On the mixture of wind speed distribution in a Nordic region

Taha B.M.J. Ouarda[1,*], Christian Charron[1]

[1] Canada Research Chair in Statistical Hydro-Climatology, INRS-ETE, 490 de la Couronne, Québec, QC, G1K 9A9, Canada

April 2018

Abstract

The assessment of wind energy potential at sites of interest requires reliable estimates of the statistical characteristics of wind speed. A probability density function (pdf) is usually fitted to short-term observed local wind speed data. It is common for wind speed data to present bimodal distributions for which conventional one-component pdfs are not appropriate. Mixture distributions represent an appropriate alternative to model such wind speed data. Homogeneous mixture distributions remain rarely used in the field of wind energy assessment while heterogeneous mixture models have only been developed recently. The present work aims to investigate the potential of homogeneous and heterogeneous mixture distributions to model wind speed data in a northern environment. A total of ten two-component mixture models including mixtures of gamma, Weibull, Gumbel and truncated normal are evaluated in the present study. The estimation of the parameter of the mixture models are obtained with the least-squares (LS) and the maximum likelihood (ML) methods. The optimization of the objective functions related to these estimation methods is carried out with a genetic algorithm that is more adapted to mixture distributions. The case study of the province of Québec (Canada), a Northern region with an enormous potential for wind energy production, is investigated in the present work. A total of 83 stations with long data records and providing a good coverage of the territory of the province are selected. To identify the most appropriate one-component distribution for the selected stations, the newly proposed method of L-moment ratio diagram (MRD) is used. The advantages of this approach are that it is simple to apply and it allows an easy comparison of the fit of several pdfs for several stations on a single diagram. One-component distributions are compared with the selected mixture distributions based on model selection criteria. Results show that mixture distributions often provide better fit than conventional one-component distributions

for the study area. It was also observed that the ML method outperforms the LS method and that the mixture model combining two Gumbel distributions using ML is the overall best model.

Keywords: wind speed distribution; moment ratio diagram; homogeneous mixture distribution; heterogeneous mixture distribution; probability density function; model selection criteria; kappa distribution.

Nomenclature

pdf	probability density function
cdf	cumulative distribution function
$f(\,)$	probability density function
P_i	cumulative empirical probability for the *i*th wind speed class interval
p_i	relative frequency for the *i*th wind speed class interval
$\hat{F}_i$	estimated cumulative probability for the *i*th wind speed class interval
$F(\,)$	cumulative distribution function
$F^{-1}(\,)$	inverse of a given cumulative distribution function
ω	mixing weight in two-component mixture distributions
θ	distribution parameters vector
W	Weibull probability distribution
E	Gumbel or extreme value type I probability distribution
G	gamma probability distribution
GEV	generalized extreme value probability distribution
GG	generalized gamma probability distribution
KAP	kappa probability distribution
LN	lognormal distribution
P3	Pearson type III distribution
ML	maximum likelihood
MM	method of moments
LS	method of least-square
MWW	mixture of two 2-parameter Weibull
MWTN	mixture of Weibull and singly truncated from below normal
MGW	mixture of gamma and Weibull
MGG	mixture of two gamma
MGTN	mixture of gamma and singly truncated from below normal
MGE	mixture of gamma and Gumbel
MEE	mixture of Gumbel and Gumbel

METN	mixture of Gumbel and singly truncated from below normal
MTNTN	mixture of two singly truncated from below normal
n	number of wind speed observations in a series of wind speed observations
N	number of class intervals
R_p^2	coefficient of determination giving the degree of fit between the estimated relative frequencies of the theoretical pdf and the empirical relative frequencies of the histogram of wind speed.
R_F^2	coefficient of determination giving the degree of fit between the theoretical cdf and the empirical cumulative probabilities of the histogram of wind speed.
RMSE	root mean square error of the predicted relative frequencies
KS	Kolmogorov-Smirnov test statistic
χ^2	Chi-square test statistic
v	wind speed
$M_{p,r,s}$	probability weighted moment of order *p*, *r*, *s*
λ_{r+1}	*r*th L-moment
ℓ_{r+1}	*r*th sample L-moment
τ_r	*r*th L-moment ratio
t_r	*r*th sample L-moments ratio
β_r	*r*th probability weighted moment where $M_{1,\mathrm{r},0}$
b_r	unbiased estimator of B_r

1. Introduction

The assessment of wind energy potential at sites of interest requires the estimation of the distribution of observed local wind speed data. For this purpose, a probability density function (pdf) is usually fit to short-term wind speed data (typically 1 hour). Selecting a PDF that correctly characterizes the wind speed distribution is crucial for reducing uncertainties in wind energy production estimates. The Weibull (W) is the most widely used and accepted distribution for the estimation of wind energy potential (Archer and Jacobson, 2003; Celik, 2003; Akpinar and Akpinar, 2005; Ahmed Shata and Hanitsch, 2006; Acker et al., 2007; Ayodele et al., 2012; Irwanto et al., 2014; Petković et al., 2014; Carrasco-Díaz et al., 2015; Dabbaghiyan et al., 2016; Yip et al., 2016). Its popularity can be attributed to its flexibility, simplicity and the fact that its parameters are easy to estimate (Tuller and Brett, 1983). However, W does not allow describing all encountered wind regimes in nature (Carta et al., 2008; Ouarda et al., 2015).

Several other distributions have been proposed in the literature for the assessment of wind energy: the gamma (G), generalized gamma (GG), inverse gamma (IG), inverse gaussian (IGA), lognormal (LN), logistic (L), log-logistic (LL), Gumbel (E), generalized extreme value (GEV), three-parameter beta (B), Pearson type III (P3), log-Pearson type III (LP3), Burr (BR), Erlang (ER), Johnson S_B, kappa (KAP) and Wakeby (WA) (Carta et al., 2009; Zhou et al., 2010; Lo Brano et al., 2011; Morgan et al., 2011; Masseran et al., 2012; Soukissian, 2013; Jung et al., 2017).

The identification of the statistical model that provides the best fit to the data represents a challenge. Traditionally, the fit was assessed using goodness-of-fit statistics and histograms of observed wind speed plotted together with candidate theoretical distributions. Recently, Ouarda et al. (2016) proposed the method of moment ratio diagram (MRD) for the selection of theoretical distributions. MRDs are commonly used in a number of fields for distribution assessment and parameter estimation (El Adlouni and Ouarda, 2007; Seckin et al., 2011), but were never applied to wind speed modeling.

With this approach, all possible values of the standardized kurtosis and the standardized skewness of the candidate distributions are usually plotted on a same graph. The sample moments of the observed data at the stations of interest are then estimated from the observations and plotted on the same graph. The selection of the appropriate distribution to fit the data sample is made based on the position of the sample moments in the graph. The advantage of using this approach is that it allows an easy comparison of the fit of several pdfs on a single graph. The approach allows also the analysis of the fit of data from several stations on the same graph. Hosking (1990) introduced the MRD using L-moment ratios instead of the conventional moment ratios. The theoretical advantages of L-moments over conventional moments are that they are able to characterize a wider range of distributions, they are more robust to the presence of outliers in the data when estimated from a sample, and are less subject to bias in the estimation (Hosking and Wallis, 1997). Ouarda et al. (2016) applied conventional MRD and L-moment MRD to wind speed data and concluded that L-moment MRD provide results that are more coherent with goodness-of-fit statistics and should be preferred over the conventional MRD. This conclusion was also obtained in other studies dealing with hydrologic data (Hosking, 1990; El Adlouni and Ouarda, 2007).

It has been shown in several studies that it is frequent for wind speed data to present distributions with bimodal regimes (Jaramillo and Borja, 2004; Shin et al., 2016; Soukissian and Karathanasi, 2017; Jung and Schindler, 2017; Mazzeo et al., 2018). In these cases, conventional pdfs are not suitable for modelling such distributions. To cope with such regimes, mixture distributions, defined as linear combinations of different distributions, were proposed by a number of authors (Carta et al., 2009; Ouarda et al., 2015; Shin et al., 2016). Proposed mixture models in the literature include mixtures of two Weibull distributions (Carta and Ramirez, 2007; Akpinar and Akpinar, 2009), two normal distributions (Jaramillo and Borja, 2004), singly truncated normal and Weibull distributions (Carta and Ramirez, 2007; Akpinar and Akpinar, 2009), two singly truncated normal distributions (Mazzeo et al., 2018; Chang, 2011), gamma and Weibull (Chang, 2011), Weibull and Gumbel distributions (Shin et al., 2016) and singly truncated normal and GEV distributions (Kollu et al., 2012). Homogeneous (the two components represent the same distribution) and heterogeneous (the two components represent two different distributions) mixture models are flexible and can provide good fit to bimodal regimes as well as unimodal regimes (Carta and Ramirez, 2007; Shin et al., 2016). Their use is gaining increasing popularity in the field of wind energy assessment and modeling.

