



Statistical Properties of Electricity Generation from a Large System of Wind Plants and Demand for Fast Regulation*

Dubravko Sabolić^{†1}, Alan Župan², Roman Malarić³

¹Croatian Transmission System Operator Ltd., Kupska 4, Zagreb, Croatia
e-mail: dubravko.sabolic@hops.hr

²Croatian Transmission System Operator Ltd., Kupska 4, Zagreb, Croatia
e-mail: alan.zupan@hops.hr

³Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, Zagreb, Croatia
e-mail: roman.malari@fer.hr

Cite as: Sabolić, D., Župan, A., Malarić, R., Statistical Properties of Electricity Generation from a Large System of Wind Plants and Demand for Fast Regulation, *J. sustain. dev. energy water environ. syst.*, 5(3), pp 447-465, 2017, DOI: <http://dx.doi.org/10.13044/j.sdewes.d5.0156>

ABSTRACT

Experimental data of total wind generation, recorded at 5 minute intervals and published by the Bonneville Power Administration for the years 2007 to 2013, were analyzed on a year by year basis. All data were normalized to total installed power of wind plants. Statistical distribution functions were obtained for the following wind generation-related quantities: total generation as percentage of total installed capacity; change in total generation power in 5, 10, 15, 20, 25, 30, 45, and 60 minutes as percentage of total installed capacity; duration of intervals with total generated power, expressed as percentage of total installed capacity, lower than certain pre-specified level. The statistical distributions obtained from the data were used to devise simple, yet accurate, theoretical models. The models presented here can be utilized in analyses related to power system economics/policy, because they describe availability of wind energy resource in simple statistical terms relevant to interactions of wind generation with electricity system, and electricity markets. After a brief display of the models, the article concentrates on static properties of the observed system's electricity generation related to its capacity credit, as well as on dynamic properties related to the demand for fast regulation (i.e., secondary and fast tertiary reserve). Both properties are important for technical planning of future electricity systems, as well as rational design of policy measures.

KEYWORDS

Wind plants, Electricity production, Measured data, Dynamic properties, Statistical modeling, Demand for regulation.

INTRODUCTION

One of most notable technical problems with massive integration of renewable sources of electricity into the existing power grid comes from the fact that most favourable technologies, i.e. wind and solar, exhibit substantial production variability on

* This paper is fully supported by Croatian Science Foundation under the project Metrological infrastructure for smart grid IP-2014-09-8826

[†] Corresponding author

all time horizons. The power system operation is mostly concerned with not-fully-predictable short-term variations in generation of power, which occur within several minutes. Technical challenges are always complemented by economic ones. Integration of renewables does cost, and the costs of variability regulation do rise with installed power. For these reasons, as well as a number of other ones, it is very important to learn about the nature of wind phenomena and associated electricity generation, which naturally follows the intensity of wind at a plant's location. To give some examples of the importance of acquiring a knowledge of wind generation statistics, which is not connected (merely) to cost aspects of short-term balancing, but to the security of grid operation, one can point to e.g., Morel *et al.* [1], who introduced a strategy to improve the transient stability of a power system with a high-penetration of intermittent renewable sources. The stability degradation arises mainly due to the highly variable outputs of these types of generators, during normal and abnormal operations.

Next, Ji *et al.* [2] analyze an aspect of influence of wind power variability on unit commitment problem and solving it using Gravitational Search Algorithm. Rotich *et al.* [3] discuss wind resource assessment and forecast planning with neural networks trained with real wind speed data. Song, Jiang and Zijun [4] discuss a Markov-switching model in wind speed forecasting, whereby utility-wide real wind speed data were used to develop the model and control its performance. Wan *et al.* [5] analyze probabilistic forecasting of wind power generation using extreme learning machine, with one-hour experimental time series of generation data from a real wind plant location in Australia to test the model. Zhang, Wang and Wang [6] give a review of methods of probabilistic forecasting of wind power generation, whereby accurate experimental data are needed both for training of certain types of algorithms, and for testing their performance. Jung and Broadwater [7] provide a similar type of review. Cannon *et al.* [8] deal with meteorological reanalysis data sets to address infrequent phenomena of extreme situations with wind production variability. In this paper, our analysis does not pertain to extreme realizations of stochastic processes of wind power generation in particular, but does point to the significance of distribution tails.

On the side of economic and social optimization of power system development with increased variability of renewables, there is also a body of literature that points to the importance of temporal analysis of wind generation. Wen *et al.* [9] investigate optimal economic allocation of energy storage systems considering wind power distribution. Zakariazadeh, Jadid and Siano [10] investigate optimal economic-environmental energy and reserve scheduling of smart distribution systems using a mathematical programming approach, where the authors model wind speed with Rayleigh distribution. Mostafaeipour *et al.* [11] provide an analysis of wind energy potential and economic evaluation in Zahedan, Iran, whereby they use Weibull distribution to model the wind variability, and consequently estimate Weibull distribution parameters from measured wind speed data. Sun *et al.* [12] propose a random fuzzy model to express probabilistic, and possibilistic, uncertainties of wind speed on adequacy of a power generation system, where wind is assumed to be Weibull-distributed. In a review by Chauhan and Saini [13], on integrated renewable energy system based power generation for stand-alone applications, the wind part of the system is modeled using Weibull distribution. In the analysis of integrated scheduling of renewable generation and demand response programs in a microgrid, Mazidi *et al.* [14] make use of Rayleigh distribution to model the wind part of the generation system.

Adaramola, Agelin-Chaab and Paul [15] use measured wind speed data along the coast of Ghana to derive parameters of theoretical Weibull distribution of speeds. Wang *et al.* [16] review and analyze estimation methods for offshore extreme wind speed statistical parameters and wind energy resources. Khahro *et al.* [17] use measured data of wind speeds from a site in Pakistan to derive Rayleigh, or Weibull, distribution

parameters for the location. Pishgar-Komleh, Keyhani and Sefeedpari [18] provide a similar type of analysis, for a geographic region in Iran. Shu, Li and Chen [19] give analysis of measured wind speed data at five weather stations in Hong Kong in one-minute temporal resolution, and estimations of Weibull distributions associated with the observed phenomena. Ouarda *et al.* [20] provide comparative analysis of goodness-of-fit of different theoretical statistical distributions to actual measured wind speed data, to conclude that the two-parameter Weibull model performs the best. Fazelpour *et al.* [21] also provide the hourly, diurnal, seasonal, monthly, and annual wind speed variations, and employ a Weibull statistical model of wind speed.

