Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 7 (2021) 248–255
www.elsevier.com/locate/egyr

Tmrees, EURACA, 28 to 30 May 2021, Athens, Greece

Probability distributions for wind speed volatility characteristics: A case study of Northern Norway

Hao Chen[a,*], Stian Normann Anfinsen[b], Yngve Birkelund[b], Fuqing Yuan[a]

[a] *Department of Technology and Safety, UiT The Arctic University of Norway, Tromsø 9019, Norway*
[b] *Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø 9019, Norway*

Received 27 July 2021; accepted 28 July 2021

Abstract

The Norwegian Arctic is rich in wind resources. The development of wind power in this region can boost green energy and also promote local economies. In wind power engineering, it is a tremendous advantage to base projects on a sound understanding of the intrinsic properties of wind resources in an area. Wind speed volatility, a phenomenon that strongly affects wind power generation, has not received sufficient research attention. In this paper, a framework for studying short-term wind speed volatility with statistical analysis and probabilistic modeling is constructed for an existing wind farm in Northern Norway. It is found that unlike the characteristics of wind power volatility, wind speed volatility cannot be described by the normal distribution. The reason is that even though the probability distribution of wind speed volatility is centrally symmetric, it is much more centrally concentrated and has thicker tails. After comparing three distributions corresponding to different sampling periods, this paper suggests utilizing the *t* distribution, with average modeling RMSE less than 0.006 and R^2 exceeding 0.995 and with the best modeling scenario of temporal resolution, the 30 mins has an RMSE of 0.0051 and an R^2 of 0.997, to more accurately and effectively explore the fluctuating characteristics of wind speed.

Peer-review under responsibility of the scientific committee of the Tmrees, EURACA, 2021.

Keywords: Wind energy; Wind speed volatility; Statistical analysis; Probability distribution; Arctic

* Corresponding author.
E-mail address: hao.chen@uit.no (H. Chen).

https://doi.org/10.1016/j.egyr.2021.07.125

Peer-review under responsibility of the scientific committee of the Tmrees, EURACA, 2021.

Nomenclature

PDF	Probability density function
CDF	Cumulative distribution function
SV	Wind speed volatility
SP	Wind speed
γ	Skewness
κ	Excess kurtosis
MLE	Maximum likelihood estimation
RMSE	Root mean square error
R^2	Coefficient of determination

1. Introduction

As an alternative to fossil fuels, wind energy has received increasing attention worldwide because of its abundant availability, widespread dispersal, and potential financial support [1]. Norway owns some of the best wind energy resources in Europe [2]. It has enormous potential for wind power generation, especially in its northern and Arctic regions.

Assessing potential wind resources – typically evaluated by measured and modeled wind speed and direction through a year or more at a certain location – are critical for evaluating the feasibility and sustainability of a wind energy project [3]. Wind is a phenomenon involving air movement and relates to the atmospheric motion state. Changes in wind characteristics are closely related to the circulation of energy and matter in the atmosphere. The most noticeable difference between wind energy and conventional energy is the volatility, stochasticity, intermittency, and uncontrollability of the former [4]. The changes in wind speed are affected by long-term atmospheric motion and micro-scale atmospheric turbulence caused by many surface factors. These cause the wind to show strong instantaneous volatility in time and space. Due to the uncertainty and intermittency of wind, wake effects between wind turbines, and the cubic relationship between wind speed and the wind turbine-generated power, a small change of wind speed can be significantly amplified in the output wind power. The random volatility of wind is regarded as an adverse factor for wind energy [5]. This intermittency brings severe challenges to the power system's safety, power quality, and the balance of power supply and demand. Therefore, studying the volatility characteristics of wind is of great significance for improving wind power forecasting accuracy, scenario generations, and overcoming the adverse effects of wind power integration in the grid [6].

However, the typical wind energy assessment methodology lacks tools to characterize wind speed volatility on sites. The volatility analysis offers additional information about wind. The wind has different volatility characteristics at different temporal scales. Although the wind has certain seasonal and diurnal characteristics, there is no fixed volatility amplitude and cycle; its volatility has no clear rules to follow.

