distribution is a special case of the Weibull distribution). Also, the rate of the convergence of the maxima of Weibull variate,  $x$ , to the Gumbel variate was discussed (Hong 1994a,b; Cook and Harris 2004, Harris 2009) in the context of extreme wind speed. A distribution model was proposed in Hong (1994a) to improve the rate of convergence. This proposed model leads to the double logarithmic of the cumulative probability

distribution of the maxima approaching to a power function of  $x$ . A similar distribution, named general Fisher Tippet Type I (FT1) penultimate distribution, was proposed by Cook and Harris (2004) and Harris (2009). Wadsworth et al. (2010) presented the rate of convergence issue in the context of the extreme value analysis with the application of the Box-Cox power transformation (BCT) (Box and Cox 1964). Their study is presented in the context of measurement scales, and is based on the asymptotic and penultimate theory in statistics. They showed that the parameter used for the BCT controls the rate of convergence (see also Teugels and Vanroelen 2004).

The main objective of the present study is to investigate the usefulness of applying the BCT with the Gumbel distribution and with the GEVD to fit the annual maximum wind speed. The use of the BCT resulted in an additional model parameter for the considered distribution models. These distributions, with an unknown power transformation parameter, are more flexible and could allow the transformed data to be better fitted. Both the Gumbel distribution and GEVD, with and without the transformation are applied to the wind speed records at 14 Canadian meteorological stations to estimate return period value of annual maximum hourly-mean wind speed. Implications of the results are discussed.

## 2. PROBABILISTIC MODELS AND MODEL PARAMETER ESTIMATION

### 2.1. Models with improved rate of convergence

The Gumbel distribution  $F_G(x)$  is (Coles 2001),

$$F_{GV}(x) = \exp(-\exp(-(x-u)/a)) \quad (1)$$

where  $x$  denotes the value of the random variable  $X$ ,  $u$  and  $a$  are the location and scale parameters. The mean of  $X$ ,  $\mu_X$ , and the standard deviation of  $X$ ,  $\sigma_X$ , are  $\mu_X = u + \gamma a$  and  $\sigma_X = a\pi/\sqrt{6}$  where  $\gamma \approx 0.5772$ ; the  $T$ -year return period value of  $X$ ,  $x_T$ , is given by,

$$x_T = u - a \ln(-\ln(1-1/T)). \quad (2)$$

In wind engineering, whether it is more appropriate to use the Gumbel distribution to fit the extreme wind speed or its squared value was debated as mentioned in the introduction section. This debate is focused on the slow convergence of the maxima of the sequence of independent identically distributed Weibull variate, defined by the cumulative distribution  $F_W(x)$ ,

$$F_W(x) = 1 - \exp(-(x/b)^k), \text{ for } x \geq 0 \quad (3)$$

to the Gumbel variable if  $k$  deviates significantly from the unity. For example, for  $k = 2$  which represents the Rayleigh distribution, a slow convergence of  $(F_W(x))^n$  to the Gumbel distribution is expected if the condition that the number of event,  $n$ , tends to infinite cannot be relied on (Hong 1994a, 1994b; Cook and Harris 2004, Harris 2009).

To ameliorate the slow convergence for the Weibull sequence, Hong (1994a) proposed the following distribution,  $F_1(x)$ ,

$$F_1(x) = \exp(-\exp(-(a_1x^{a_1} - a_3x^{-a_4} - a_5))), \text{ for } x \geq 0 \quad (4)$$

where  $a_i > 0$  for  $i = 1$  to 4 are distribution parameters. For the maxima of Weibull variate with known  $k$ ,  $a_2$  in Eq. (4) equals  $k$ . Since the consideration of the term  $a_3x^{-a_4}$  in the model is to ensure that  $F_1(x)$  tends to zero as  $x$  tends to zero (because  $X$  is a non-negative defined random variable), Eq. (4) can be simplified by arbitrarily assigning  $a_3 = 0.01$  and  $a_4 = 1$  resulting in,

$$F_1(x) = \exp(-\exp(-(a_1x^{a_1} - 0.01x^{-1} - a_5))), \quad (5)$$

for  $x \geq 0$ .

This simplification does not affect the estimated  $x_T$  in the upper tail of the distribution if the mean of  $X$  is sufficiently greater than zero. In fact, an application of a further simplified version of Eq. (4) with  $a_3$  equal to zero, denoted by  $G_2(x)$ ,

$$G_2(x) = \exp(-\exp(-(a_1x^{a_1} - a_5))), \text{ for } x \geq 0, \quad (6)$$

was presented in Hong (1994b) for annual extreme wind speed and wave height at offshore locations.

The above mentioned slow convergence problem was also discussed by Cook and Harris (2004) in the context of extreme value analysis of annual maximum wind speed. They proposed the FT1 penultimate distribution, which is the same as Eq. (6) but with different parameterization.