Wind energy assessment and mapping studies remain relatively limited in Nordic environments. The case study of the province of Québec, Canada, is used in the present work to evaluate the suitability of homogeneous and heterogeneous mixture distributions for fitting wind speed data. Ten different mixture distribution models mixing W, G, E and truncated Normal are fitted to the wind speed at the stations of the study area. The choice of distribution functions used in mixture models is based on previous studies (e.g. Shin et al., 2016) as well as the asymptotic behavior of the tails of the distributions considered (El Adlouni et al., 2008), although the focus is not only on extreme wind speeds for energy generation. A large number of meteorological stations distributed all across the province are used for this objective. The province of Québec represents a region with an enormous undeveloped potential for wind energy production. A very limited number of studies evaluated the potential for wind energy in the province of Québec. In all these studies, only W was used (Ilinca et al., 2003; HE&AWS, 2005) except for a limited scope study where only two stations were explored (Ouarda and Charron, 2018). The method of the L-moment ratio diagram is also applied for evaluating the adequacy to the data of a selection of one-component distributions commonly used to model wind speed data. Note that, in their current state, MRD cannot be used to represent mixture distributions and thus, only one-component distributions are represented in these diagrams. Two-component homogenous and heterogeneous mixture distribution functions combining G, E, singly truncated normal (TN) and W as well as one-component W and KAP were fitted to the wind speed data of the study area. Validation of goodness-of-fit was made using criteria commonly used in the field of wind energy assessment. The results of the analysis are illustrated for a selection of 10 stations representing the study area. These stations provide a good illustration of the range of behaviors of wind speed distributions in the province of Québec.

The paper is organized as follows: Section 2 presents the theoretical background on L-moment ratio diagrams, one component probability distributions and mixture models. Section 3 presents the methodology of the study, including the representation of pdfs in MRDs, the estimation of distribution parameters and the model evaluation criteria. The case study dealing with wind speed data in the province of Quebec is presented in Section 4. The results are presented in Section 5, and the conclusions and future research directions are finally discussed in section 6 of the paper.

2. Theoretical background

2.1. L-moment ratio diagrams

L-moments introduced by Hosking (1990) represent an alternative to the conventional moments for the characterization of the shapes of probability distributions. The advantages of L-moments over conventional moments are that they are able to characterize a wider range of distributions, are more robust to the presence of outliers in the data and are less subject to bias in estimation (Hosking, 1990). For a given random variable *X* with a cumulative distribution function $F(X)$, the probability weighted moments (PWMs) are defined by (Greenwood et al., 1979):

$$M_{p,r,s} = \mathrm{E}[X^{p}\{F(X)\}^{r}\{1-F(X)^{s}\}]. \tag{1}$$

A useful special case of the PWM used in the definition of L-moments is given by:

$$\beta_r = M_{1,r,0} = \mathrm{E}[X\{F(X)\}^{r}] = \int_0^1 x(u)u^r du \tag{2}$$

where $x(u)$ is the quantile function of *X*. The L-moments of *X* are defined by (Hosking, 1990):

$$\lambda_{r+1} = \sum_{k=0}^{r} p^*_{r,k}\beta_k\,,\ r = 0,1,2,... \tag{3}$$

where

$$p^*_{r,k} = (-1)^{r-k}\binom{r}{k}\binom{r+k}{k}\,. \tag{4}$$

L-moments are directly interpretable as measures of the shape of distributions. The dimensionless versions of the L-moments, the L-variation, L-skewness and L-kurtosis, are respectively given by:

$$\begin{aligned}\tau_2 &= \lambda_2 / \lambda_1\\ \tau_3 &= \lambda_3 / \lambda_2\\ \tau_4 &= \lambda_4 / \lambda_2\end{aligned} \quad . \tag{5}$$

An important property makes the L-moments especially useful for the assessment of the goodness-of-fit with MRD: if the mean of the distribution exists, then all L-moments exist and the L-moments uniquely define the distribution (Hosking, 1997). L-moment ratios τ_4 vs. τ_3 in Eq. (5) are usually plotted in MRD for the assessment of the goodness-of-fit. A distribution function with one shape parameter, two shape parameters, or three or more shape parameters, is respectively represented as a point, a curve or an area in the MRD. The pdfs that are represented on the MRD of this study are given in Table 1 with their domain and number of parameters.

For a given data sample, the estimated values of the L-moment ratios can be obtained. Sample L-moment ratios are then plotted on the MRD to evaluate the adequacy of the pdfs represented in the MRD to the data samples. For an ordered sample of size *n*, $x_1 \leq x_2 \leq \cdots \leq x_n$, the sample L-moments are defined by:

$$\ell_{r+1} = \sum_{k=0}^{r} p^{*}_{r,k} b_k, \qquad r = 0,1,\ldots,n-1 \tag{6}$$

where

$$b_r = n^{-1} \binom{n-1}{r}^{-1} \sum_{j=r+1}^{n} \binom{j-1}{r} x_j . \tag{7}$$

The sample L-moment ratios analogous to the L-moment ratios in Eq. (5) are defined by:

$$\begin{aligned} t_2 &= \ell_2 / \ell_1 \\ t_3 &= \ell_3 / \ell_2 . \\ t_4 &= \ell_4 / \ell_2 \end{aligned} \tag{8}$$

2.2. One-component probability distributions

In this study, the whole range of distributions commonly considered in wind energy assessment and modeling were examined using MRD. The one-component W and KAP were identified as the only one-component probability distributions which provide a good fit to the wind speed data. Their pdfs were fitted to the wind speed data of the case study. W is the most used and recognized pdf for analysis of wind speed data. The pdf of W is given by:

$$f_{\mathrm{W}}(x) = \frac{k}{\alpha}\left(\frac{x}{\alpha}\right)^{k-1} \exp\left[-\left(\frac{x}{\alpha}\right)^{k}\right] \tag{9}$$

where $x > 0$, $\alpha > 0$ is a scale parameter and k is a shape parameter. The cumulative distribution function (cdf) of W is given by:

$$F_{\mathrm{W}}(x) = 1 - \exp\left[-\left(\frac{x}{\alpha}\right)^{k}\right]. \tag{10}$$

KAP, introduced by Hosking (1994) is a four-parameter distribution that includes the generalized logistic, the GEV, and the generalized Pareto distributions as special case. KAP was shown to lead to very good fit to wind speed data in previous studies (Shin et al., 2016; Jung et al., 2017). The pdf of the KAP is given by:

$$f_{\mathrm{KAP}}(x;\mu,\alpha,k,h) = \alpha^{-1}[1 - k(x-\mu)/\alpha]^{1/k-1}[F_{\mathrm{KAP}}(x)]^{1-h} \tag{11}$$

and the cdf of KAP is given by:

$$F_{\mathrm{KAP}}(x;\mu,\alpha,k,h)=(1-h(1-k(x-\mu)/\alpha)^{1/k})^{1/h} \tag{12}$$

where μ is a location parameter, α is a scale parameter, *h* and *k* are shape parameters.

2.3 Singly truncated from below distributions

The singly truncated from below normal (TN) distribution is often adopted instead of the conventional normal distribution in models for wind speed data (Carta et al., 2009). The reason is that N allows negative values of wind speed which is not possible. Truncated distributions are used to restrict the domain of the distributions. The truncation of the tails of the distribution was also shown in previous studies to be robust to extreme observations in the sample and to lead to improved estimates of the distribution moments, parameters and quantiles (see for instance Ouarda and Ashkar, 1998). In the context wind speed modeling, the restriction $x \geq 0$ is applied. TN has also the advantage over W of been able to represent calm frequencies as it is defined for $x=0$. However, adding a constraint to the Normal support makes the inference more complex when other distributions are considered. If $F_{\mathrm{N}}(x;\mu,\alpha)$ and $f_{\mathrm{N}}(x;\mu,\alpha)$ are the cdf and pdf of the normal distribution, the pdf and cdf of the TN are defined by:

$$f_{\mathrm{TN}}(x;\mu,\alpha)=\frac{1}{I_0(\mu,\alpha)}\,f_{\mathrm{N}}(x;\mu,\alpha), \tag{13}$$

$$F_{\mathrm{TN}}(x;\mu,\alpha)=\frac{1}{I_0(\mu,\alpha)}\int_0^x f_{\mathrm{N}}(x;\mu,\alpha)=\frac{F_{\mathrm{N}}(x;\mu,\alpha)-F_{\mathrm{N}}(0;\mu,\alpha)}{I_0(\mu,\alpha)} \tag{14}$$

where $x \geq 0$ and the function $I_0(\mu,\alpha)=\int_0^\infty f_{\mathrm{N}}(x;\mu,\alpha)=1-F_{\mathrm{N}}(0;\mu,\alpha)$ ensures that the integral of the pdf of TN is equal to one. Similarly, the truncated pdf and cdf of E can also be obtained by using Eq. (13) and (14) and replacing $f_{\mathrm{N}}(x;\mu,\alpha)$ by $f_{\mathrm{E}}(x;\mu,\alpha)$.

2.4. Mixture probability distributions

Mixture distributions are defined as linear combinations of two or several distributions. For a mixed distribution with *d* components, the pdf is given by:

$$f(v;\omega,\theta)=\sum_{i=1}^{d}\omega_i f_i(v;\theta_i)\,. \tag{15}$$

where θ_i are the parameters of the *i*th distribution, $f_i(v;\theta_i)$ are independently distributed *i*th components and ω_i are mixing parameters such that $\sum_{i=1}^{d}\omega_i=1$. In the case of a two-component mixture distribution, the mixture density function is then:

$$f(v;\omega,\theta_1,\theta_2)=\omega f_1(v;\theta_1)+(1-\omega)f_2(v;\theta_2)\,. \tag{16}$$

where $0<\omega<1$ is the mixing weight, and θ_1 and θ_2 are vectors of parameters for the first and second component of the distribution.