The probability density function (PDF) is an effective quantification to describe wind randomness and uncertainty [7]. Much research has used the probability density function in wind engineering [8]. However, most of the research concerns evaluation of historical wind speed distribution. To illustrate, Mahmood and colleagues [9] used the Weibull distribution to assess wind speed data from a site in Iraq successfully.

Studies centered on statistical analyses of volatility in wind energy, and those who exist have mainly considered wind power volatility directly are few and far between, although a handful exists. For instance, Lange [10] analyzed the uncertainty in wind power prediction using the statistical distributions and found that wind speed prediction error is normally distributed. Bludszuweit [11] looked into the statistical distributions of wind power errors forecasted by the persistence model. It proposed an indirect algorithm based on the Beta distribution based on one-year measured data from two different wind farms. Zhang [12] presented a versatile distribution for fitting wind power predictive errors and compared the distribution with benchmarking distributions of normal and Beta. Inspired by probability distributions of wind power volatility, it is also possible to use statistics to analyze the wind speed volatility using classical ideal distribution functions to model the histogram of the wind volatility and capture its nature.

This paper uses different probability density functions and skewness and kurtosis moments to characterize short-term wind speed volatility at various temporal scales for a wind farm in Northern Norway. The statistical modeling of volatility assists in documenting wind's internally volatile features, especially for the wind in a cold climate and complex terrain.

2. Data preparation

This paper draws on data from a wind power station located in the Norwegian Arctic, whose coordinates are 70°5′56″N, 20°3′54″E, and its designed capacity is 54 MW. The hub height of the turbines is 80 m above the ground, and the rotor diameter is 90 m. The farm has eighteen Vestas V90-3.0 3.0 MW turbines with 45 m long rotor and the hub height is 80 m above the ground. The wind farm is surrounded by hills and fronts a fjord. The wind park company provides measurements, taken by the wind mast with the same height of turbines, of wind speed. The original wind speed data are from 0:00 on 1st January 2017 to 23:00 on 31st December 2017 with 10 min temporal resolution. The number of measured data points is 52,560. Wind speed data with a reduced temporal resolution of 30 min and 60 min are obtained by interpolations. The size of the dataset with 30 min and 60 min resolution is 17,520 and 8,760 data points, respectively. The wind Speed Volatility (SV) is calculated as the first-order differential by Eqs. (1):

$$SV_i = SP_i - SP_{i-1} \tag{1}$$

where SP_i and SP_{i-1} are wind speed at time t and one temporal resolution before t_i.

3. Methodology

The sample skewness (γ) and sample excess kurtosis (κ) are common shape-parameters that describe the historical distributions of variables, and they are defined as:

$$\gamma = T^{-1} \sum_{t=1}^{T} \left(X_t - \overline{X}\right)^3 / s^3 \tag{2}$$

$$\kappa = T^{-1} \sum_{t=1}^{T} \left(X_t - \overline{X}\right)^4 / s^4 - 3 \tag{3}$$

where T is the size of the data sample, $\overline{X}$ is the sample mean, and s is the sample standard division. γ measures whether the PDF of a random variable "leans" to one side of the mean. A distribution is left-skewed when γ is negative and right-skewed when γ is positive. κ measures the "peakedness" of a distribution. The so-called excess kurtosis defined in Eq. (3) is measured relative to the normal distribution, which attains a value of $\kappa = 0$. Therefore, excess kurtosis is a measure of departure from normality and reflects the sharpness of the peak [13]. A distribution is leptokurtic when $\kappa > 0$, indicating that the PDF is sharper and steeper than the normal distribution, and it is platykurtic when $\kappa < 0$.

The PDF of a random variable is a statistical model that describes the probability of occurrence of this variable at a specific point in each observation interval. The cumulative distribution function (CDF) specifies the probability that the variable is less than or equal to a specific value [14]. In this section, three commonly used ideal PDFs are chosen as the candidates for modeling the SV.

For the normal distribution, its PDF (4) and CDF (5) are expressed by:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{4}$$

$$F(x;\mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{5}$$

where μ is the mean, σ is the standard division, and *erf* (.) is the error function.