Wadsworth et al. (2010) discussed the extreme value analysis in the context of measurement scales. They indicated that data from the same underlining physical process could naturally be measured on more than one scale and, that different limiting distributions may be appropriate if the transformation between the measurement scales is not linear. They demonstrated the validity of applying extreme value distribution to the transformed data (i.e., in the transformed scale) obtained through the application of the BCT to the data in the original scale. It was shown theoretically that the rate of convergence of penultimate approximations can be improved if the analysis is carried out in the transformed scale. The improvement was also discussed by Teugels and Vanroelen (2004); the application of the BCT together with the Gumbel distribution (minima) was considered by Achcar et al. (1987) for survival time data.

By applying the BCT (Box and Cox 1964), the samples in the transformed  $Y$  scale,  $\{y_i\}$ , is given by

$$y_i = \begin{cases} (x_i^\lambda - 1) / \lambda & \text{for } \lambda \neq 0 \\ \ln(x_i) & \text{for } \lambda = 0 \end{cases}, \quad (7)$$

where the exponent  $\lambda$  is the transformation parameter,  $\{x_i\}$  represents the samples in the original measurement scale  $X$ , and  $y_i$  is a monotonic increasing function of  $x_i$ . If the variable  $Y$  (in the transformed measurement scale) is modeled as a Gumbel variate (see Eq. (1)), the incorporation of the transformation shown in Eq. (7) for a given  $\lambda$  results in the cumulative distribution function,  $F_2(y)$ ,

$$F_2(y) = \exp\left(-\exp\left(-\frac{y-u}{a}\right)\right) \\ = \exp\left(-\exp\left(-\frac{x^\lambda - \lambda u - 1}{\lambda a}\right)\right). \quad (8)$$

This distribution is identical to Eq. (6) but with different parameterization. Comparison of Eqs. (4) and (8) indicates that the functional forms of these distributions are identical except the term  $a_3 x^{-a_4}$  in Eq. (4). This term makes Eq. (4) slightly more complicated, but it ensures that  $F_1(x)$  is a properly defined continuous probability distribution function. For example, without such a term  $F_2(y)$  does not approach to zero as  $x$  tends to zero if  $X$  is a non-negatively distributed random variable and  $\lambda > 0$ . It is noteworthy that  $F_1(x)$  can also be interpreted as the results of the application of the polynomial transformation  $y_i = \lambda_1 (x_i^{\lambda_2} - x_i^{-\lambda_3})$  with the Gumbel distribution but with re-parameterization, where  $\lambda_i > 0$  for  $i = 1$  to 3. This transformation is a monotonic transformation; it is more flexible than the transformation shown in Eq. (7) but with the burden of two additional parameters.

It was mentioned in the introduction that the GEVD is also a popular distribution used to model annual maximum wind speed. This distribution,  $F_{GE}(x)$ , is given by,

$$F_{GE}(x) = \exp\left(-\left(1 - \kappa(x - \xi) / \alpha\right)^{1/\kappa}\right), \text{ for } \kappa \neq 0, \quad (9)$$

where  $\xi$ ,  $\alpha$  and  $\kappa$  are the location, scale and shape parameters. For  $\kappa < 0$ ,  $\xi + \alpha / \kappa \leq x \leq \infty$ , and for  $\kappa > 0$ ,  $-\infty < x < \xi + \alpha / \kappa$ . Eq. (9) tends to the Gumbel distribution if  $\kappa$  tends to 0. The  $T$ -year return period value  $x_T$  for the GEVD is given by,

$$x_T = \xi - (\alpha / \kappa) \times \left(1 - (-\ln(1 - 1/T))^\kappa\right). \quad (20)$$

Again, to potentially improve the distribution fitting, the extreme analysis can be carried out in the transformed  $Y$  scale defined by Eq. (7) rather than the original  $X$  scale (Wadsworth et al. 2010), where  $Y$  is considered

to follow the generalized extreme value distribution. This resulted in,

$$F_3(y) = \exp\left(-\left(1 - \kappa(x^\lambda - \lambda\xi - 1) / (\lambda\alpha)\right)^{1/\kappa}\right), \quad (31)$$

for  $k \neq 0$ , which has four model parameters  $\xi$ ,  $\alpha$ ,  $\kappa$  and  $\lambda$ .

Based on the asymptotic and penultimate theory, Wadsworth et al. (2010) demonstrated that the best rate of convergence for the Weibull sequence is achieved if  $\lambda$  is taken equal to the shape parameter  $k$  (see Eq. (4)). They also recommended re-parameterization of the model parameters ( $\xi$ ,  $\alpha$ ,  $\kappa$ ,  $\lambda$ ) so the inference in a Bayesian setting is more effective. The re-parameterization is done based on  $\lambda$  and the parameters that are for the GEVD applied to the untransformed samples of  $X$ .