Similarly to the previous study of Shin et al. (2016), the G, W, E, and TN distributions were adopted as density components of mixture distributions. In all, 10 mixture distributions are obtained with the combination of the different components considered and are denoted by: MGW, MGE, MGTN, MWW, MWE, MWTN, MEE, MEN and MTNTN. The pdfs of these mixture models are presented in Table 2.

3. Methodology

3.1. Representation of the pdfs in MRD

In this section we explain how selected pdfs in Table 1 are represented in the MRD. The distribution E, having no shape parameter, is defined by a dot on the MRD. Distributions E, GEV, W, P3 and G have a single shape parameter, and plot as a line in the MRD. For the previous distributions, polynomial approximations of τ_4 as function of τ_3 are available in Hosking and Wallis (1997) and are used to plot the lines corresponding to these distributions in the MRD. Distributions GG, LP3 and KAP having two shape parameters define areas in the MRD and bounds of these areas are represented in the MRD. Analytical expressions of these bounds are generally not available. In that case, the following numerical method is applied: For a given pdf with two shape parameters *h* and *k*, a position parameter μ and/or a scale parameter α, parameters *h* and *k* are varied over a large range within the feasibility domain of the given pdf and with small intervals ($h = h_1, h_2, \ldots, h_n; k = k_1, k_2, \ldots, k_m$). Parameters μ and α are given arbitrary values because they are independent of L-moment ratios τ_3 and τ_4. For each generated pair of values (h_i, k_j), the corresponding pairs of moment ratios ($\tau_{3,i,j}, \tau_{4,i,j}$) are computed and are plotted on the L-moment ratio diagram. Afterwards, the contours of the regions defined by these points are defined.

For KAP, the expressions of L-moment ratios τ_3 and τ_4 as a function of its distribution parameters are given in Hosking and Wallis (1997). Explicit expressions of L-moments as a function of the distribution parameters of the GG and LP3 are not available. In this case, B_1, B_2 and B_3 are estimated by numerically integrating the distribution in Eq. (2) and λ_2 λ_3 and λ_4 are obtained using Eq. (5).

3.2. Parameter estimation

Parameters of the W are estimated in the present work with the method of moments (MM). Parameters of KAP are estimated here with the method of L-moments (Hosking, 1997). Algorithm for this method can be found in Hosking (1996).

The parameters of the mixture distributions are commonly estimated with the least-square (LS) method (Carta and Ramirez, 2007; Shin et al., 2016; Jung and Schindler, 2017) and the maximum likelihood (ML) method (Carta et al., 2009; Shin et al., 2016). The least-squares are defined by:

$$SSE = \sum_{i=1}^{N} \left[P_i - F(v_{\max,i} \mid \theta) \right]^2 \tag{17}$$

where P_i is the cumulative empirical probability of the ith group, $v_{\max,i}$ is the maximum wind speed of the ith group and θ is a parameter vector. Observed wind speed data are arranged into N class intervals $[0, v_1), [v_1, v_2), ..., [v_{N-1}, v_N]$. Relative frequencies p_i are computed for each class interval and $P_i = \sum_{j=1}^{i} p_j$ is the cumulative empirical probability at the ith class.

The maximum likelihood method was applied on observed wind speeds. It is proposed here to use the maximum likelihood with the class interval approach. Given an underlying distribution $f(x;\theta)$ for the wind speed, the likelihood is given by (Carter et al., 1971):

$$L(\theta) = C \prod_{i=1}^{N} p_i^{n_i} \tag{18}$$

where $C = n! \Big/ \prod_{i=1}^{N} n_i!$ and n_i is the number of observations in the ith class interval. It is generally more convenient to optimize the log-likelihood given by:

$$\log L(\theta) = \log(C) + \sum_{i=1}^{N} n_i \log p_i \ . \qquad (19)$$

To optimize the least-squares function in Eq. (17) and the log-likelihood function in Eq. (19), a genetic algorithm (GA) is used. GA has been used in different fields for the optimization of a given objective function (Hassanzadeh et al., 2011). A particularity of GA which makes it attractive for solving the problem associated to the estimation of the parameters of mixture distributions is that it does not require defining initial values for the parameters, which is difficult in the case of mixture distributions (Ouarda et al., 2015).

The approaches presented here for parameter estimation are sensitive to the discretization interval selection. The intervals should have small extent but also contain enough observations, which is not possible especially for small sample sizes. The choice of intervals depends also on the sensitivity of the anemometer. For less precise anemometers, it is not possible to use very fine intervals.

3.3. Validation

The chi-square test statistic (χ^2), the coefficient of determination (R^2), the RMSE and the Kolmogorov-Smirnov test statistic (KS) are used for the validation of the goodness-of-fit of the different models. These criteria are frequently used for the evaluation of the goodness-of-fit (Ouarda et al., 2016). Before the computation of the statistics, wind speed data are arranged in N class intervals and relative frequencies p_i are computed at each class interval.

Two indices are used to define R^2. The first index is defined by:

$$R_F^2 = 1 - \frac{\sum_{i=1}^{N} (P_i - \hat{F}_j)^2}{\sum_{i=1}^{N} (P_i - \bar{P})^2} \tag{20}$$

where $\hat{F}_i$ is the predicted cumulative probability of the theoretical distribution at the ith class interval, $P_i = \sum_{j=1}^{i} p_j$ is the empirical cumulative probability at the ith class interval and $\bar{P} = \frac{1}{N}\sum_{i=1}^{N} P_i$. The second index is defined by:

$$R_p^2 = 1 - \frac{\sum_{i=1}^{N} (p_i - \hat{p}_i)^2}{\sum_{i=1}^{N} (p_i - \bar{p})^2} \tag{21}$$

where $\hat{p}_i = F(v_i) - F(v_{i-1})$ is the estimated probability at the ith class interval, v_{i-1} and v_i are the lower and upper limits of the ith class interval and $\bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i$.

The RMSE is a measure of the error in the estimation of the relative frequencies and is given by:

$$\mathrm{RMSE} = \left[\sum_{i=1}^{N} (p_i - \hat{p}_i)^2 / N \right]^{1/2} . \tag{22}$$

The χ^2 test statistic is a measure the adequacy of a given theoretical distribution to a data sample and is expressed as:

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(O_i - E_i\right)^2}{E_i} \tag{23}$$

where O_i is the observed frequency in the *i*th class interval and E_i is the expected frequency in the *i*th class interval. When E_i for a given class interval is very small, it is combined with the adjacent class interval in order to avoid the situation where E_i has an excessive weight. The KS statistic corresponds to the largest difference between the predicted and the observed distribution and is given by:

$$D = \max_{1 \le i \le N} \left| P_i - \hat{F}_i \right|. \tag{24}$$

A lower value of χ^2, RMSE or KS, and a higher value of R^2 indicate a better fit.

4. Nordic environment case study

4.1. Region of study

The province of Quebec (Canada) covers a territory of over 1.5 million km^2 and has an enormous potential for wind energy production. In a study commended by the government of Quebec, it was concluded that the exploitable potential in Quebec is close to 4 000 000 MW (HE&AWS, 2005). The majority of energy production in Quebec comes traditionally from hydroelectricity (Barbet et al., 2006). Because of the large existing and potential hydroelectric resources, the development of other renewable sources of energy has been considerably delayed. Nevertheless, an increasing interest for renewable energy and especially for wind energy harvesting is observed recently. The government of Quebec requested in its new energy policy to support the development of new wind farm projects on the territory (Gouvernement du Québec, 2016). Wind generation is nowadays considered as a viable alternative for energy supply in remote rural areas, especially in the Northern regions.

A very limited number of studies dealt with modeling wind speed and assessing the wind energy resources in the province of Quebec (Ilinca et al., 2003; HE&AWS, 2005; Hundecha et al., 2008). Ilinca et al. (2003) and HE&AWS (2005) evaluated the potential for wind energy in the province based solely on the W distribution. Hundecha et al. (2008) studied the changes in the annual maximum 10-m wind speed in and around the Gulf of St. Lawrence, Canada, through a nonstationary extreme value analysis. The study was based on the North American Regional Reanalysis (NARR) dataset as well as observed data from a selection of stations located on and around the Gulf of St. Lawrence.

4.2. Wind speed data

The wind speed data used in this study were obtained from "Environment and Climate Change Canada", the Federal ministry of the Environment. Meteorological data are available freely at http://climate.weather.gc.ca. Observed data consist of mean hourly wind speeds observed at 10 m above the ground for meteorological stations distributed across the province of Quebec. Stations with identical coordinates were combined together. Stations of the database with at least one complete year of data were selected. A total of 83 stations covering most of the territory of the province of Québec were selected. The geographical location of the selected stations is illustrated in Fig. 1. The majority of the stations are located in the southern part of the province of Quebec on both sides of the Saint-Lawrence River. The network density is significantly higher in the southern part of the province due to the concentration of major urban agglomerations and economic activities in this region.

The lengths of the data series at the stations range from 1 to 65 years with a median of 21 years of data. The calm frequencies at the stations are very low. 10 stations having long data series and a good distribution across the study area were selected to illustrate the results of the present study. These stations are considered representative of the whole data base. A detailed description of the selected stations is presented in Table 3 with information concerning the period of record, the geographical location and the statistical characteristics of the wind speed data. The selected 10 stations are illustrated in the map of Fig. 1 with red dots. The rest of the stations are illustrated with black dots.