Research problem

When studying temporal characteristics of wind generation, one can take two possible generic courses: to examine wind speed features and convert them into power production ones by technical analysis, or to observe directly power production as a function of time. Both approaches are legitimate and have their justifications; however, the former is more frequently encountered in the literature. As the directly measured power generation data have been less analyzed up to now, it is useful to observe and describe generation from a fairly large wind plant system.

The primary research goal was to obtain simple theoretic distribution functions to model actual statistical distributions of wind generation-related phenomena with sufficient accuracy, so that they can be used for various analyses in the fields of either power system operation, or perhaps power system economics/policy. The quantities of interest in this research were: total power generated within a (relatively) large system, and short-term temporal changes in generated power. They are important because they require flexible generation capacities, usually provided by conventional generating units (e.g. hydro plants with water reservoirs, or gas-fired plants), to be kept aside energy market in order to be able provide physical balance of the intermittency introduced by certain renewable sources, like wind, or solar.

Another research goal relates to the fact that, often, publicly available electricity generation time-series contain data recorded in one hour intervals, which is too rough for analyses of demand for ancillary services, especially secondary and fast tertiary regulation, caused by intra-hour generation variability. Thus, the question whether one can estimate the statistical parameters of, say, 15 minute power variations, based on available 60 minute data, was found worth investigating.

For discussion of various methods for capacity credit calculation, see [22]. The present paper, deals with a parameter simply called “intrinsic capacity”, defined as a percentage of the total installed power of a group of generation facilities which can be counted on with certain default risk.

Next, the following statistical measures to describe the dynamic behavior of wind generation systems on the longer-term scale (i.e., in time periods from several hours up) shall be defined:

- The level-crossing rate at certain pre-specified level: the number of times the generation power drops below a certain pre-defined level in a year;
- Distribution of duration of intervals with total generated power, expressed as percentage of total installed capacity, lower than a certain pre-specified level: The level-crossing rate itself does not sufficiently describe longer-term dynamics of intermittent generation. However, at the same pre-specified level, one can determine the number of intervals during which the total power output is below that level, as well as the duration of each of these intervals. Therefore, one can compute their statistical distribution, as well as other descriptive statistical parameters, like expectation, variance, etc.

Structure of data

All the data analyzed in this research were taken from the web page of the Bonneville Power Administration (BPA), a federal nonprofit agency, the power system operator, based in the Pacific Northwest region of the United States of America. The BPA makes them freely available to general public. For more information on BPA, visit <http://www.bpa.gov>. The Excel files with historic data of wind generation in 5-minute increments from 2007 on are available at the address: <http://transmission.bpa.gov/business/operations/wind/>. At the same address, a number of other data sets are also available but for the purpose of this research, only the wind generation data were used.

The system of wind generation plants under jurisdiction of the BPA from was growing ever since the first plant was put in operation during 1998. As from the beginning of 2007, the system grew to 4,515 MW.

For reader's convenience, Table 1 gives a brief overview of these data. In this research, the detailed schedule was used to normalize the wind generation power to the total installed capacity.

Table 1. A brief overview of the increase in total installed wind plant power under BPA's authority from 2007 to 2013

Year	Total installed capacity at the end of the year [MW]
2006	722
2007	1,301
2008	1,599
2009	2,617
2010	3,372
2011	3,788
2012	4,711
2013	4,515

Most of the contributing wind plants are situated in a relatively small mountain area of, roughly, 50 × 60 kilometers, except for several plants situated at somewhat greater distance. An analysis to see whether the parameters of statistical distributions obtained were somehow related to changes in spatial distribution of the plants over time was performed. No apparent correlations were found.

METHODS

The basis for statistical analyses presented in this paper was a time series of wind generation power, measured in 5 minute increments during the years 2007 to 2013. These readings were normalized to total installed power of wind plants in the same area, as described in previous sections. Consequently, all the data prepared in this way belonged to an interval [0, 100%]. Statistical analyses were performed on a year-by-year basis. Therefore, in any given year the total number of experimentally obtained data points was equal to $N = (12 \text{ intervals per hour} \times 24 \text{ hours} \times 365 \text{ days})/\text{year} = 105,120$ readings per year (or 105,408 for a leap year).

Statistical distributions of experimental data were estimated in the following way: The domain containing the values between experimentally established minimum and maximum of the population was divided to 100 equally spaced bins. The membership of

each bin was counted to establish the frequency distribution, as well as the cumulative frequency distribution, of the population. By dividing the two by N , the distribution density function, as well as the probability distribution function, were obtained.

The formulae presented in Table 2 were derived in the following way: after a careful inspection of the experimentally obtained distribution functions, based on author's experience in data fitting, subsequent try-and-error selections of candidate theoretical models (the formulae from Table 2, see next section) that could reproduce the experimental distributions sufficiently well were performed. The models found in this way were then subjected to least-squares optimization computer routines.

Once the parameters of optimal distribution were estimated, goodness-of-fit assessments between experimental and theoretical distribution functions were carried out. Given a large number of individual readings, $N > 100k$, and a relatively small number of degrees of freedom (typically, 97) as compared to N , the Chi-square goodness-of-fit test could not be passed with a sufficient statistical significance in any of the cases. As the Chi-square test [23] depends heavily on the size of population, experimentally obtained distribution density functions would have to be virtually ideally slick to pass that kind of test. Yet, this was not the case. Instead, an approach based on distance between empirical and theoretical probability distribution functions, that is, the well known Kolmogorov-Smirnov test, was taken [24].

Unlike other distribution functions, which were evaluated in 100 points each, the distribution functions for duration of intervals with total generated power lower than a certain pre-specified level were evaluated in 30 points (i.e., the interval durations were distributed in 30 bins) because the total number of data points was merely between 300 and 400, and it was different for each year and each pre-specified level.

RESULTS

For the sake of brevity, statistical distributions derived from raw data in this research, as well as other quantitative relations, are stated in Table 2. Methods of testing goodness-of-fit of the distributions will not be presented here, although some results will be displayed. The details of methods used can be found in [25], or obtained from the corresponding author upon request. The raw data used for analyses can be obtained from the authors, too.