## 2.2. Estimation of model parameters

There are several methods, including the method of moments, the method of maximum likelihood (MML), generalized least-squares method, and the method of L-moments that can be used to estimate the distribution parameters for the Gumbel distribution and GEVD. The generalized least-squares method and MML are preferred for the Gumbel distribution based on the minimum bias and minimum root-mean-square-error (RMSE) (Hong et al. 2013). For the GEVD, the MML provides the estimator with the minimum RMSE, although the method can give unrealistic predictions if the sample size is small (Hosking 1990, Martin and Stedinger 2000).

Since the analytical equations for the moments or  $L$ -moments of  $X$  (i.e., on the original  $X$  scale) are not readily available for the distributions shown in Eqs. (4), (5), (8) and (11), the application of the method of moments and the method of L-moments is not straight forward. A simple procedure to estimate the distribution parameters for model shown in Eqs. (4) and (5) is to carry out regression analysis for the data plotted in the Gumbel probability. In such a case, a plotting position needs to be selected. For example, the Gringorten formula for the  $i$ -th rank

of the samples,  $(i - 0.44) / (n + 0.12)$ , can be used (Cunnane 1978), where  $n$  represents the sample size. The plotting positions for the generalized extreme value distribution were presented in the literature, including Arnell et al. (1986). However, they are less amenable to use because of their dependency on the distribution parameter or skewness. Therefore, for simplicity the Gringorten plotting position is considered for the numerical examples presented in the following section; the model parameters are estimated by minimizing  $\varepsilon$  (i.e., least-squares (LS) method) defined as,

$$\varepsilon = \sum_{i=1}^n (y_{oi} - y_{pi})^2, \quad (42)$$

where  $y_{oi}$  denotes the sample value in the original ( $X$ ) or transformed ( $Y$ ) space,  $y_{pi} = F^{-1}((i - 0.44) / (n + 0.12))$ , and  $F^{-1}(\bullet)$  denote the inverse function of the considered cumulative distribution model  $F(\bullet)$  (i.e., Eqs. (1), (5), (8), (9) or (11)).

If the MML is used for the distribution fitting, the required log-likelihood functions for the Gumbel distribution and GEVD are well known and, those for  $F_1(x)$ ,  $F_2(y)$  and  $F_3(y)$ , denoted by  $\mathcal{L}_1$ ,  $\mathcal{L}_2$  and  $\mathcal{L}_3$  respectively, are presented in Table 1. The log-likelihood function includes the Jacobian of the BCT if it is written in the original measurement scale  $X$ . The maximization of  $\mathcal{L}_1$ ,  $\mathcal{L}_2$  and  $\mathcal{L}_3$  can be carried out using an available nonlinear constrained optimization methods (Bertsekas 1999), including those implemented in the Solver in Microsoft Excel.

## 3. ESTIMATION OF ANNUAL MAXIMUM WIND SPEED

For the application, the wind records at 14 meteorological stations across Canada are considered, and the fitting by using the Gumbel distribution, the GEVD,  $F_1(x)$  shown in Eq. (5),  $F_2(y)$  shown in Eq. (8), and  $F_3(y)$  shown in Eq. (11) is carried out. Each of the 14 stations represents a location in the airport of a capital

city of the Canadian provinces and territories or the national capital (Hong et al. 2013, 2014). The annual maximum wind speed for each station obtained was extracted from the Environment Canada (EC) HLY01 database. The wind speed measurements at each station were adjusted for anemometer height using the power law (NRCC 2010), and for exposure or roughness considering surrounding terrain conditions (ESDU 2002, Mara et al. 2013). It is considered that the adjusted wind speeds are representative of the annual maximum hourly-mean wind speed at 10 m height for open country terrain.

Table 1. Log-likelihood of the considered models shown in Eqs. (4), (8) and (11)

Dist.	Log-likelihood
$F_1(x)$	$\mathcal{L}_1 = \sum_{i=1}^n \left( \ln \left( a_1 a_2 x_i^{a_2-1} + a_3 a_4 x_i^{-a_2-1} \right) - z_i - \exp(-z_i) \right)$ <p>where <math>z_i = a_1 x_i^{a_2} - a_3 x_i^{-a_2} - a_4</math>. For Eq. (5), <math>a_3 = 0.01</math> and <math>a_4 = 1</math> are assigned.</p>
$F_2(y)$	$\mathcal{L}_2 = -\sum_{i=1}^n \left( \ln a + z_i + \exp(-z_i) \right),$ <p>where <math>z_i = (y_i - u) / a</math>; or</p> $\mathcal{L}_2 = \sum_{i=1}^n \left( \ln(x_i^{\lambda-1} / a) - z_i - \exp(-z_i) \right)$ <p>where <math>z_i = (x_i^\lambda - \lambda u - 1) / (\lambda a)</math>.</p>
$F_3(y)$	$\mathcal{L}_3 = \sum_{i=1}^n \left( -\ln a + (1/k - 1) \ln(z_i) - z_i^{1/k} \right),$ <p>where <math>z_i = 1 - k(y_i - u) / a</math>; or</p> $\mathcal{L}_3 = \sum_{i=1}^n \left( \ln(x_i^{\lambda-1} / a) + (1/k - 1) \ln(z_i) - z_i^{1/k} \right),$ <p>where <math>z_i = 1 - k(x_i^\lambda - \lambda u - 1) / (\lambda a)</math></p>