5. Results

Fig. 2 presents the L-moment ratio diagram with the selected one-component pdfs. KAP, covering the largest area in the MRD, is thus the most flexible pdf, followed by LP3 and GG. The GEV, W, G, P3 and LN, plot as lines and are similar for values of τ_3 around the zero value. E is a special case of the GEV. The distributions G, P3 and W are special cases of the GG, and the distributions GEV and E are special cases of the KAP distribution. Sample L-moments were computed using Eq. (8) and (6) for each station of the study area and were plotted on the MRD. It can be observed in Fig. 2 that the curve defined by W passes through the cloud of points defined by the sample L-moment ratios. For the other distributions defining a curve (G, P3, GEV and LN), the lines are located over the cloud of points and the distributions are thus inadequate for representing wind speed data at the stations of the case study. It can hence be concluded that W is the most suitable pdf with one shape parameter.

All 83 stations are located within the regions that are bounded by the pdfs of the distributions possessing two shape parameters (GG, LP3 and KAP), and thus, these pdfs can represent appropriate models for the wind speed data of the Quebec stations. Even though on average W represents a good model, it may not be suitable for data sets located far from the curve defined by W. In these cases, GG, LP3 and KAP provide a better fit and are more appropriate.

MRDs are useful tools for studying the fit of one-component probability distributions commonly used in the field of wind energy assessment. However, they are not able to identify distributions with bimodal or multimodal regimes. In some cases where bimodality is detected, the use of mixture distributions is necessary. Future research efforts can focus on the extension of the MRD approach to bimodal and some mixture distributions.

The one-component distributions W and KAP as well as the selected mixture distributions were fitted to the wind speed series of the case study. The class interval is set to 1 m/s for the computation of the least square in Eq. (17), for the log-likelihood in Eq. (19) and for the computation of the goodness-of-fit criteria in Section 4.3. For the present study, important improvements in the fit were obtained by using TN instead of the conventional N in the mixture models. Consequently, the results using TN are presented here. In the case of E, no improvement was obtained with the truncated E and thus results with the conventional E in the mixture models are presented here.

The goodness-of-fit criteria presented in Section 4.3 were computed at all stations and results are presented in Fig. 3 with box plots. According to the criteria, the one-component KAP performs better than W. However KAP has two more outlying observations than W for χ^2.

Mixture models using the ML approach perform significantly better than the corresponding models using the LS approach. In general, we do not observe a big difference in the performances of the various mixture models using the LS approach. All mixture models using the ML approach perform better that the one-component W and KAP and according to the RMSE, R_p^2 and χ^2 criteria, the majority of mixture models using the LS approach perform also better than W and KAP. The overall best model is obtained with MEE/ML. Mixture models including TN generally lead to lower performances. Bimodality may not always be present in wind speed data series and this explains the general good performances obtained by one-component distributions W and KAP.

In Fig. 4, the histograms of the observed wind speed data at the 10 stations selected to illustrate the results for the prince of Quebec are presented. W and KAP as well as the first and second mixture distribution models providing the best fit to the data according to χ^2 are superimposed in these histograms. Table 4 lists the four best models at each station according to each criterion. Mixture distributions have in general higher ranks than one-component distributions. The flexibility and advantages of mixture distribution models are illustrated in Fig. 4 as one-component distributions are shown not to be suitable to model all stations. For instance, mixture distributions are necessary to model stations Quebec, Cap-Madeleine, Bagotville, Val-d'Or and Parc national des Pingualuit. Even for stations presenting unimodal behaviors, such as the stations of Montréal, Cape Whittle, Mont-Joli, Kuujjuarapik and Nitchequon, mixture distributions provide also, in general, the best goodness-of-fit statistics according to the ranks in Table 4.

6. Conclusions and future work

In this study, we evaluate the suitability of a selection of homogeneous and heterogeneous two-component distributions as well as a number of one-component distributions including W and KAP to model wind speed data in a Nordic environment. The case study is represented by 83 meteorological stations distributed throughout the wide territory of the province of Québec, Canada. The approach consisted first in using the L-moment ratio diagrams to assess the one-component distributions that best fit the data. Among the pdfs defining a curve (probability distributions having one shape parameter) on the MRD, W is the pdf leading to the best fit. For the distributions with two shape parameters, GG, LP3 and KAP, areas of feasibility are defined in the MRD diagram and these distributions can represent better alternatives for the stations whose data samples are located the farthest from the curve defined by W.

MRDs are not able to represent distributions presenting bimodal behaviors. Mixture distributions can be used to model such behavior. A selection of 10 two-component distributions mixing W, G, E and TN were fitted to the wind speed at the stations of the study area. The parameters were estimated with the LS and ML methods. Results were compared to the fit obtained by the most adequate one-component distributions: the W and KAP distributions. Global results indicated that mixture models provide better goodness-of-fit than the W and KAP according to the performance criteria used. It was found that the ML method outperforms the LS method according to all criteria. Mixture distributions are flexible and can efficiently model both bimodal and unimodal behaviors.

The results of the present study show that mixture distribution models have the potential to improve the estimation of energy generation potential at stations presenting bimodal regimes and even at stations presenting unimodal regimes. Improved accuracy in wind energy potential assessment can help with site selection and with the design and management of wind farms. The proposed methods are general and can be transposed to other regions especially those where pronounced bimodal regimes are observed.

10 stations with a good distribution across the study area were selected to illustrate the results of the study. The histograms of the fitting of the one-component and mixture models to the wind speed data at the 10 selected stations are presented. The analysis of the histograms of the wind speed data at each station have shown that a bimodal behavior was observed in about 5 stations. For these stations, mixture distributions reveal to be necessary in order to adequately model the wind speed distributions.

It is important to note that the mixture models applied in this study present additional complexity in comparison to simpler models such the one-component Weibull. Parameter estimation for mixture models requires advanced optimization method such as the genetic algorithm used here. This method takes more time to process than other optimization methods. This can be cumbersome when the mixture approach is applied to a large number of stations for instance.

The MRD approach needs to be extended to bimodal and mixture distributions in order to be useful for the whole range of distributions of interest for wind energy assessment and modeling. Future work should also focus on the analysis of the non-stationarity in wind speed data (presence of trends, jumps and cycles) in the province of Quebec in order to provide reliable estimates of the future potential for wind energy generation. The frequency analysis models used in the present study and in most literature dealing with wind speed modeling are based on the hypothesis of the stationarity of the wind speed regime. Unfortunately, such assumption is often invalid, and past wind speed observations are not necessarily representative of the future wind speed regime. Increasing attention is being devoted to the development of non-stationary frequency modeling tools for climatic variables, which take into consideration information about climate change (see for instance Lee and Ouarda, 2011; Chandran et al., 2016).

Future work should also focus on the extension of homogeneous and heterogeneous mixture models to the non-stationary case. The resulting models will have distribution parameters that are dependent on the values of covariates that may represent time or climate indices. A non-stationary frequency analysis of wind speed data in the province of Quebec can also integrate low frequency climate oscillation indices as covariates to take into consideration information concerning the impact of these climatic indices on the inter-annual variability in wind speed in the region. Such models are becoming increasingly popular in climatology and renewable energy modeling (see for instance Ouachani et al., 2013; Naizghi and Ouarda, 2017) and would allow understanding the teleconnections of wind characteristics with various global climate indices and examining the long-term variability of wind speed in the province. Thiombiano et al. (2017) have already identified the Arctic Oscillation (AO) and the Pacific North American (PNA) climate indices as the dominant indices in the region. These indices can be integrated relatively easily in the models developed in the present work.

Acknowledgements

Financial support for the present study was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors wish to thank Environment and Climate Change Canada for having supplied the wind speed data used in this study. The authors are grateful to the Editor-in-Chief, Dr. Marc Rosen, and to two anonymous reviewers for their comments which helped improve the quality of the manuscript.

References

Acker, T.L., Williams, S.K., Duque, E.P.N., Brummels, G., Buechler, J., 2007. Wind resource assessment in the state of Arizona: Inventory, capacity factor, and cost. Renewable Energy, 32(9): 1453-1466. doi:10.1016/j.renene.2006.06.002.

Ahmed Shata, A.S., Hanitsch, R., 2006. Evaluation of wind energy potential and electricity generation on the coast of Mediterranean Sea in Egypt. Renewable Energy, 31(8): 1183-1202. doi: 10.1016/j.renene.2005.06.015.

Akpinar, E.K., Akpinar, S., 2005. An assessment on seasonal analysis of wind energy characteristics and wind turbine characteristics. Energy Conversion and Management, 46(11-12): 1848-1867. doi:10.1016/j.enconman.2004.08.012.

Akpinar, S., Akpinar, E.K., 2009. Estimation of wind energy potential using finite mixture distribution models. Energy Conversion and Management, 50(4): 877-884. doi:10.1016/j.enconman.2009.01.007.

Archer, C.L., Jacobson, M.Z., 2003. Spatial and temporal distributions of U.S. winds and wind power at 80 m derived from measurements. Journal of Geophysical Research: Atmospheres, 108(D9): 4289. doi:10.1029/2002jd002076.

Ayodele, T.R., Jimoh, A.A., Munda, J.L., Agee, J.T., 2012. Wind distribution and capacity factor estimation for wind turbines in the coastal region of South Africa. Energy Conversion and Management, 64: 614-625. doi:10.1016/j.enconman.2012.06.007.

Barbet, M., Bruneau, P., Ouarda, T.B.M.J., Gingras, H., 2006. REGIONS – Software for regional flood estimation. HYDRO-2006 conference: Maximizing the benefits of hydropower, Porto-Carras, Greece, 25th -28th September 2006.