Critical values for these tests depend, however, on the type of probability distribution. Therefore, the simplest possible approach was taken: the empirical distribution was regarded as fitting well enough to a theoretical one if the absolute value of Kolmogorov-Smirnov test statistic was 0.02 or less [26]. The critical values for Kolmogorov-Smirnov test can also be found in [26]. For completely specified continuous distributions and statistical significance level of $1 - \alpha$, with number of bins, b , larger or equal to 35, these values can be estimated by applying the formula: $[-0.5 \times \ln(\alpha/2)/b]^{0.5}$. They were used as a proxy to assess if it was reasonable to regard our 0.02 criterion adequate for evaluation of goodness-of-fit. For $1 - \alpha = 0.01$ and $b = 100$, the critical value is 0.163. For $1 - \alpha = 0.01$ and $b = 35$ the critical value is 0.269. Obviously, the 0.02 "critical value" is much more stringent than these, and despite the fact that it was not possible to perform a fully rigorous Kolmogorov-Smirnov analysis, it is safe to assume no make mistake was made by inferring that an empirical curve which stays within ± 0.02 limit off of a theoretical one was well enough modeled by the latter.

The results are presented in Figures 1 to 20, and Tables 3 to 8, and brief explanations for them are listed systematically in Table 2, which is to be read together with methodical explanations of this and the previous section.

Table 2. Statistical distributions, and other quantitative relations, derived in this research

No.	Variable (X , if not speci-fied otherwise)	Formulae	Fig./ Table	Notes
1.	Total generated power	$z(x) = e^{-B} x^A$ $\ln(z) = A \ln x + B$ $y(x) = A e^{-B} x^{A-1}$	Figures 1-6 Table 3	Figure 3 gives the Kolmogorov-Smirnov test statistic values for the whole domain of values (0 to 100% of total installed wind plant power). The ± 0.02 borders are also shown. Note that the models are generally better in the lower, and the middle, part of the distribution function domain, while they do not reproduce the upper tail very well. For practical purposes, the distribution function for values of generation near the 100% of total installed capacity may not be so important because the system reliability is usually not jeopardized when there is an abundance of energy in it. Thus, we can concentrate on the lower tail of the distribution. We found the same type of model is appropriate. The corresponding parameters are listed for the years 2007 to 2013 in Table 3 as well. Figure 4 gives the 2012 distribution (both measured and theoretic) as an example. The goodness-of-fit analysis indicated a very good fit (Figure 5).
2.	r is cumulated time normalized to 1 year, thus, $r \in [0, 1]$	$VaR(r) = e^{-B/A} r^{1/A}$	Figure 7	Variable r is defined on a domain $[0, 1]$, and represents the total cumulated time, normalized to one year, during which the total generated power stays below certain level, p , also normalized to the total installed wind plant power. In other words, $p(r)$ is a Value-at-Risk (VaR) function, where p is the value, and r is the risk. Figure 7 gives as an example of the 2012 $VaR(r)$. For small values of p , a percentage of the total installed power of a group of generation facilities which can be counted on up to a certain default risk, r , increases very slowly with r . That is, for small p , VaR increases slowly with risk. Note that this property differs from many familiar VaR functions, which is not favourable as regards capacity credit.
3.	Module of change in ge-neration po-wer, normali-zed to the total installed wind plant power	$z(x) = \ln[1 + (e - 1)(1 - e^{-B/A} x^A)]$ $y(x) = \frac{(e - 1)(A/B)}{(1 - e^{-B/A} x^A)^2} [1 + (e - 1)(1 - e^{-B/A} x^A)^2]$	Figures 8-12 Table 4	Short-term temporal changes in generation power are a very important feature of any generation sub-system with significant intermittency because they set demand for secondary and fast tertiary regulation. Variable X was defined as module of change in generation power, normalized to the total installed wind plant power. The range of X values was an interval $[0, 100\%]$. Statistical distributions were derived for 5, 10, 15, 20, 25, 30, 45, and 60 minute changes. Additional relations between A and B parameters (see row 3 of this Table). The variable T_k has a very simple meaning: it is merely a time interval equal to k minutes. For instance, $T_{30} = 30$ min. Figure 13 displays natural logarithm of the ratio A/B as a function of natural logarithm of the ratio between relevant time interval, T_k , and the 5-minute interval, T_5 , for the years 2007 to 2013. Figure 14 shows the experimentally obtained relation between $\ln(A/A_{60})$ and $\ln(T_k/T_{60})$ for each year 2007-2013. The second-order polynomial regression was used to fit the data. All of the regression functions were forced to run through $(0, 0)$, which resulted in only two parameters for each year, P and Q . Figure 15 shows the experimentally obtained relation between $\ln(B/B_{60})$ and $\ln(T_k/T_{60})$ over the same period. Obviously, the linear regression functions (not shown for the sake of clarity), specified for each year, can fit the data very accurately. Just like in the A -parameter case, these regressions must be forced to run through $(0, 0)$, so that they all have merely one parameter per year, R . Table 5 lists the P , Q , and R parameters, as well as A_{15}/A_{60} and B_{15}/B_{60} factors, for the years 2007 to 2013. It also gives the averages and standard deviations of these quantities. P and Q can vary considerably over the years. However, applied mathematical operations largely dump these variations, so that the resulting ratio A_{15}/A_{60} shows a mild standard deviation of 6.45%, as compared to the average value. The R parameter does not vary much with time, so that the resulting ratio B_{15}/B_{60} also shows a mild standard deviation of 5.31% relative to the average value. To conclude, if only the 60 minute incremental data are known, one still can estimate the 15 minute distribution parameters, A_{15} and B_{15} , by taking about 85% of A_{60} , and about 36% of B_{60} , respectively. More generally, parameters of statistical distributions of 15 minute (or any other k minute) changes can be estimated relatively precisely from 60 minute parameters. See more details in [25].
4.	T_k – interval of k minutes, $k \in \{5, 10, 15, 20, 25, 30, 45, 60\}$	$B_k = A_k e^D (T_k/T_5)^C$ $A_k = A_{60} (T_k/T_{60})^Q \exp[-P \ln^2(T_k/T_{60})]$ $A_{15} = A_{60} 0.25^Q \exp[-1.9218 P]$ $B_k = B_{60} (T_k/T_{60})^R$ $B_{15} = B_{60} 0.25^R$	Figures 13-15 Table 5	
5.	M_{reg} , "regulation multiplier"	$M_{reg} = (U/V)r^V + W r$ (U, V and W are real and positive constants)	Figures 16 and 17	Conceptually, M_{reg} shows how many times can total installed power of wind plants exceed available 15 minute flexible regulation reserve, if the system operator deems default risk of r percent acceptable, nothing else being variable in the power system. Because of inherent variability of load, other generating plants, and cross border exchange, and because regulation reserve may not always be available in its maximum value, in the reality the multiplier would be somewhat smaller. Here we give only formula for small values of r . More details in [25].
6.	Duration of intervals with total generated power, expressed as percentage of total installed capacity, lower than certain pre-specified level	$z(x) = (1 - e^{-Bx^A})^A$ $y(x) = AB e^{-Bx^A} (1 - e^{-Bx^A})^{A-1}$	Figures 18-20 Tables 6-8	The concepts were borrowed from radio engineering, where electromagnetic field strength at the receiving antenna, when a signal is transmitted over time-varying channels, exhibits changes in time, from rather slow ones, to very fast, and anything in between [27]. Figure 18 explains the definition of level-crossing rate: if the level L is crossed downwards R_L times per year, the level-crossing rate equals R_L per year. This quantity is an indicator of the wind generation volatility. It is a one single number assigned to a whole year. Table 6 gives the R_L figures for the years 2007-2013. Figure 19 gives ten cumulative distribution functions for ten pre-specified levels L , for the year 2013, while Figure 20, for the sake of clarity, shows only one of these curves (that for the level of 5%) together with the corresponding theoretic curve $z(x)$. Table 7 contains the A parameters calculated for the years 2007-2013, and for each of the ten levels. Table 8 gives the B parameters of the same distributions.