The intentions of the distribution fitting analysis presented below are to illustrate that the use of the BCT can lead to consistent estimate of return period value of wind speed whether the wind speed or squared wind speed are used for distribution fitting. It also serves to identify potential trends on the  $\lambda$  value for  $F_2(y)$  and  $F_3(y)$  or on the  $a_2$  value for  $F_1(x)$ ;

First, the distribution fitting is carried out by using the wind speed data, and the return period value is estimated and denoted by  $x_{T,1}$ , where  $T$  represents the return period and the subscript 1 indicates that the analysis is performed based on the wind speed data. The obtained  $x_{T,1}$  is shown in Table 2 and the model parameters controlling the shape of the distributions or the rate of convergence (i.e.,  $a_2$  for Eq. (5),  $\lambda$  for Eq. (8) and  $\kappa$  for Eq.(9)) are presented in Table 3. Although results are obtained for all 14 stations, results for five stations are presented in Tables 2 and 3 and in the remaining tables to save space.

The results indicate that there are many empty cells associated with the application of GEVD with the BCT. This is because no convergence was achieved in finding the distribution model parameters. If distribution model parameters are found, the application of this model has the tendency of lowering the return period values as compared to the ones obtained by directly applying the GEVD. Therefore, the application the BCT with the GEVD by using the LS method or the MML is considered ineffective for the considered extreme wind data, and it will not be considered further. Moreover, the use of the Bayesian methodology through the Markov chain Monte Carlo procedure to assessing the distribution parameters (Wadsworth et al. 2010) is beyond the scope of this study.

For a given distribution, in general, the use of different fitting methods can leads to different  $x_{T,1}$ . The differences are smallest for the Gumbel model because it is less flexible than other considered models; the difference increase as  $T$  increases. The use of Eq. (5) and Eq. (8) leads to almost identical results in almost all cases. For a given station,  $a_2$  and  $\lambda$  are greater than unity if  $\kappa$  for the GEVD (without considering the BCT) is away from zero and greater. This is important since the models shown in Eqs. (5) and (8) for  $a_2 > 1$  or  $\lambda > 1$  have a light-tailed behavior as compared to the Gumbel model and without an upper bound, while the GEVD with  $\kappa > 0$  has an upper bound of  $\xi + \alpha / \kappa$ .

Table 2. Selected Canadian meteorological stations and estimated return period values of hourly-mean wind speed (km/hour). Analysis is carried out based on wind speed by minimizing  $\varepsilon$  shown in Eq. (12) (denoted as LS in the table) and by using the MML.

Station	T (years)	Gumbel, Eq. (1)		Eq. (5)		Eq. (8)		GEVD, Eq. (9)		Eq. (11)	
		LS	MML	LS	MML	LS	MML	LS	MML	LS	MML
Victoria	50	80.5	84.0	78.4	78.4	78.3	78.0	76.9	75.5	76.9	---
	500	94.1	99.9	87.2	87.2	87.1	86.3	82.3	80.5	82.1	---
	1000	98.2	104.7	89.6	89.6	89.5	88.5	83.5	81.6	83.2	---
Edmonton	50	81.5	81.3	82.1	80.5	82.1	80.5	82.1	79.4	---	80.5
	500	95.0	94.7	97.6	92.6	97.6	92.6	97.8	91.4	---	93.1
	1000	99.1	98.8	102.4	96.2	102.4	96.2	102.7	94.9	---	96.9
Toronto	50	95.8	98.3	94.3	94.4	94.3	94.5	92.7	91.1	92.5	---
	500	109.0	113.1	104.0	104.1	104.0	104.2	98.8	96.7	96.9	---
	1000	112.9	117.6	106.7	106.7	106.6	106.8	100.1	98.0	97.7	---
Quebec	50	88.0	91.0	86.3	86.5	86.4	86.4	84.3	82.5	83.3	---
	500	101.5	106.4	95.9	96.0	96.0	95.7	89.4	87.1	85.2	---
	1000	105.5	111.0	98.5	98.6	98.7	98.2	90.4	88.0	85.4	---
St. John's	50	121.1	119.3	123.7	120.4	125.8	120.4	126.4	119.6	---	---
	500	141.3	138.5	152.1	141.4	165.6	141.4	167.4	141.0	---	---
	1000	147.3	144.2	161.8	147.9	181.9	147.9	183.5	147.7	---	---

Table 3. Parameters controlling the shape of the distributions (analysis is carried out based on wind speed. U.B. represents the upper bound if the GEVD is used).