The logistic distribution resembles the normal distribution in shape but has heavier tails (higher κ). The PDF (6) and CDF (7) of the logistic distribution are given [15], respectively, by:

$$f(x;\mu,s) = \frac{e^{-(x-\mu)/s}}{s\left(1 + e^{-(x-\mu)/s}\right)^2} \tag{6}$$

$$F(x;\mu,s) = \frac{1}{1 + e^{-(x-\mu)/s}} \tag{7}$$

where μ is a location parameter and s is a scale parameter. The mean equals μ, and the variance is $s^2\,\pi^2/3$.

The PDF (8) and CDF (9) of the t distribution are determined via the following functions [16]:

$$f(x;\nu)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{\nu\pi}}\frac{1}{\left(1+\frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}} \tag{8}$$

$$F(x;\nu)=\int_{-\infty}^{x}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{\nu\pi}}\frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}dt \tag{9}$$

where $\nu > 0$ is the number of degrees of freedom and $\Gamma(.)$ is the Gamma function.

Since histograms are discrete distributions, a nonparametric method of simulating distributions based on the data itself, the kernel distribution, can approximate discrete historical distributions to the empirical distribution of samples taken at infinitely small intervals. Figuratively, it is called smoothing PDFs and is determined by a smoothing function and a bandwidth. In this study, the smoothing function is the Gaussian function, and the bandwidth values 0.025, which can extract wind speed information with high precision and without adding sampling noise.

3.1. Parameter estimation

The PDF parameter estimation means an ideal probability distribution model can statistically describe the distribution of SV data. The parameters of the model are estimated by training the SV data with proper estimation approaches. This study uses the Maximum Likelihood Estimation (MLE) approach to determine parameters for the above three PDFs.

4. Experiments

The procedure for modeling the PDF of SV at different temporal scales is illustrated in Fig. 1. The raw wind speed data are interpolated and calculated by Eq. (1) to create SF data sequences for different temporal scales. These data are then tested for their normality, and their histograms are plotted. Moreover, their smoothing PDFs are created by the kernel distribution. Then, fitted distribution models corresponding to all SF datasets on different temporal scales are created, whose parameters are obtained with the MLE method. Finally, the fitted PDF models are tested with the goodness-of-fit and compared with the corresponding smoothing PDFs.

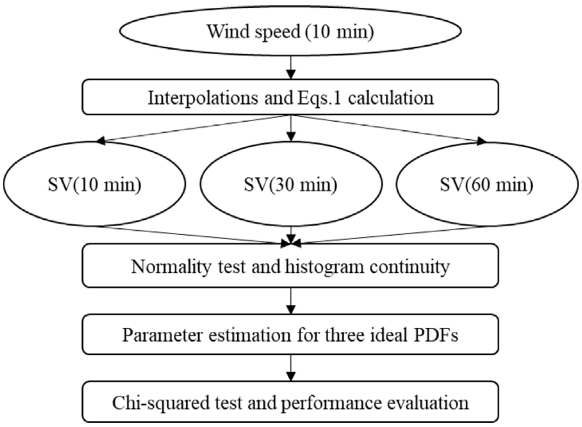


Fig. 1. Procedure for the SV probabilistic modeling.

Pearson's chi-square goodness-of-fit test for PDF models is a nonparametric test that evaluates how likely a data sample has been drawn from a given PDF [17]. The chi-square test divides data into k bins and defines the following null hypothesis: H_0: $\{X_1, X_2, \ldots, X_n\}$ follows the given probability distribution. The alternative hypothesis is: H_1: $\{X_1, X_2, \ldots, X_n\}$ do not follow this distribution. The test statistic is defined by Eqs. (10):

$$\chi^2=\sum_{i=1}^{k}(O_i-E_i)^2/E_i \tag{10}$$

where O_i is the observed count and E_i is the expected count for bin i based on the hypothesized PDF.

To evaluate the performance of different PDFs for SV modeling, the Root Mean Square Error (RMSE) and the coefficient of determination (R^2) are applied to calculate the probability density difference between smoothing PDFs and corresponding fitted PDF models. RMSE is a negatively oriented metric, meaning that smaller values indicate better fitting performance. Meanwhile, the second is positively oriented, and its range is zero between one.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(smoothing_i - modeling_i\right)^2}{n}} \tag{11}$$

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{12}$$

where n is the total number of sampling by the kernel function with 0.025 m/s bandwidth that is related to SV ranging from −10 m/s to 10 m/s, and it equals 800. SS_{res} is the sum of squares of residuals (deviations fitted from smoothing PDFs based on histograms) and SS_{tot} is the total sum of squares (overall squared differences between the smoothing PDF values at the sampling points and their averages).