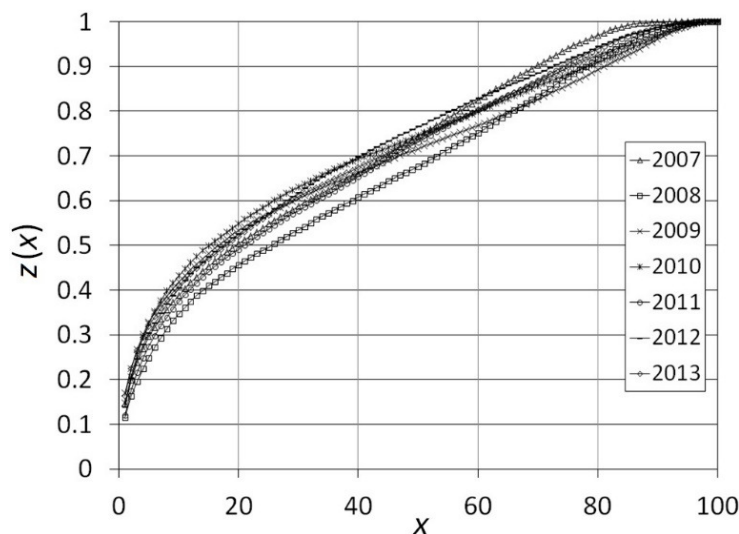


Figure 1. Experimental probability distribution functions, $z(x)$, of total generated power, expressed as percentage of total installed wind capacity, x , for the years 2007 to 2013

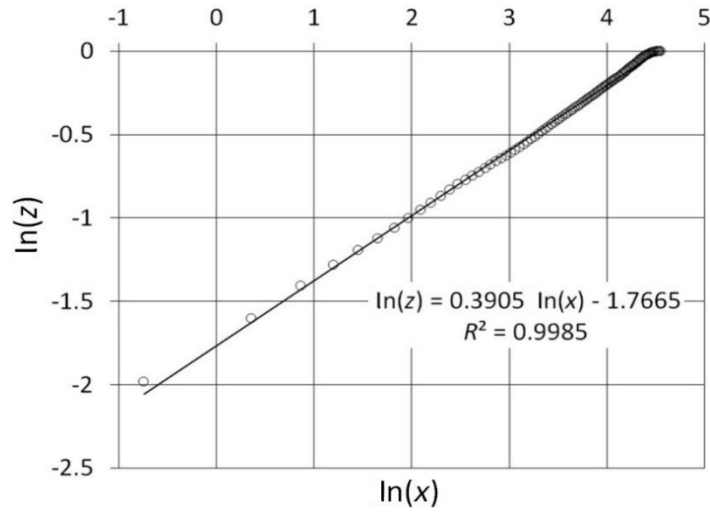


Figure 2. Log-log scatter plot of 2012 probability distribution from Figure 1 with associated linear trend line

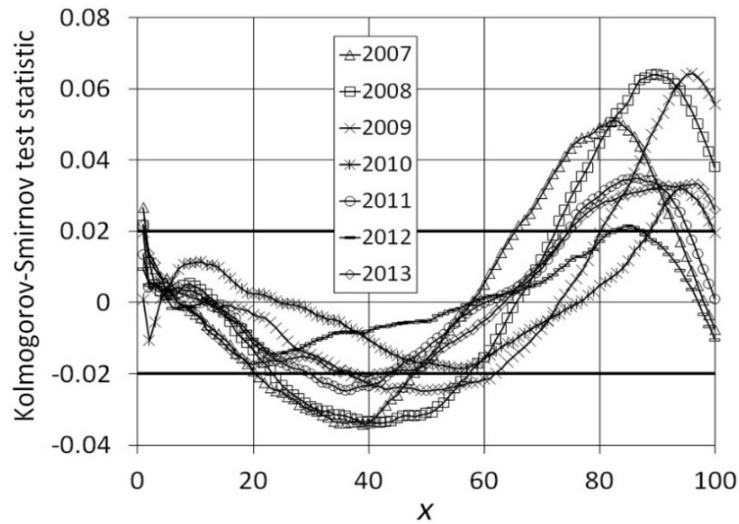


Figure 3. Kolmogorov-Smirnov test statistic values for the whole domain of values (0 to 100% of total installed wind plant power), the ± 0.02 borders are emphasized with thick gridlines

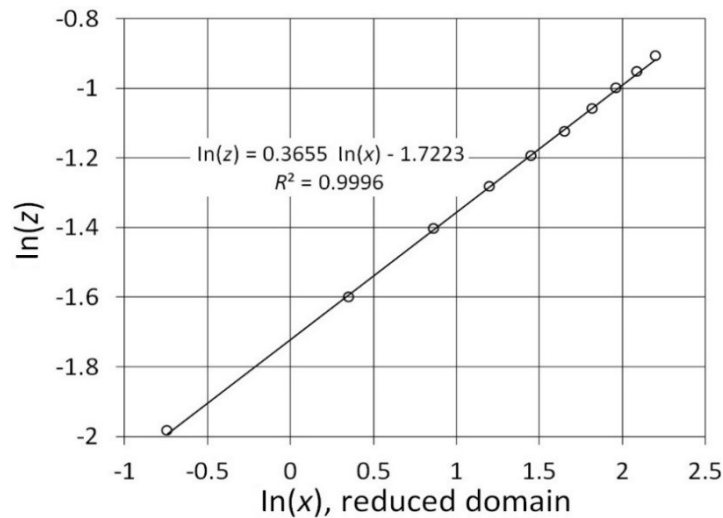


Figure 4. Log-log scatter plot of 2012 probability distribution of total generated power, expressed as percentage of total installed wind plant capacity, for $x \leq 10\%$