Station	$a_2$ for Eq. (5)		$\lambda$ for Eq. (8)		$\kappa$ for GEVD, Eq. (9)			
	LS	MML	LS	MML	LS		MML	
					$\kappa$	U.B.	$\kappa$	U.B.
Victoria	2.68	2.68	2.64	2.90	0.23	89.3	0.24	87.4
Edmonton	0.60	1.27	0.60	1.27	-0.04	$\infty$	0.02	298.5
Toronto	2.44	2.49	2.45	2.47	0.19	108.4	0.20	106.1
Quebec	2.41	2.50	2.41	2.57	0.24	95.3	0.26	92.7
St. John's	0.08	0.76	-0.67	0.76	-0.18	$\infty$	-0.02	$\infty$

In some cases could be unrealistically low in some cases. For 10 out of 14 considered stations,  $a_2$  and  $\lambda$  are within 1.5 to 2.9, indicating a slow convergence to the Gumbel distribution. The slow convergence observed is consistent with that shown in Hong (1994a, b), Cook and Harris (2004), and Harris (2009). For these 10 cases, the return period values predicted by Eqs. (5) or (8) are smaller than those predicted by the Gumbel model. The overestimation by using the Gumbel model is 4%, 10% and 12% for  $T$  equal to 50, 500 and 1000 years if the least-squares method is used. These values become 9%, 17% and 19% if the MML is used.

To further appreciate the quality of fit, plots of the fitted distributions based on MML by

using the Gumbel model, Eq. (5), Eq. (8) and GEVD are presented in Figure 1. The figure reinforces that Eqs. (5) and (8) provide similar fit, that the GEVD may be associated with an unrealistically low upper bound, and that the differences between the fitted distributions in the upper tail region can be very significant.

The distribution fitting for the wind speed data is repeated for the squared wind speed data by considering the models shown in Eqs. (1), (5), (8) and (9). In this case the return period value of wind speed,  $x_{T,2}$ , is calculated by taking the square root of the estimated return period value of squared wind speed, where subscript 2 emphasizes that the estimates are obtained by using squared wind speed as data (i.e., using the