5. Results and discussion

We use three different PDFs, the normal distribution, the logistic distribution, and the *t* distribution, to model the volatility of wind speed over various temporal intervals for the wind farm in Northern Norway. The results are presented as follows.

5.1. Statistics for SV data

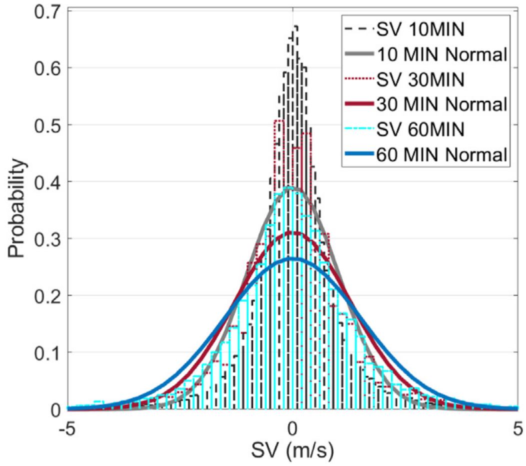
Table 1. The statistics of SV data at different temporal scales.

Temporal resolution (min)	Mean (m/s)	Standard deviation (m/s)	Min (m/s)	Max (m/s)	Skewness	Kurtosis
10	0.0000	1.0200	−9.6000	16.2000	0.3741	8.2444
30	−0.0003	1.2808	−9.7000	11.6333	0.2429	5.1468
60	−0.0007	1.5034	−10.4167	12.2833	0.1930	4.4025

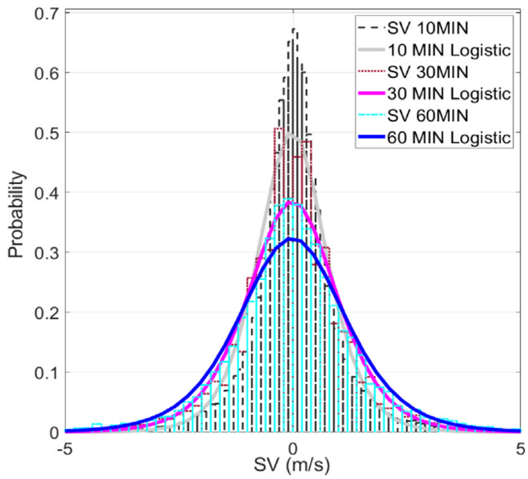
The descriptive statistics for SV data are shown in Table 1. The mean value of the SV data is very close to zero at all temporal resolutions, which indicates that the wind speed volatility is, in general, trendless and oscillates back and forth around the zero points. As the sampling time grows, the SV data standard deviation increases, and their γ and κ decrease. The increase in standard deviation is understandable since SV is more variable over more extended periods. The γ of all three SV datasets is slightly positive, which means the right tails of the distributions are longer than the left ones, and their mass is concentrated slightly to the left. Both γ and κ decrease with time spacing, and so the data become increasingly normal. The negative correlation of the γ with sampling time indicates that the histogram of the SV data becomes more symmetrical as the time spacing increases. The three SV datasets have positive κ, which shows that all of them are leptokurtic and morphologically steeper or thicker tails than the normal distribution. Large κ values can occur in two situations: the probability mass is concentrated near the mean, and occasionally, there are some data in the dataset that are away from the mean, or the mass of probability is concentrated at the tails of the distribution. κ values that increase with temporal resolution also illustrate the decline in the concentration and the size of extreme values away from the means of SV datasets. Therefore, based on the above analysis, it is reasonable to assume that merely using the normal distribution to describe SV is inaccurate.