Carrasco-Díaz, M., Rivas, D., Orozco-Contreras, M., Sánchez-Montante, O., 2015. An assessment of wind power potential along the coast of Tamaulipas, northeastern Mexico. Renewable Energy, 78: 295-305. doi:10.1016/j.renene.2015.01.007.

Carta, J.A., Ramírez, P., 2007. Use of finite mixture distribution models in the analysis of wind energy in the Canarian Archipelago. Energy Conversion and Management, 48(1): 281-291. doi:10.1016/j.enconman.2006.04.004.

Carta, J.A., Ramirez, P., Velazquez, S., 2008. Influence of the level of fit of a density probability function to wind-speed data on the WECS mean power output estimation. Energy Conversion and Management, 49(10): 2647-2655. doi:10.1016/j.enconman.2008.04.012.

Carta, J.A., Ramirez, P., Velazquez, S., 2009. A review of wind speed probability distributions used in wind energy analysis Case studies in the Canary Islands. Renewable & Sustainable Energy Reviews, 13(5): 933-955. doi:10.1016/j.rser.2008.05.005.

Carter, W.H., Bowen, J.V., Myers, R.H., 1971. Maximum likelihood estimation from grouped Poisson data. Journal of the American Statistical Association, 66(334): 351-353. doi:10.1080/01621459.1971.10482267.

Celik, A.N., 2003. Energy output estimation for small-scale wind power generators using Weibull-representative wind data. Journal of Wind Engineering and Industrial Aerodynamics, 91(5): 693-707. doi:10.1016/s0167-6105(02)00471-3.

Chandran, A., Basha, G., Ouarda, T.B.M.J., 2016. Influence of climate oscillations on temperature and precipitation over the United Arab Emirates. International Journal of Climatology, 36(1): 225-235. doi:10.1002/joc.4339.

Chang, T.P., 2011. Estimation of wind energy potential using different probability density functions. Applied Energy, 88(5): 1848-1856. doi:10.1016/j.apenergy.2010.11.010.

Dabbaghiyan, A., Fazelpour, F., Abnavi, M.D., Rosen, M.A., 2016. Evaluation of wind energy potential in province of Bushehr, Iran. Renewable and Sustainable Energy Reviews, 55: 455-466. doi:10.1016/j.rser.2015.10.148.

El Adlouni, S., Bobée, B., Ouarda, T.B.M.J., 2008. On the tails of extreme event distributions in hydrology. Journal of Hydrology, 355(1-4): 16-33. doi:10.1016/j.jhydrol.2008.02.011.

El Adlouni, S., Ouarda, T.B.M.J., 2007. Orthogonal projection L-moment estimators for three-parameter distributions. Advances and Applications in Statistics, 7(2): 193-209.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R., 1979. Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. Water Resources Research, 15(5): 1049-1054. doi:10.1029/WR015i005p01049.

Gouvernement du Québec, 2016. Politique énergétique 2030. Retrieved from: http://politiqueenergetique.gouv.qc.ca/wp-content/uploads/politique-energetique-2030.pdf

Hassanzadeh, Y., Abdi, A., Talatahari, S., Singh, V.P., 2011. Meta-Heuristic Algorithms for Hydrologic Frequency Analysis. Water Resources Management, 25(7): 1855-1879. doi:10.1007/s11269-011-9778-1.

HE&AWS, 2005. Inventaire du potentiel éolien exploitable du Québec, Hélimax Énergie inc., AWS Truewind, LLC. Retrieved from: http://www.mrn.gouv.qc.ca/publications/energie/eolien/vent_inventaire_inventaire_2005.pdf

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 (Methodological), 52(1): 105-124. doi:10.2307/2345653.

Hosking, J.R.M., 1994. The four-parameter kappa distribution. IBM Journal of Research and Development, 38(3): 251-258. doi:10.1147/rd.383.0251.

Hosking, J.R.M., 1996. Fortran routines for use with the method of L-moments, version 3.04. 20525, IBM Research Division, Yorktown Heights, N.Y.

Hosking, J.R.M., Wallis, J.R., 1997. Regional frequency analysis: An approach based on L-Moments. Cambridge University Press, New York, 240 pp.

Hundecha, Y., St-Hilaire, A., Ouarda, T.B.M.J., El Adlouni, S., Gachon, P., 2008. A Nonstationary Extreme Value Analysis for the Assessment of Changes in Extreme Annual Wind Speed over the Gulf of St. Lawrence, Canada. Journal of Applied Meteorology and Climatology, 47(11): 2745-2759. doi:10.1175/2008jamc1665.1.

Ilinca, A., McCarthy, E., Chaumel, J.-L., Rétiveau, J.-L., 2003. Wind potential assessment of Quebec Province. Renewable Energy, 28(12): 1881-1897. doi:10.1016/S0960-1481(03)00072-7.

Irwanto, M., Gomesh, N., Mamat, M.R., Yusoff, Y.M., 2014. Assessment of wind power generation potential in Perlis, Malaysia. Renewable and Sustainable Energy Reviews, 38: 296-308. doi:10.1016/j.rser.2014.05.075.

Jaramillo, O.A., Borja, M.A., 2004. Wind speed analysis in La Ventosa, Mexico: a bimodal probability distribution case. Renew. Energy, 29(10): 1613-1630. doi:10.1016/j.renene.2004.02.001.

Jung, C., Schindler, D., 2017. Global comparison of the goodness-of-fit of wind speed distributions. Energy Conversion and Management, 133: 216-234. doi:10.1016/j.enconman.2016.12.006.

Jung, C., Schindler, D., Laible, J., Buchholz, A., 2017. Introducing a system of wind speed distributions for modeling properties of wind speed regimes around the world. Energy Conversion and Management, 144: 181-192. doi:10.1016/j.enconman.2017.04.044.

Kollu, R., Rayapudi, S.R., Narasimham, S., Pakkurthi, K.M., 2012. Mixture probability distribution functions to model wind speed distributions. International Journal of Energy and Environmental Engineering, 3(1): 1-10. doi:10.1186/2251-6832-3-27.

Lee, T., Ouarda, T.B.M.J., 2011. Prediction of climate nonstationary oscillation processes with empirical mode decomposition. Journal of Geophysical Research: Atmospheres, 116: D06107. doi:10.1029/2010jd015142.

Lo Brano, V., Orioli, A., Ciulla, G., Culotta, S., 2011. Quality of wind speed fitting distributions for the urban area of Palermo, Italy. Renewable Energy, 36(3): 1026-1039. doi:10.1016/j.renene.2010.09.009.

Masseran, N., Razali, A.M., Ibrahim, K., 2012. An analysis of wind power density derived from several wind speed density functions: The regional assessment on wind power in Malaysia. Renewable & Sustainable Energy Reviews, 16(8): 6476-6487. doi:10.1016/j.rser.2012.03.073.

Mazzeo, D., Oliveti, G., Labonia, E., 2018. Estimation of wind speed probability density function using a mixture of two truncated normal distributions. Renewable Energy, 115: 1260-1280. doi:10.1016/j.renene.2017.09.043.

Morgan, E.C., Lackner, M., Vogel, R.M., Baise, L.G., 2011. Probability distributions for offshore wind speeds. Energy Conversion and Management, 52(1): 15-26. doi:10.1016/j.enconman.2010.06.015.

Naizghi, M.S., Ouarda, T.B.M.J., 2017. Teleconnections and analysis of long-term wind speed variability in the UAE. International Journal of Climatology, 37(1): 230-248. doi:10.1002/joc.4700.

Ouachani, R., Bargaoui, Z., Ouarda, T.B.M.J., 2013. Power of teleconnection patterns on precipitation and streamflow variability of upper Medjerda Basin. International Journal of Climatology, 33(1): 58-76. doi:10.1002/joc.3407.

Ouarda, T.B.M.J., Ashkar, F., 1998. Effect of Trimming on LP III Flood Quantile Estimates. Journal of Hydrologic Engineering, 3(1): 33-42. doi:10.1061/(ASCE)1084-0699(1998)3:1(33).

Ouarda, T.B.M.J.,Charron, C., 2018. Distributions of wind speed in a northern environment, 2018 9th International Renewable Energy Congress (IREC), Hammamet, Tunisia, pp. 1-3. DOI:10.1109/IREC.2018.8362453

Ouarda, T.B.M.J., Charron, C., Chebana, F., 2016. Review of criteria for the selection of probability distributions for wind speed data and introduction of the moment and L-moment ratio diagram methods, with a case study. Energy Conversion and Management, 124: 247-265. doi:10.1016/j.enconman.2016.07.012.

Ouarda, T.B.M.J., Charron, C., Shin, J.Y., Marpu, P.R., Al-Mandoos, A.H., Al-Tamimi, M.H., Ghedira, H., Al Hosary, T.N., 2015. Probability distributions of wind speed in the UAE. Energy Conversion and Management, 93: 414-434. doi:10.1016/j.enconman.2015.01.036.

Petković, D., Shamshirband, S., Anuar, N.B., Saboohi, H., Abdul Wahab, A.W., Protić, M., Zalnezhad, E., Mirhashemi, S.M.A., 2014. An appraisal of wind speed distribution prediction by soft computing methodologies: A comparative study. Energy Conversion and Management, 84: 133-139. doi:10.1016/j.enconman.2014.04.010.

Seckin, N., Haktanir, T., Yurtal, R., 2011. Flood frequency analysis of Turkey using L-moments method. Hydrological Processes, 25(22): 3499-3505. doi:10.1002/hyp.8077.