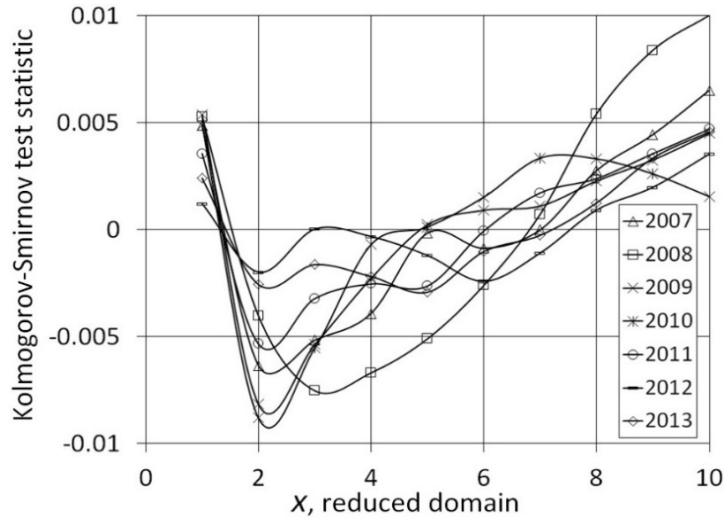


Figure 5. Kolmogorov-Smirnov test statistic values for the reduced domain (0 to 10% of total installed wind plant power)

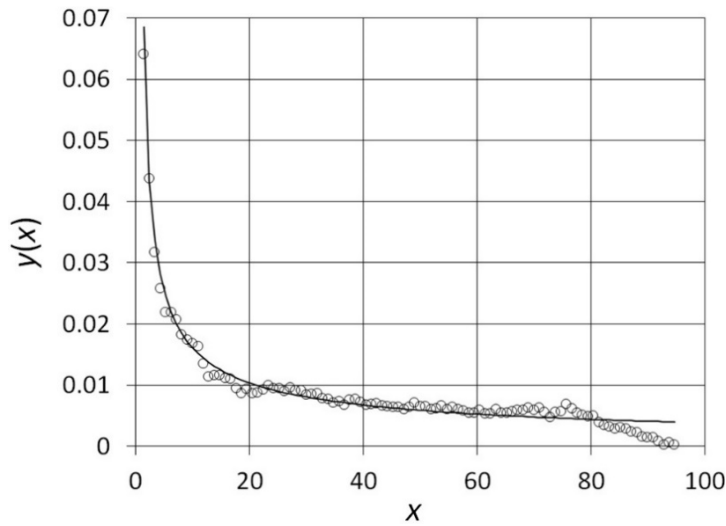


Figure 6. An example of probability density function that corresponds to the model from Figure 1, for 2012

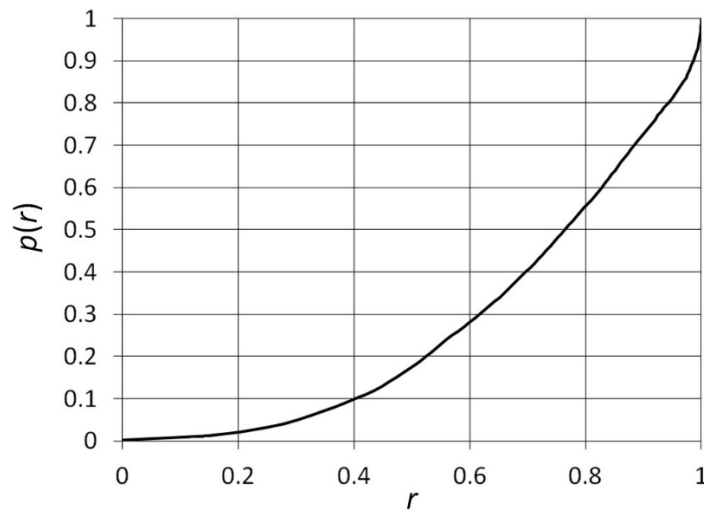


Figure 7. Value-at-Risk (VaR) function for the level of wind generated power for 2012, normalized to the total installed wind plant capacity (p), the default risk level is denoted by r

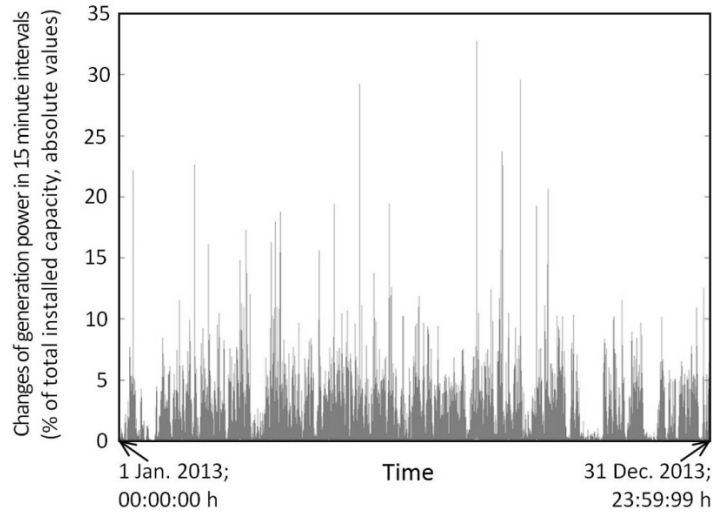


Figure 8. Absolute values of 15 minute changes of wind generation power, expressed as percentage of total installed wind plant capacity, for the year 2013

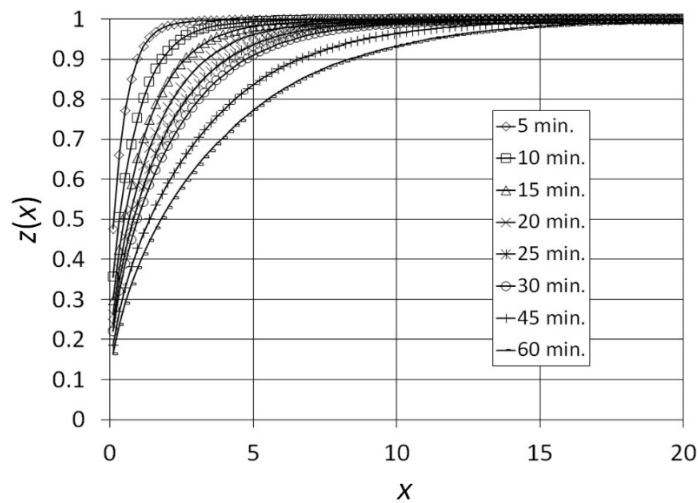


Figure 9. Experimental probability distribution functions, $z(x)$, of short-time changes of wind generation power in several time intervals, expressed as percentage of total installed wind plant capacity, x , for the year 2013