5.2. PDF modeling fitting and test

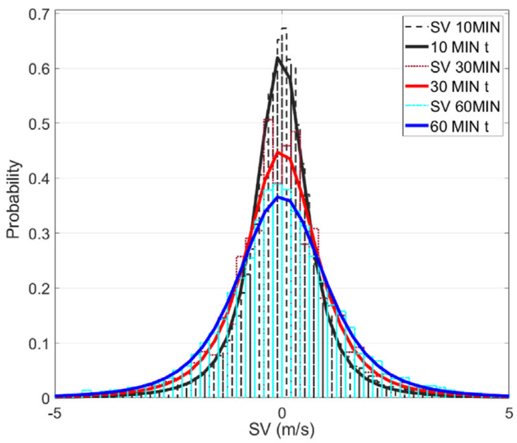
Fig. 2 shows histograms of SV data and fitted model PDFs with parameters that have been estimated by the MLE approach. From Fig. 2, it can be seen visually that the mode PDF value falls significantly as the sampling time increases. The *t* and logistic distributions fit the shape of histograms better than the normal distribution for all three temporal resolution cases.



(a)



(b)



(c)

Fig. 2. The histograms and estimated PDFs curve graph for SV data ((a) is normal, (b) is logistic, and (c) is *t* distribution).

Table 2. The parameters for PDFs and p values of the chi-square test (Values less than 10^{-8} are approximately expressed as zero).

Temporal resolution (min)	Normal (μ, σ)	Logistic (μ,s)	t (ν)	Normal p	Logistic p	t p
10	(0,1.02)	(0,0.50)	2.50	0	0	0
30	(0,1.28)	(0,0.65)	3.00	0	0	0
60	(0,1.50)	(0,0.77)	3.31	0	0	0.062

The standard deviation of the normal distribution, the scale parameter of logistic distribution, and the degrees of freedom of t distribution are shown in Table 2 to correlate positively with the time resolution, proving that curves of all three distributions become lower and broader.

Pearson's chi-square test is a rigorous statistical test. According to the test, it can be concluded whether there exists a statistically significant difference between a theoretical distribution model and the observed frequency distribution of specified discrete events in the data sample. The hypothesis tests areas above Section 5.2 and with a significance level of 5%. The p values of chi-square tests are also shown in Table 2. Only the p-value for the t distribution corresponding to the SV data with 30 min is above 0.05, indicating that the dataset statistically follows the t distribution with a degree of freedom equals 3.31. Given that rigorous statistical tests do not give a complete picture of the accuracy of probabilistic models. We will introduce quantitative analysis to evaluate these models in the following sub-section.

5.3. Performance evaluation

Table 3. RMSE and R^2 of different PDF models.

Temporal resolution (min)	RMSE			R^2		
	Normal	Logistic	t	Normal	Logistic	t
10	0.0401	0.0226	0.0058	0.8738	0.9596	0.9971
30	0.0286	0.0150	0.0051	0.9106	0.9748	0.9970
60	0.0236	0.0121	0.0055	0.9241	0.9795	0.9954

Real-world data will often have problems with passing a rigorous statistical test. In engineering practice, evaluation metrics from regression analysis are commonly adopted to assess the quality of PDF modeling. The RMSE and R^2 between empirical or smoothing PDFs of SV data and different fitted PDF models for various temporal resolutions are shown in Table 3. It is found that although most of the PDF models do not pass the chi-square test, the R^2 of all logistic and t distribution models surpasses 0.95, which generally means that these PDFs provide a sound fit. Except for the normal distribution, the other two distributions can display probabilistic characteristics of the SV dataset. Regarding performance differences between different PDF models, the t distribution is superior to other distributions for all sampling time datasets in both RMSE and R^2. Almost all PDF curves are highly centrally concentrated and have heavy and long tails, which potentially embodies the risk of wind ramp events. The t distribution satisfactorily embodies these features. Besides, the logistic distribution performs better than the normal distribution in all cases, suggesting that it can also deliver relatively satisfactory probabilistic modeling for describing SV.

Concerning the comparison of various time resolutions, the RMSE and R^2 of normal and logistic distributions respectively decrease and increase with the sampling time. This demonstrates that both PDFs more easily characterize the SV data's statistical distributions with the rising sampling time. Meanwhile, the RMSE and R^2 of the t distribution are very stable and do not fluctuate much with sampling interval slightly volatile features. Overall, the t distribution is proven to be a more desirable probabilistic model to represent wind speed volatility in comparison with the normal logistic distributions.