Shin, J.-Y., Ouarda, T.B.M.J., Lee, T., 2016. Heterogeneous mixture distributions for modeling wind speed, application to the UAE. Renewable Energy, 91: 40-52. doi:10.1016/j.renene.2016.01.041.

Soukissian, T., 2013. Use of multi-parameter distributions for offshore wind speed modeling: The Johnson SB distribution. Applied Energy, 111: 982-1000. doi:10.1016/j.apenergy.2013.06.050.

Soukissian, T.H., Karathanasi, F.E., 2017. On the selection of bivariate parametric models for wind data. Applied Energy, 188: 280-304. doi:10.1016/j.apenergy.2016.11.097.

Thiombiano, A.N., El Adlouni, S., St-Hilaire, A., Ouarda, T.B.M.J., El-Jabi, N., 2017. Nonstationary frequency analysis of extreme daily precipitation amounts in Southeastern Canada using a peaks-over-threshold approach. Theoretical and Applied Climatology, 129(1): 413-426. doi:10.1007/s00704-016-1789-7.

Tuller, S.E., Brett, A.C., 1983. The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis. Journal of Climate and Applied Meteorology, 23.

Yip, C.M.A., Gunturu, U.B., Stenchikov, G.L., 2016. Wind resource characterization in the Arabian Peninsula. Applied Energy, 164: 826-836. doi:10.1016/j.apenergy.2015.11.074.

Zhou, J.Y., Erdem, E., Li, G., Shi, J., 2010. Comprehensive evaluation of wind speed distribution models: A case study for North Dakota sites. Energy Conversion and Management, 51(7): 1449-1458. doi:10.1016/j.enconman.2010.01.020.

Table 1. List of probability density functions, domains, and list of parameters.

Name	Probability density function	Domain	Number of parameters
E	$f_{\mathrm{E}}(x;\mu,\alpha)=\frac{1}{\alpha}\exp\left[-\frac{x-\mu}{\alpha}-\exp\left(-\frac{x-\mu}{\alpha}\right)\right]$	$-\infty<x<+\infty$	2
W	$f_{\mathrm{W}}(x;\alpha,k)=\frac{k}{\alpha}\left(\frac{x}{\alpha}\right)^{k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{k}\right]$	$0\le x<\infty$	2
G	$f_{\mathrm{G}}(x;\alpha,k)=\frac{\alpha^{k}}{\Gamma(k)}x^{k-1}\exp(-\alpha x)$	$0\le x<\infty$	2
LN	$f_{\mathrm{LN}}(x;\mu,\alpha)=\frac{1}{x\alpha\sqrt{2\pi}}\exp\left[-\frac{(\ln x-\mu)^{2}}{2\alpha^{2}}\right]$	$0\le x<\infty$	2
GG	$f_{\mathrm{GG}}(x;\alpha,k,h)=\frac{\lvert h\rvert\alpha^{hk}}{\Gamma(k)}x^{hk-1}\exp(-\alpha x)^{h}$	$0\le x<\infty$	3
P3	$f_{\mathrm{P3}}(x;\mu,\alpha,k)=\frac{\alpha^{k}}{\Gamma(k)}(x-\mu)^{k-1}\exp\left[-\alpha(x-\mu)\right]$	$\mu\le x<\infty$	3
GEV	$f_{\mathrm{LN}}(x;\mu,\alpha,k)=\frac{1}{\alpha}\left[1-\frac{k}{\alpha}(x-u)\right]^{\frac{1}{k}-1}\exp\left\{-\left[1-\frac{k}{\alpha}(x-u)\right]^{1/k}\right\}$	$u+\alpha/k\le x<\infty$ if $k<0$ $-\infty<x\le u+\alpha/k$ if $k>0$ $-\infty<x<\infty$ if $k=0$	3
LP3	$f_{\mathrm{LP3}}(x;\mu,\alpha,k)=\frac{g\lvert\alpha\rvert}{x\Gamma(k)}\left[\alpha(\log_{a}x-\mu)\right]^{k-1}\exp\left[-\alpha(\log_{a}x-\mu)\right]$ where $g=\log_{a}e$	$e^{\mu/g}\le x<\infty$ if $\alpha>0$ $0\le x\le e^{\mu/g}$ if $\alpha<0$	3
KAP	$f_{\mathrm{KAP}}(\mu,\alpha,k,h)=\alpha^{-1}[1-k(x-\mu)/\alpha]^{1/k-1}[F_{\mathrm{KAP}}(x)]^{1-h}$ where $F_{\mathrm{KAP}}(x)=(1-h(1-k(x-\mu)/\alpha)^{1/k})^{1/h}$	$\mu+\alpha(1-h^{-k})/k\le x\le\mu+\alpha/k$ if $h>0, k>0$ $\mu+\alpha(1-h^{-k})/k\le x<\infty$ if $h>0, k<0$ $-\infty<x<\mu+\alpha/k$ if $h\le 0, k>0$ $\mu+\alpha/k\le x<\infty$ if $h\le 0, k<0$	4

μ: location parameter
α: scale parameter
k: shape parameter
h: second shape parameter (GG, KAP)
$\Gamma(\,)$: gamma function

Table 2. List of mixture probability density functions, domains.

Name	Probability density function	Domain
MGG	$f_{\mathrm{GG}}(x;\alpha_1,k_1,\alpha_2,k_2)=\omega f_{\mathrm{G}}(x;\alpha_1,k_1)+(1-\omega)\,f_{\mathrm{G}}(x;\alpha_2,k_2)$	$0<x<+\infty$
MGW	$f_{\mathrm{GW}}(x;\alpha_1,k_1,\alpha_2,k_2)=\omega f_{\mathrm{G}}(x;\alpha_1,k_1)+(1-\omega)\,f_{\mathrm{W}}(x;\alpha_2,k_2)$	$0<x<+\infty$
MGE	$f_{\mathrm{GE}}(x;\alpha_1,k,\mu,\alpha_2)=\omega f_{\mathrm{G}}(x;\alpha_1,k)+(1-\omega)\,f_{\mathrm{E}}(x;\mu,\alpha_2)$	$-\infty<x<+\infty$
MGTN	$f_{\mathrm{GN}}(x;\alpha_1,k,\mu,\alpha_2)=\omega f_{\mathrm{G}}(x;\alpha_1,k)+(1-\omega)\,f_{\mathrm{TN}}(x;\mu,\alpha_2)$	$0\le x<+\infty$
MWW	$f_{\mathrm{WW}}(x;\alpha_1,k_1,\alpha_2,k_2)=\omega f_{\mathrm{W}}(x;\alpha_1,k_1)+(1-\omega)\,f_{\mathrm{W}}(x;\alpha_2,k_2)$	$0<x<+\infty$
MWE	$f_{\mathrm{WE}}(x;\alpha_1,k,\mu,\alpha_2)=\omega f_{\mathrm{W}}(x;\alpha_1,k)+(1-\omega)\,f_{\mathrm{E}}(x;\mu,\alpha_2)$	$-\infty<x<+\infty$
MEE	$f_{\mathrm{EE}}(x;\mu_1,\alpha_1,\mu_2,\alpha_2)=\omega f_{\mathrm{E}}(x;\mu_1,\alpha_1)+(1-\omega)\,f_{\mathrm{E}}(x;\mu_2,\alpha_2)$	$-\infty<x<+\infty$
METN	$f_{\mathrm{EN}}(x;\mu_1,\alpha_1,\mu_2,\alpha_2)=\omega f_{\mathrm{E}}(x;\mu_1,\alpha_1)+(1-\omega)\,f_{\mathrm{TN}}(x;\mu_2,\alpha_2)$	$-\infty<x<+\infty$
MTNTN	$f_{\mathrm{NN}}(x;\mu_1,\alpha_1,\mu_2,\alpha_2)=\omega f_{\mathrm{TN}}(x;\mu_1,\alpha_1)+(1-\omega)\,f_{\mathrm{TN}}(x;\mu_2,\alpha_2)$	$0\le x<+\infty$

Table 3. Wind speed characteristics at the 10 stations selected to illustrate the results for the province of Quebec.

Station number	Station name	Period	Lat (°)	Long (°)	Alt (m)	Calm (%)	Median (m/s)	CV (-)	Skewness (-)	Kurtosis (-)
1	Quebec	1953/01-2017/10	46.79	-71.39	74.40	7.61	3.6	0.69	0.75	3.57
2	Montreal	1953/01-2017/10	45.52	-73.42	27.40	6.39	4.2	0.63	0.73	3.78
3	Cape Whittle	1995/01-2017/10	50.16	-60.06	7.00	0.35	6.9	0.53	0.79	3.68
4	Cap-Madeleine	1994/01-2017/10	49.25	-65.32	29.00	1.51	5.3	0.62	0.79	3.53
5	Kuujjuarapik	1957/01-2017/10	55.28	-77.75	12.20	5.24	4.4	0.60	0.65	3.55
6	Mont-Joli	1953/01-2017/10	48.60	-68.22	52.40	3.98	4.7	0.59	0.66	3.39
7	Bagotville	1953/01-2017/10	48.33	-71.00	159.10	7.72	3.6	0.68	0.69	3.25
8	Val-D’or	1955/01-2010/12	48.06	-77.79	337.40	6.69	3.1	0.62	0.61	3.37
9	Nitchequon	1959/01-1985/12	53.20	-70.90	536.10	5.61	3.9	0.63	0.72	3.68
10	Parc National des Pingualuit	2011/01-2017/10	61.31	-73.67	503.40	4.27	5.6	0.64	0.52	2.98

CV denotes coefficient of variation

Table 4. Ranked pdfs giving the best fit for each goodness-of-fit statistic and at each of the 10 selected stations to illustrate the results.