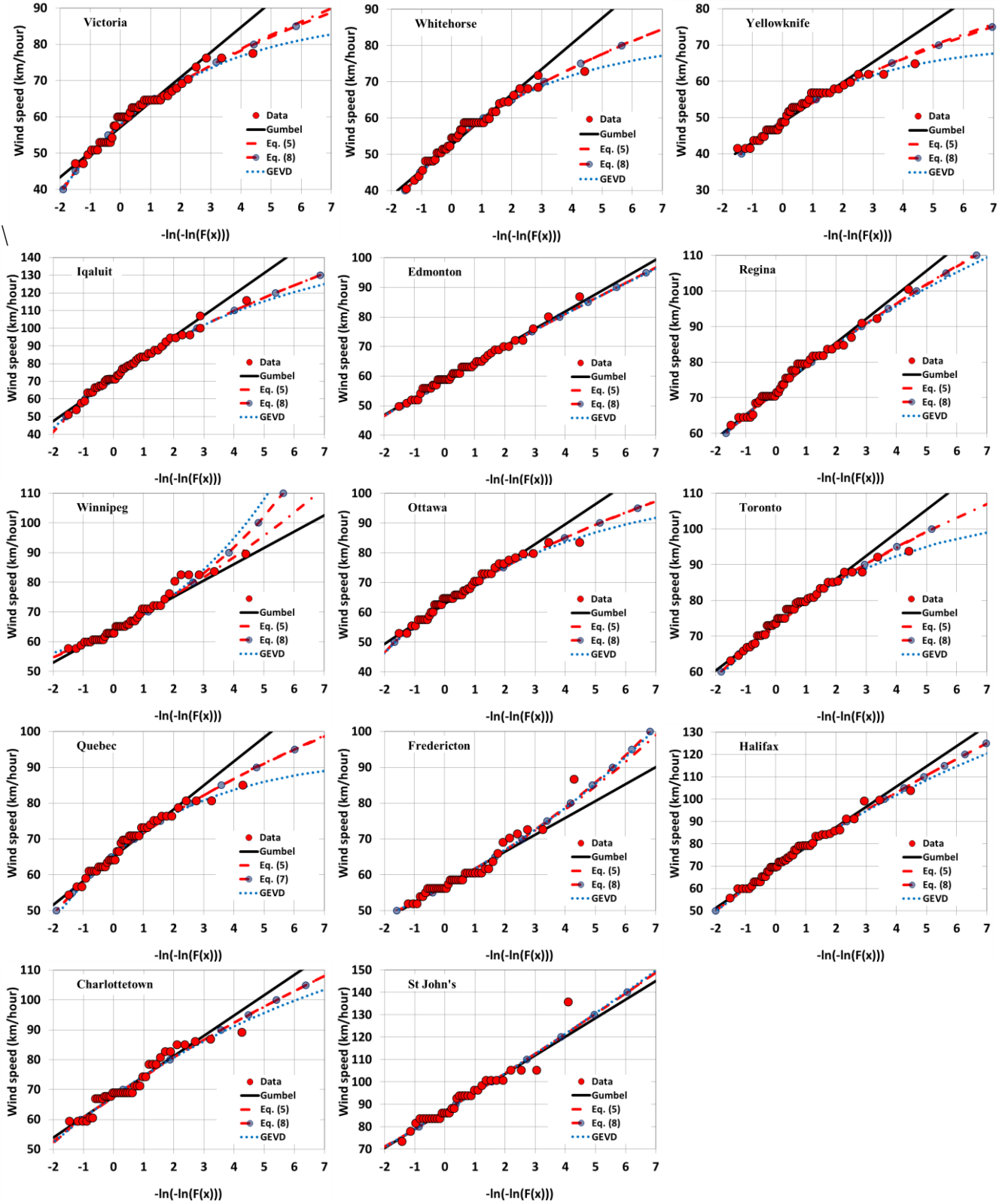


Figure 1. Fitted probability distributions using the MML for 14 Canadian meteorological stations.

wind velocity pressure as the measurement scale). The relative difference  $\delta = ((x_{T,1} - x_{T,2})/x_{T,2})$  between the estimated return period values by using the wind speed data and the squared wind speed data are shown in Figure 2 and Table 4 for  $T$  equal to 50, 500 and 100 years. The fitted parameters controlling the shape of the distributions are presented in Tables 5.

Figure 2 indicates that relative difference  $\delta$  increases as  $T$  increases, especially if the Gumbel distribution is applied. However, if Eqs. (5) and (8) are used, the relative difference  $\delta$  is always small. This emphasizes the adequacy of using Eqs. (5) and (8) in model the extreme wind speed or squared wind speed data since the estimated return period value is insensitive to the choice of

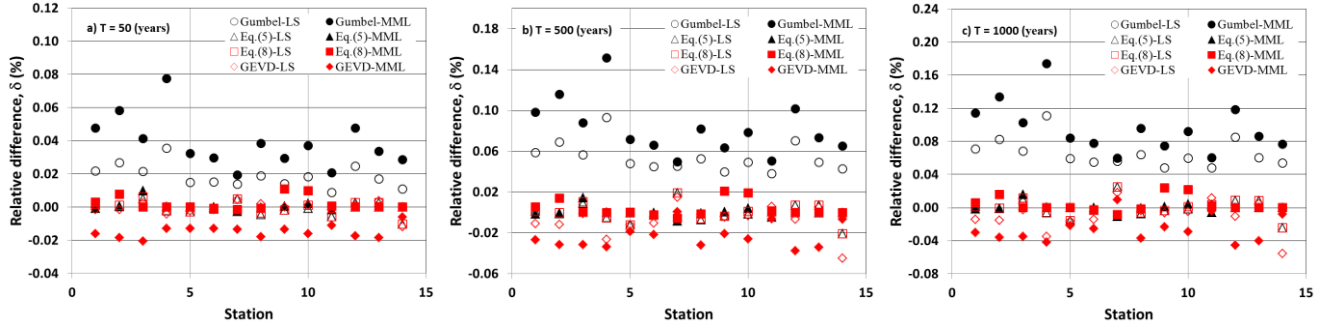


Figure 2. Relative difference,  $\delta = ((x_{T,1} - x_{T,2})/x_{T,2})$ , in estimated return period value by using wind speed data and squared wind speed data (the order of the stations is the same as that listed in Table 2).

Table 4. Statistics of the relative differences  $\delta$  (i.e.,  $\delta = ((x_{T,1} - x_{T,2})/x_{T,2})$ ) for different return period  $T$ , and combinations of the probabilistic models and distribution fitting methods.