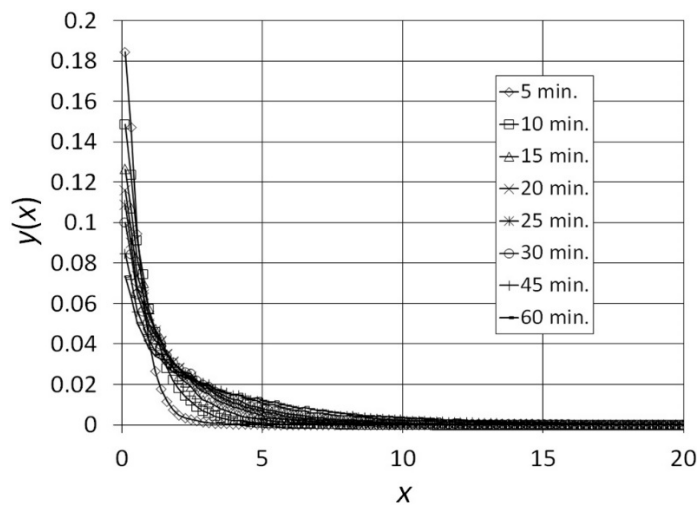


Figure 10. Experimental probability density functions, $y(x)$, that correspond to the probability distribution functions from Figure 9

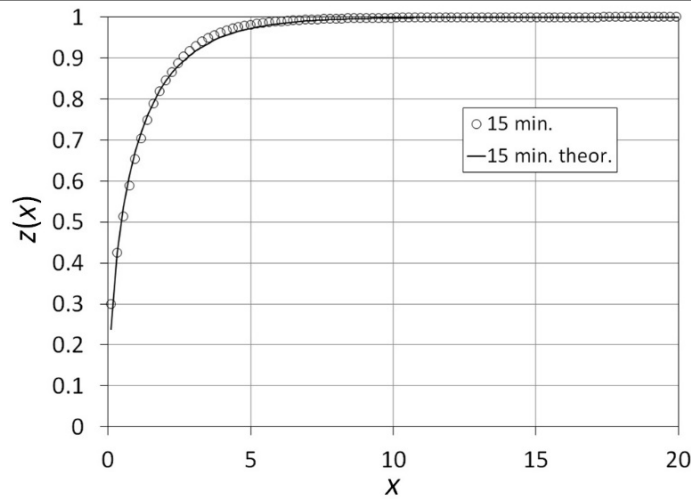


Figure 11. An example of the fit between the probability distribution function found from the 2013 experimental data (15 minute changes), and the theoretical function derived from it (see Table 2, row No. 3)

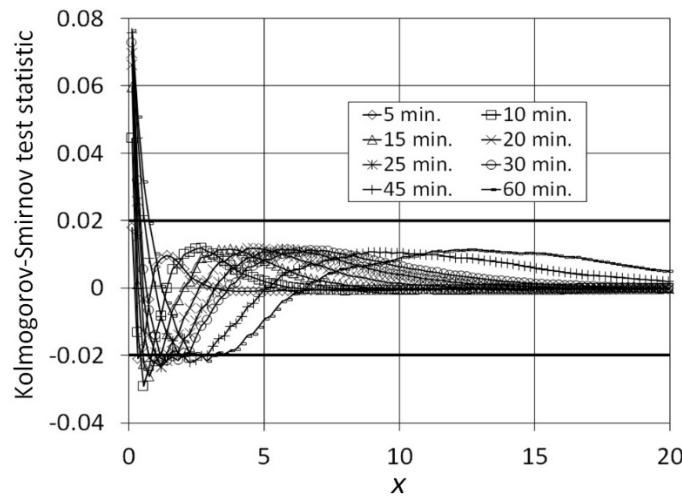


Figure 12. An example of the Kolmogorov-Smirnov test statistics together with ± 0.02 boundaries for generation power changes in 2013

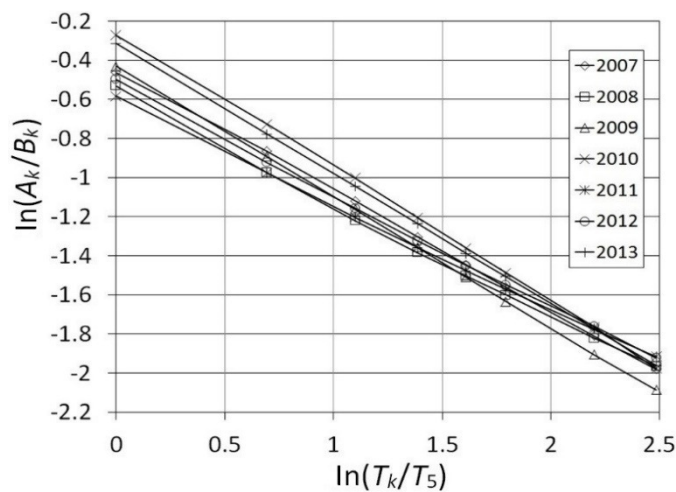


Figure 13. Natural logarithm of the ratio A/B as a function of natural logarithm of the ratio between relevant time interval, T_k , $k \in \{5, 10, 15, 20, 25, 30, 45, 60\}$, and the 5 minute interval, T_5 , for the years 2007 to 2013

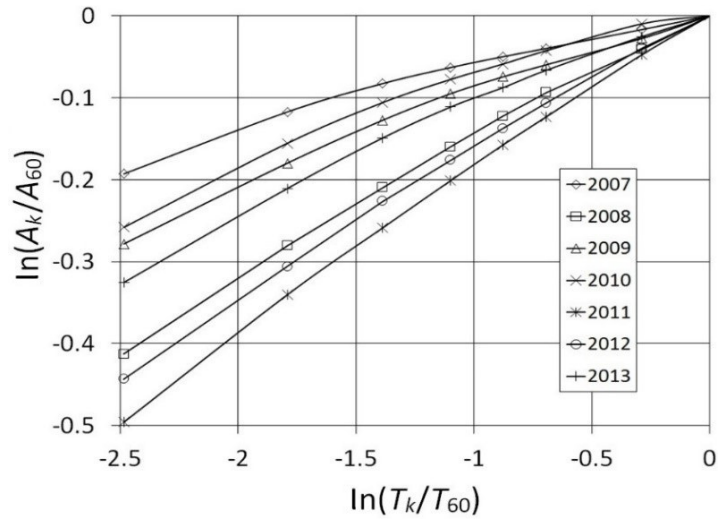


Figure 14. Experimentally obtained relation between $\ln(A_k/A_{60})$ and $\ln(T_k/T_{60})$ over the 2007 to 2013 period