Station name	Rank	RMSE	R_F^2	R_p^2	Chi-2	KS
Quebec	1	MEE/ML (0.0068)	MGG/ML (0.9997)	MEE/ML (0.9843)	MEE/ML (10173)	MGG/ML (0.013)
	2	MGG/ML (0.0078)	MEE/ML (0.9996)	MGG/ML (0.9795)	MGG/ML (10847)	MEE/ML (0.016)
	3	MGW/ML (0.0088)	MGW/ML (0.9995)	MGW/ML (0.9741)	MGW/ML (14105)	MGE/ML (0.019)
	4	MGTN/ML (0.0096)	MGE/ML (0.9994)	MGTN/ML (0.9693)	MGE/ML (15586)	MGW/ML (0.019)
Montreal	1	MGE/ML (0.0090)	MGW/ML (0.9996)	MGE/ML (0.9695)	MGE/ML (12443)	MTNTN/ML (0.016)
	2	MGW/ML (0.0091)	MWTN/ML (0.9996)	MGW/ML (0.9691)	MGG/ML (12487)	MGW/ML (0.017)
	3	MGG/ML (0.0092)	MGG/ML (0.9996)	MGG/ML (0.9685)	MWTN/ML (12650)	MGE/ML (0.017)
	4	MWTN/ML (0.0093)	MWW/ML (0.9996)	MWTN/ML (0.9672)	MGW/ML (12766)	MWTN/ML (0.017)
Cape Whittle	1	MWE/ML (0.0057)	MGW/ML (0.9999)	MWE/ML (0.9779)	MWE/ML (3626)	MTNTN/ML (0.009)
	2	MGW/ML (0.0057)	MWE/ML (0.9999)	MGW/ML (0.9774)	MGW/ML (3655)	MGE/ML (0.011)
	3	MGE/ML (0.0059)	METN/ML (0.9999)	MGE/ML (0.9759)	MGG/ML (3903)	MGG/ML (0.011)
	4	MGE/LS (0.0060)	MWTN/ML (0.9999)	MGE/LS (0.9756)	MGE/ML (3952)	MGW/ML (0.012)
Cap-Madeleine	1	MWE/LS (0.0076)	MGE/ML (0.9996)	MWE/LS (0.9722)	MWE/LS (5269)	MGE/ML (0.016)
	2	MGE/ML (0.0090)	MGW/ML (0.9993)	MGE/ML (0.9614)	MGE/ML (5574)	MGG/ML (0.019)
	3	MEE/LS (0.0090)	MGG/ML (0.9993)	MEE/LS (0.9611)	MEE/LS (6245)	MGW/ML (0.022)
	4	MGE/LS (0.0102)	MWE/ML (0.9992)	MGE/LS (0.9509)	MGG/ML (7012)	MWE/LS (0.027)
Kuujjuarapik	1	MWW/ML (0.0069)	MGG/ML (0.9998)	MWW/ML (0.9808)	MGG/ML (8168)	MGG/ML (0.014)
	2	MGG/ML (0.0074)	MWW/ML (0.9997)	MGG/ML (0.9778)	MEE/ML (8823)	MWE/ML (0.016)
	3	MEE/ML (0.0076)	MGE/ML (0.9997)	MEE/ML (0.9767)	MGE/ML (8855)	MGE/ML (0.017)
	4	MGE/ML (0.0076)	MWE/ML (0.9997)	MGE/ML (0.9766)	MWW/ML (9583)	MEE/ML (0.017)
Mont-Joli	1	MGG/ML (0.0084)	MWTN/ML (0.9997)	MGG/ML (0.9690)	MGG/ML (12851)	MGG/ML (0.013)
	2	MGE/ML (0.0086)	MGTN/ML (0.9997)	MGE/ML (0.9674)	MWE/ML (12896)	MGTN/ML (0.016)
	3	MGTN/ML (0.0091)	MGG/ML (0.9997)	MGTN/ML (0.9631)	MGE/ML (13438)	MGW/ML (0.016)
	4	MGW/ML (0.0091)	MWW/ML (0.9997)	MGW/ML (0.9630)	MWW/ML (13984)	MWTN/ML (0.018)
Bagotville	1	MWTN/ML (0.0072)	MWTN/ML (0.9997)	MWTN/ML (0.9823)	MWTN/ML (8648)	MWTN/ML (0.015)
	2	MTNTN/ML (0.0076)	MTNTN/ML (0.9997)	MTNTN/ML (0.9804)	MTNTN/ML (9721)	MTNTN/ML (0.016)
	3	MEE/ML (0.0083)	METN/ML (0.9996)	MEE/ML (0.9768)	MEE/ML (11044)	MGG/ML (0.018)
	4	METN/ML (0.0085)	MGW/ML (0.9996)	METN/ML (0.9754)	METN/ML (11393)	MGW/ML (0.019)
Val-D’or	1	MEE/ML (0.0097)	MGE/ML (0.9995)	MEE/ML (0.9796)	MGE/ML (6179)	MEE/ML (0.017)
	2	MGE/ML (0.0101)	MEE/ML (0.9995)	MGE/ML (0.9777)	MEE/ML (6544)	MGE/ML (0.021)
	3	MGG/ML (0.0134)	MGW/ML (0.9993)	MGG/ML (0.9611)	MGG/ML (10973)	MGG/ML (0.022)
	4	MGW/ML (0.0135)	MGG/ML (0.9992)	MGW/ML (0.9606)	MGW/ML (11446)	MGW/ML (0.025)
Nitchequon	1	MWW/ML (0.0064)	MWW/ML (0.9998)	MWW/ML (0.9845)	MWW/ML (2681)	MWW/ML (0.013)
	2	MGW/ML (0.0066)	MGW/ML (0.9998)	MGW/ML (0.9835)	MGW/ML (3195)	MGTN/ML (0.015)
	3	MGTN/ML (0.0066)	MGTN/ML (0.9998)	MGTN/ML (0.9833)	MGTN/ML (3254)	MGW/ML (0.015)
	4	MWTN/ML (0.0071)	MGG/ML (0.9997)	MWTN/ML (0.9809)	MGG/ML (3852)	MWTN/ML (0.016)
Parc National des Pingualuit	1	MWW/ML (0.0047)	MWW/ML (0.9999)	MWW/ML (0.9866)	MWW/ML (459)	MWW/ML (0.009)
	2	MGG/ML (0.0058)	MGG/ML (0.9998)	MGG/ML (0.9796)	MGG/ML (592)	MGG/ML (0.009)
	3	MEE/ML (0.0060)	METN/ML (0.9997)	MEE/ML (0.9787)	MEE/ML (709)	MWE/ML (0.012)
	4	MGE/ML (0.0068)	MWE/ML (0.9997)	MGE/ML (0.9723)	MGE/ML (831)	MGE/ML (0.014)

The corresponding goodness-of-fit statistic value is display in parenthesis after the pdf name.

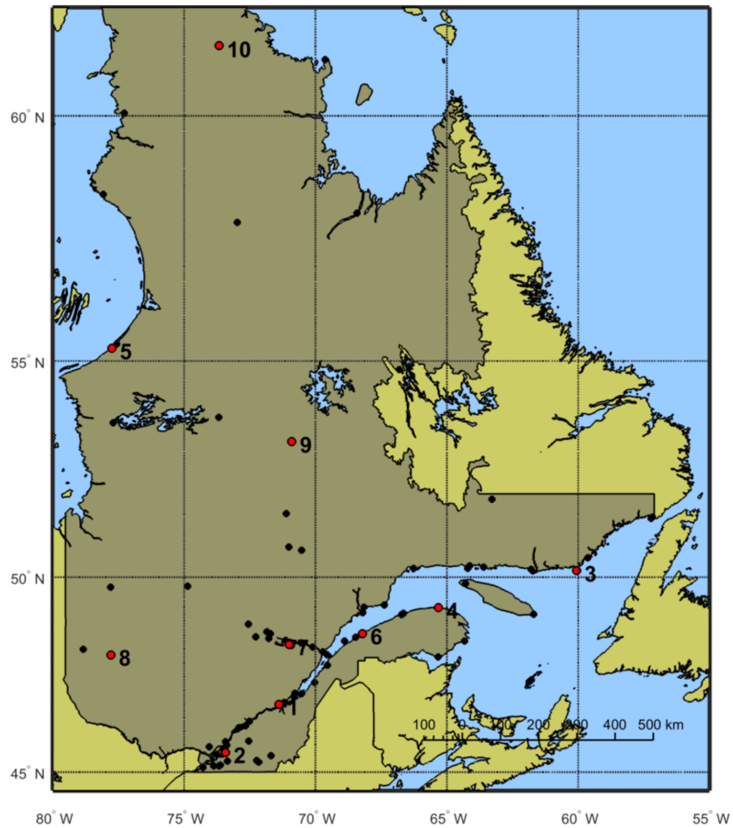

Figure 1. Spatial distribution of the 83 meteorological stations in the province of Quebec. The 10 stations selected to illustrate the results are represented by red dots.