$T$	Parameter	Gumbel, Eq. (1) (%)		Eq. (5) (%)		Eq. (8) (%)		GEVD, Eq. (9) (%)	
		LS	MML	LS	MML	LS	MML	LS	MML
50	Maximum	3.5	7.7	0.5	1.0	0.6	1.1	0.4	-0.6
	Minimum	0.9	1.9	-1.0	-0.3	-1.0	-0.2	-1.2	-2.1
	Mean	1.9	3.9	-0.1	0.1	-0.1	0.2	-0.1	-1.5
	Standard deviation	0.7	1.4	0.4	0.3	0.4	0.4	0.4	0.4
500	Maximum	9.3	15.1	1.9	1.4	1.9	2.1	1.5	0.1
	Minimum	3.8	5.0	-2.1	-0.9	-2.1	-0.6	-4.5	-3.7
	Mean	5.4	8.2	-0.1	0.0	0.0	0.4	-0.8	-2.3
	Standard deviation	1.4	2.6	0.9	0.5	0.9	0.8	1.4%	1.1
1000	Maximum	11.1	17.4	2.5	1.6	2.5	2.4	1.8	1.0
	Minimum	4.8	6.0	-1.6	-1.1	-1.6	-0.8	-3.5	-4.5
	Mean	6.6	9.6	0.0	0.0	0.0	0.4	-1.0	-2.6
	Standard deviation	1.6	2.9	1.1	0.5	1.1	0.9	1.7	1.5

Table 5. Parameters controlling the shape of the distributions (analysis is carried out based on squared wind speed. U.B. represents the upper bound if the GEVD is used).

Station	$a_2$ for Eq. (5)		$\lambda$ for Eq. (8)		$\kappa$ for GEVD, Eq. (9)			
	LS	MML	LS	MML	LS		MML	
					$\kappa$	U.B.	$\kappa$	U.B.
Victoria	1.30	1.33	1.30	1.50	0.13	98.7	0.14	97.0
Edmonton	0.21	0.64	0.21	0.64	-0.14	$\infty$	-0.07	$\infty$
Toronto	1.20	1.25	1.20	1.51	0.13	116.2	0.12	116.3
Quebec	1.20	1.29	1.20	1.50	0.17	100.2	0.16	100.7
St. John's	0.01	0.38	0.01	0.38	-0.34	$\infty$	-0.11	$\infty$



measurement scales used to describe the same physical process. Also, if the GEVD is used,  $\delta$  is not very large. The maximum, minimum, mean and standard deviation of  $\delta$  shown in Table 4 further confirms the above observations. More importantly, it shows that the maximum  $|\delta|$  can be as large as 3.5%, 9.3% and 11.1% for  $T= 50, 500$  and 1000 years if the Gumbel distribution is used to fit the wind speed and squared wind speed data using the LS method. These values become 7.7%, 15.1% and 17.4% if the MML is employed. This has implications in structural reliability as the factored design wind load for Canadian design practice (NRCC 2010) corresponds to about 500 year return period wind speed. For all three  $T$  values considered, the maximum  $|\delta|$  is less than 2.5% if Eq. (5) is used to fit the wind speed or squared wind speed with LS method or MML; the maximum  $|\delta|$  is less than 4.5% if GEVD is used. Also, in terms of minimum absolute value of the mean of  $\delta$ , or the minimum standard deviation of  $\delta$ , Eqs. (5) and (8) are preferable to the Gumbel distribution and GEVD.

Comparison of parameters shown in Table 5 to those shown in Table 3 indicates that the shape parameters  $a_2$  and  $\lambda$  shown in Table 5 are in general approximately equal to one-half of those shown in Table 3, especially if the values in Table 3 are greater than unity. It implies that the tail behaviour of the distribution of squared wind speed is consistent whether it is obtained by fitting the transformed data (i.e., squared wind speed) or it is obtained by transforming the distribution fitted to the wind speed data. However, this consistency cannot be achieved if the Gumbel distribution is employed directly to the wind speed data and the squared wind speed data as discussed in Hong (1994a, b), Cook and Harris (2004) and Harris (2009).

#### 4. CONCLUSIONS

We applied the BCT with the Gumbel distribution and the generalized extreme value distribution to fit the annual maximum wind data at a few Canadian locations. It is shown that the

application of the BCT with the Gumbel distribution leads to the distribution that has already been developed for the maximum of an independent and identically distributed Weibull sequence.

Using wind data from 14 meteorological stations in Canada, and carrying out extreme value analysis, it was concluded that the application of the Gumbel distribution with the BCT leads to very consistent estimate of the return period value of wind speed independent of whether the wind speed or squared wind speed data are used for the fitting. This consistency is insensitive to the range of return period of practical importance for structural reliability. This consistency is not achieved if the Gumbel distribution is directly applied to wind speed or squared wind speed data. In such a case the largest relative difference can be as high as 4%, 10% and 12% for  $T$  equal to 50, 500 and 1000 years if the least-squares method is used. These values become 9%, 17% and 19% if the method of maximum likelihood is used. Analysis results also showed that the exponent parameter in the BCT for 10 out of 14 stations is around 2, indicating that the Gumbel model could be more appropriate for the squared wind speed than for the wind speeds at 10 out of 14 considered stations.