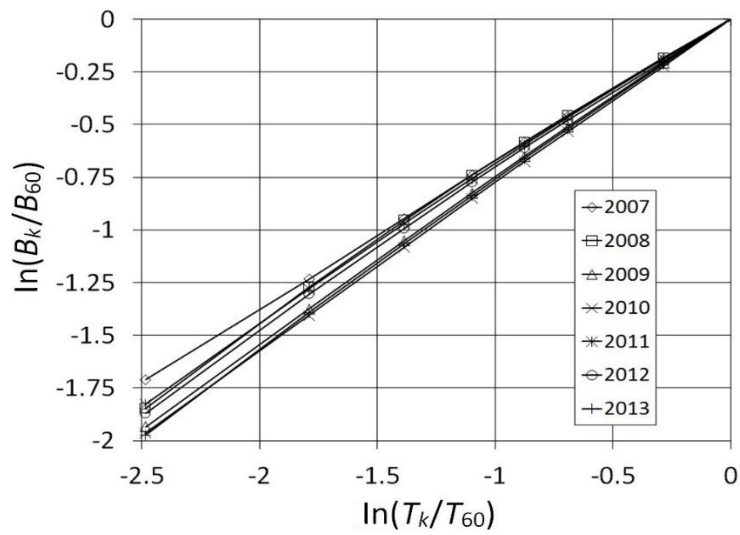


Figure 15. Experimentally obtained relation between $\ln(B_k/B_{60})$ and $\ln(T_k/T_{60})$ over the 2007 to 2013 period

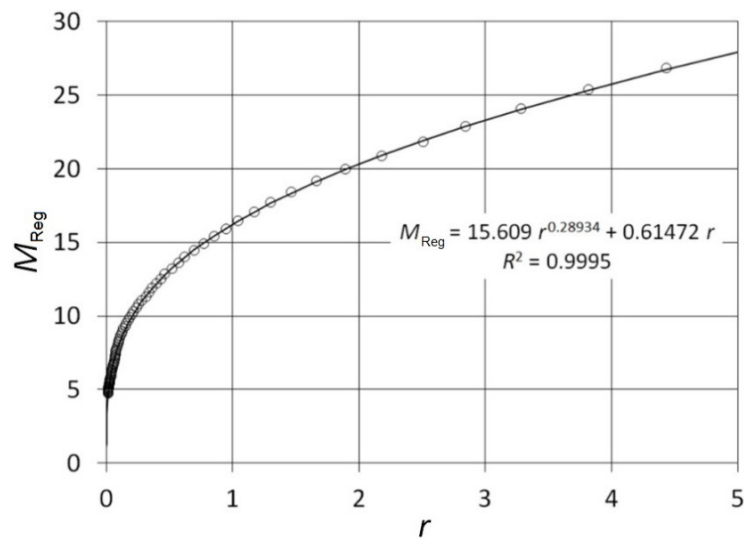


Figure 16. The fit between experimental and corresponding theoretical M_{Reg} function for 2013

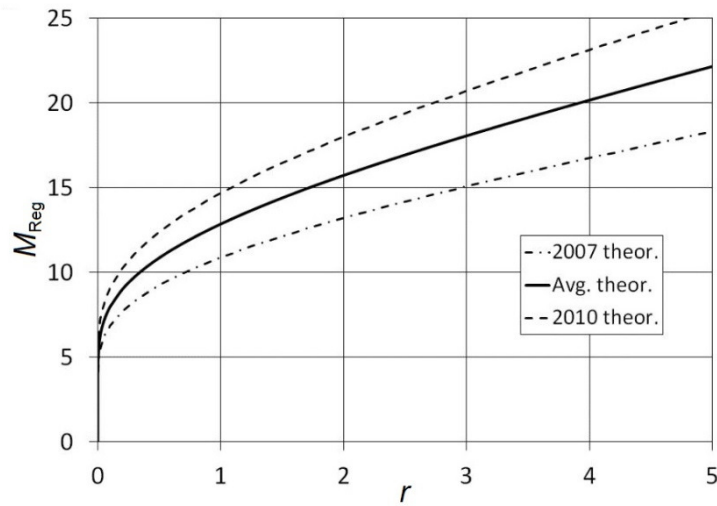


Figure 17. Theoretic M_{Reg} functions for extreme years (2007 and 2010) and the geometric mean between them

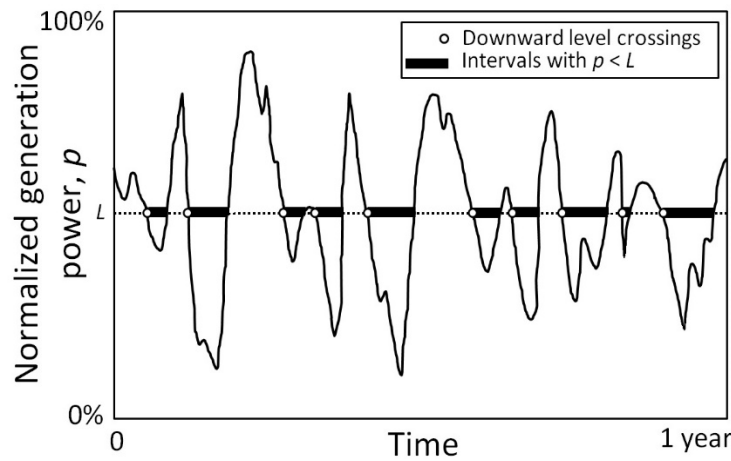


Figure 18. Explanation of level-crossing rate and intervals with normalized generation power lower than certain level