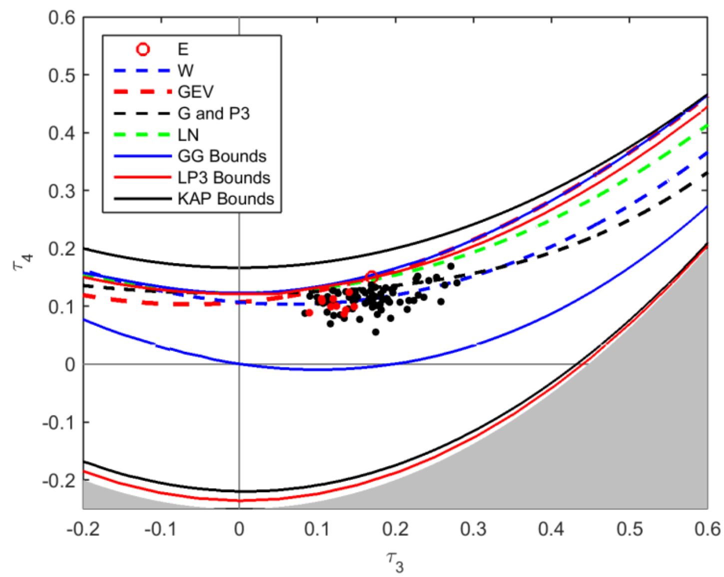

Figure 2. L-moment ratio diagram with the selected pdfs. Sample L-moments are represented by black dots for all stations. Red dots denote the sample L-moments of the 10 stations selected to illustrate the resuls in the rest of the paper.

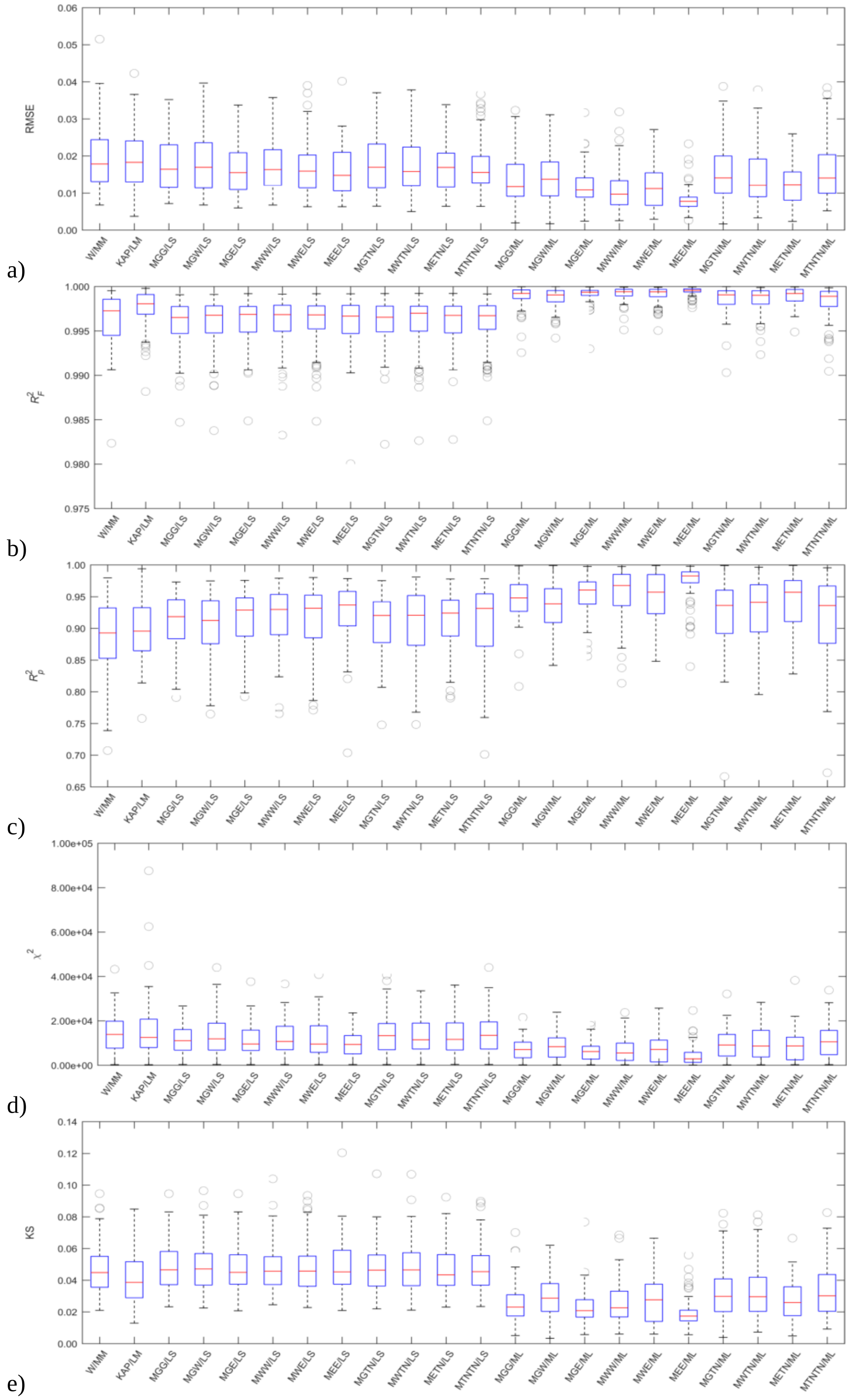

Figure 3. Box plots of the goodness-of-fit statistics.

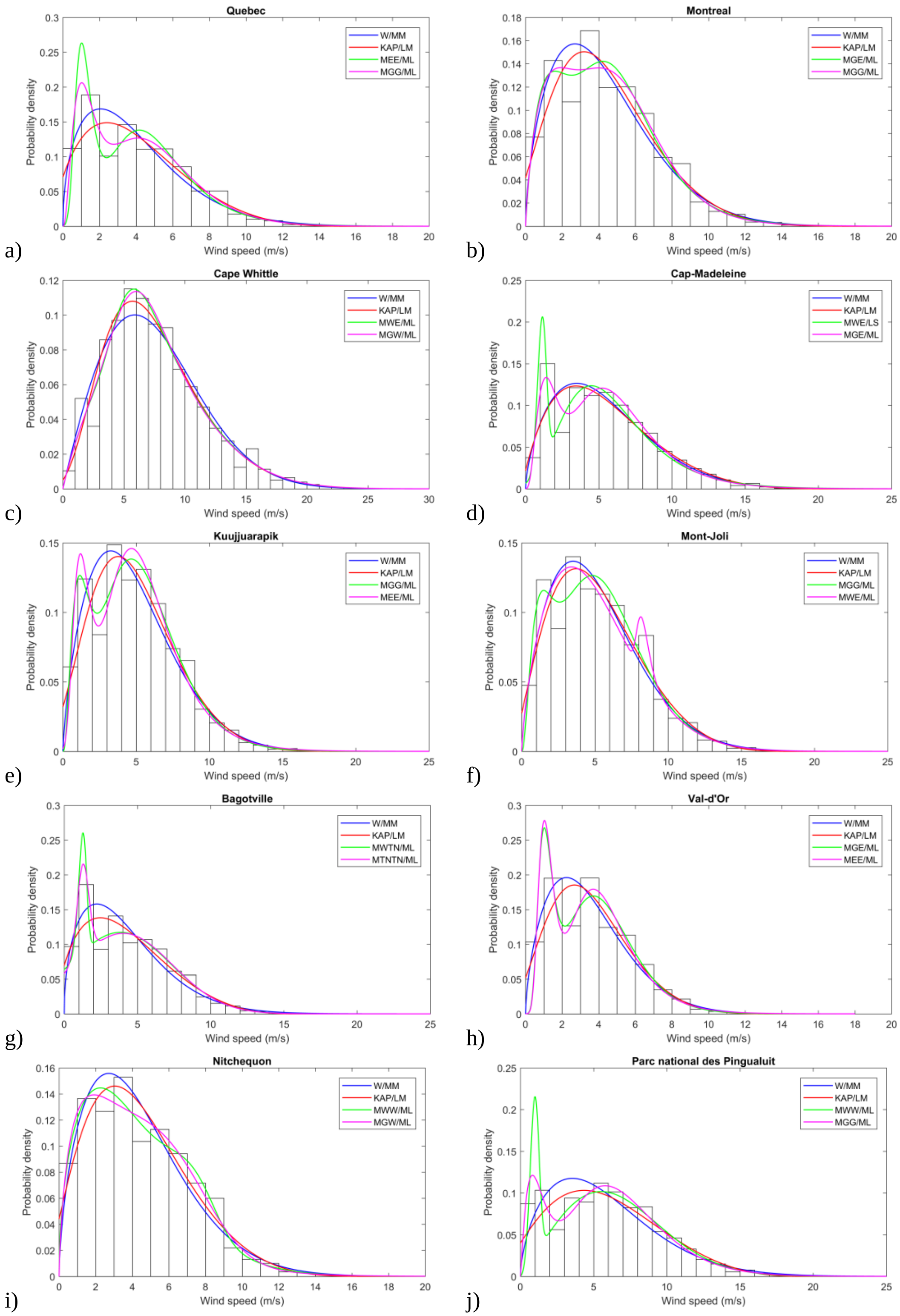

Figure 4. Wind speed frequency histograms for the 10 stations selected to illustrate the results. The two mixture distributions giving the best fit with respect to χ^2, and W and KAP are superimposed.