6. Conclusion

Statistical characterization of wind volatility is vital to effectively conduct practical assessments of wind resources for wind power development. In the present paper, we focus on statistical modeling of wind speed volatility for a wind farm inside the Norwegian Arctic region. Based on the statistical analysis and PDFs modeling results, the following conclusions can be drawn.

The probability distribution of wind volatility is overall centrally symmetrical but quite different from the normal distribution. In our cases, wind volatility is slightly left-skewed and has sharper peaks compared to the normal distribution. However, as the temporal resolution of sampling decreases, its probability distribution becomes closer to the normal distribution. Although most PDF models fail a rigorous nonparametric goodness-of-fit test based on the raw data of complex wind phenomena, the logistic and *t* distributions deliver R^2 exceeding 0.95 and RMSE approaching zero, suggesting that both distributions provide good characterizations of wind speed short-term volatility in wind energy engineering practice. Moreover, the *t* distribution has a notable advantage, and its performance is very stable with sampling time. Therefore, this paper recommends explicitly applying the *t* distribution to modeling wind speed volatility based on our results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study is supported by the Department of Technology and Safety, UiT The Arctic University of Norway. We extend our gratitude to Dr. Reidar Staupe-Delgado for his comments on an earlier version of this manuscript. We also thank Troms Kraft AS for kindly offering the measured wind data.

References

[1] Samadianfard S, et al. Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm. Energy Rep 2020;6:1147–59.
[2] Blindheim B. Implementation of wind power in the Norwegian market; the reason why some of the best wind resources in Europe were not utilised by 2010. Energy Policy 2013;58:337–46.
[3] Ferreira M, Santos A, Lucio P. Short-term forecast of wind speed through mathematical models. Energy Rep 2019;5:1172–84.
[4] Notton G, et al. Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. Renew Sustain Energy Rev 2018;87:96–105.
[5] Yang T. Optimal sizing of the hybrid energy storage system aiming at improving the penetration of wind power. In: 2016 IEEE PES Asia-Pacific power and energy engineering conference (APPEEC). IEEE; 2016, p. 2358–62.
[6] Moghaddam SZ. Generation and transmission expansion planning with high penetration of wind farms considering spatial distribution of wind speed. Int J Electr Power Energy Syst 2019;106:232–41.
[7] Li J-N, Qiao Y, Lu Z-x, Li J, Xu F. Research on statistical modeling of large-scale wind farms output volatilitys in different spacial and temporal scales. Power Syst Prot Control 2012;40(19):7–13.
[8] Jung C, Schindler D. Wind speed distribution selection–A review of recent development and progress. Renew Sustain Energy Rev 2019;114:109290.
[9] Mahmood FH, Resen AK, Khamees AB. Wind characteristic analysis based on Weibull distribution of Al-Salman site, Iraq. Energy Rep 2020;6:79–87.
[10] Lange M. On the uncertainty of wind power predictions—Analysis of the forecast accuracy and statistical distribution of errors. J Sol Energy Eng 2005;127(2):177–84.
[11] Bludszuweit H, Domínguez-Navarro JA, Llombart A. Statistical analysis of wind power forecast error. IEEE Trans Power Syst 2008;23(3):983–91.
[12] Zhang Z-S, Sun Y-Z, Gao DW, Lin J, Cheng L. A versatile probability distribution model for wind power forecast errors and its application in economic dispatch. IEEE Trans Power Syst 2013;28(3):3114–25.
[13] Joanes D, Gill C. Comparing measures of sample skewness and kurtosis. J R Stat Soc D 1998;47(1):183–9.
[14] Feller W. An introduction to probability theory and its applications, Vol. 2. John Wiley & Sons; 2008.
[15] Di Crescenzo A, Martinucci B. A damped telegraph random process with logistic stationary distribution. J Appl Probab 2010;47(1):84–96.
[16] Theodossiou P. Financial data and the skewed generalized *t* distribution. Manage Sci 1998;44(12-part-1):1650–61.
[17] Brodsky E, Darkhovsky BS. Nonparametric statistical diagnosis: Problems and methods. Springer Science & Business Media; 2013.