Application of the generalized extreme value distribution with the BCT is also attempted. However, the application is ineffective for the extreme data considered because of the difficulty in finding the distribution model parameters by using the least-squares method or the method of maximum likelihood. This situation may be improved by using Bayesian methodology and requires future scrutiny.

#### 5. ACKNOWLEDGEMENTS

Financial support received from National Science and Engineering Research Council of Canada and the University of Western Ontario is much appreciated. I thank T.C. Eric Ho for his constructive comments and suggestions.

## 6. REFERENCES

- Achcar, J.A. Bolfarine, H. and Pericchi, L.R. (1987). Transformation of survival data to an extreme value distribution, *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol. 36, No. 2/3, pp. 229-234.
- Arnell, N.W., Beran, M., and Hosking, J.R.M., (1986). Unbiased plotting positions for the general extreme value distribution, *J. Hydrol.* 86, 59–69.
- Bertsekas, D.P. (1999). *Nonlinear programming*, Athena Scientific, (2nd Edition).
- Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistics Society. Series B (Methodological)*, 26:211{252.
- Coles, S. (2001). *An introduction to statistical modelling of extreme values*. Springer, London.
- Cook, N.J. (1982). Towards better estimation of extreme winds. *Journal of Wind Engineering and Industrial Aerodynamics*, 9:295-323.
- Cook, N.J. and Harris, R.I. (2004). Exact and general FT1 penultimate distributions of extreme wind speeds drawn from tail-equivalent Weibull parents. *Structural Safety*, 26:391-420.
- De Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction* (Springer Series in Operations Research and Financial Engineering)
- ESDU (2002). Computer program for wind speeds and turbulence properties: flat or hilly sites in terrain with roughness changes. *Engineering Science Data Unit (ESDU) Data Item No. 01008*.
- Frank, H. (2001). Extreme winds over Denmark from the NCEP/NCAR reanalysis. Technical Report Risø-R-1238(EN), Risø National Laboratory.
- Harris, R.I. (2009). XIMIS, a penultimate extreme value method suitable for all types of wind climate, *Journal of Wind Engineering and Industrial Aerodynamics*
- Holmes, J.D. and Moriarty, W.W. (1999). Application of the generalized Pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, 83: 1-10.
- Hong, H.P. (1994a). A note on extremal analysis, *Structural Safety*, Vol. 13 No. 4. pp. 227-233.
- Hong, H.P. (1994b). Estimate of extremal wind and wave loading and safety level of offshore structures, *Risk Analysis*, Proceeding of a Symposium (ed. Nowak, A. S.), University of Michigan, Ann Arbor, Michigan, USA.
- Hong, H.P., Li, S.H. and Mara, T. (2013) Performance of the generalized least-squares method for the extreme value distribution in estimating quantiles of wind speeds, *Journal of Wind Engineering & Industrial Aerodynamics*, 119: 121–132.
- Hong, H.P., Mara, T.G., Morris, R., Li, S.H. and Ye, W. (2014). Basis for recommending an update of wind velocity pressures in the 2010 National Building Code of Canada, *Canadian Journal of Civil Engineering*, 41(3), 206-221..
- Hosking, J. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc. B Met.*, 52(1), 105–124.
- Kasperski, M. (2002). A new wind zone map of Germany. *Journal of Wind Engineering and Industrial Aerodynamics*, 90: 1271–1287.
- Mara, T.G., Hong, H.P. and Morris, R.J. (2013). Effect of corrections to historical wind records on estimated extreme wind speeds, CSCE-2013 conference, Montreal, Canada.
- Martin, E.S. and Stedinger, J.R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resource Research*, 36: 737-744.
- NRCC (2010). *National Building Code of Canada*. Institute for Research in Construction, National Research Council of Canada, Ottawa, Ontario.
- Palutikof, J.P. Brabson, B.B., Lister, D.H. and Adcock, S.T. (1999). A review of methods to calculate extreme wind speeds, *Meteorol. Appl.* 6, 119–132.
- Peterka, J.A. and Shahid, S. (1998). Design Gust Wind Speeds in the United States, *J. Struct. Eng.*, 124, 207-214
- Simiu, E., Heckert, N.A., Filliben, J.J., and Johnson, S.K. (2001). Extreme wind load estimates based on the gumbel distribution of dynamic pressures: an assessment. *Structural Safety*, 23:221-229.
- Teugels, J.L. and Vanroelen G. (2004). Box–Cox transformations and heavy-tailed distributions. *J. Appl. Probab.* 41 213–227.
- Wadsworth, J.L., Tawn, J.A. and Jonathan P. (2010). Accounting for choice of measurement scaling in extreme value modeling, the *Annals of Applied Statistics*, 2010, Vol. 4, No. 3, 1558–1578.