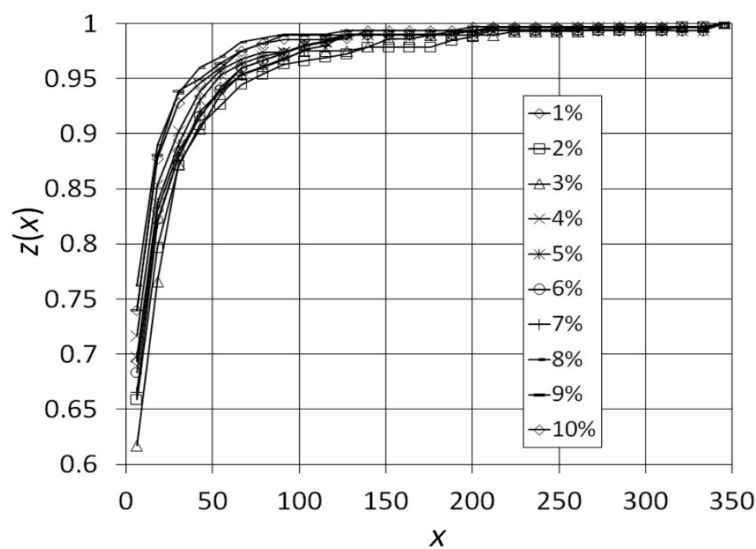


Figure 19. Probability distribution functions, $z(x)$, of duration of intervals with total generated power lower than a pre-defined level, for 2013, expressed as percentage of total installed plant capacity, x

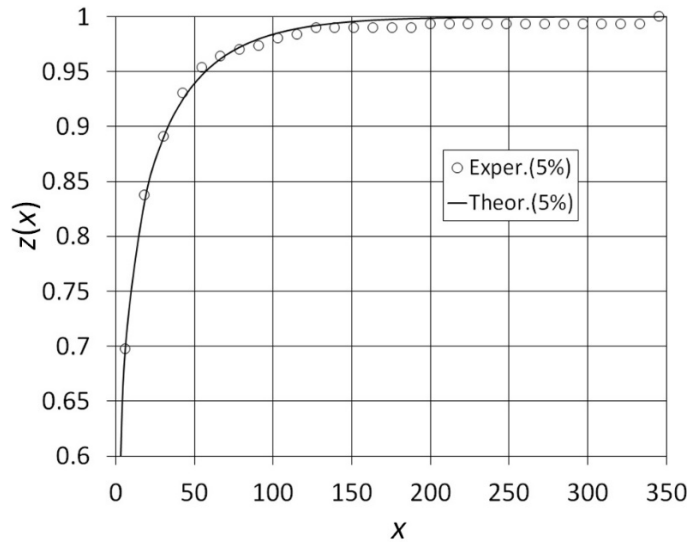


Figure 20. The 5% level curve from Figure 19 together with the corresponding theoretic curve

Table 3. *A* and *B* parameters of distributions from Table 2, row No. 1, for both complete and reduced domain of *x*

Complete domain: $x \in [0, 100\%]$							
Parameter	2007	2008	2009	2010	2011	2012	2013
<i>A</i>	0.402838	0.439875	0.347887	0.35989	0.422034	0.390461	0.365416
<i>B</i>	-1.87783	-2.06074	-1.64213	-1.6616	-1.93211	-1.76643	-1.69203
Reduced domain: $x \in [0, 10\%]$							
Parameter	2007	2008	2009	2010	2011	2012	2013
<i>A</i>	0.337334	0.379821	0.314231	0.376816	0.390471	0.365449	0.314955
<i>B</i>	-1.75044	-1.9419	-1.5758	-1.6763	-1.86602	-1.72231	-1.59877

Table 4. *A* and *B* parameters of theoretic probability distribution functions (see Table 2, row No. 3) for changes of generation power in intervals of *k* minutes, $k \in \{5, 10, 15, 20, 25, 30, 45, 60\}$, and for the years 2007-2013, the changes of generation power are expressed as percentages of total installed wind capacity

Parameter <i>A</i>							
<i>k</i> [min]	2007	2008	2009	2010	2011	2012	2013
5	0.6899	0.5513	0.5567	0.5889	0.4967	0.5018	0.5401
10	0.7440	0.6294	0.6143	0.6525	0.5802	0.5756	0.6057
15	0.7703	0.6755	0.6474	0.6858	0.6297	0.6233	0.6443
20	0.7855	0.7099	0.6687	0.7054	0.6670	0.6553	0.6691
25	0.7960	0.7371	0.6826	0.7189	0.6968	0.6813	0.6852
30	0.8041	0.7587	0.6927	0.7306	0.7210	0.7026	0.6998
45	0.8228	0.7996	0.7153	0.7549	0.7777	0.7517	0.7290
60	0.8366	0.8331	0.7353	0.7623	0.8155	0.7816	0.7478
Parameter <i>B</i>							
<i>k</i> [min]	2007	2008	2009	2010	2011	2012	2013
5	1.0970	0.9390	0.8573	0.7738	0.8918	0.8247	0.7397
10	1.7679	1.6681	1.5002	1.3513	1.5398	1.4505	1.3206
15	2.3606	2.2911	2.0711	1.8687	2.0992	1.9831	1.8325
20	2.8945	2.8349	2.5954	2.3529	2.5971	2.4651	2.3029
25	3.3816	3.3178	3.0887	2.8077	3.0527	2.9040	2.7479
30	3.8311	3.7620	3.5529	3.2337	3.4758	3.3082	3.1597
45	5.0266	4.9486	4.8127	4.4073	4.5916	4.3739	4.2911
60	6.0686	5.9435	5.9262	5.5109	5.5367	5.3480	5.3179

Table 5. The P , Q , and R parameters (see Table 2, row No. 4) extracted from the experimental data, for the years 2007 to 2013, together with factors that transpose 60 minute distribution parameters A and B to 15 minute ones, the factor A_{15}/A_{60} equals $0.25^Q e^{-1.9218P}$, while the factor B_{15}/B_{60} equals 0.25^R , the table displays average values, as well as standard deviations normalized to average, of the aforementioned quantities, too

Year	2007	2008	2009	2010	2011	2012	2013	Avg.	St.dev./avg. [%]
P	0.0137	0.0152	0.0165	0.0244	0.0124	0.0136	0.0203	0.0166	25.99
Q	0.0427	0.1288	0.0706	0.0432	0.1678	0.1449	0.0804	0.0969	51.82
A_{15}/A_{60}	0.9180	0.8124	0.8785	0.8987	0.7738	0.7969	0.8603	0.8484	6.45
R	0.6839	0.7135	0.7668	0.7841	0.7154	0.7321	0.7798	0.7394	5.17
B_{15}/B_{60}	0.3875	0.3719	0.3454	0.3372	0.3709	0.3624	0.3392	0.3592	5.31

Table 6. The level-crossing rates, R_L , for the years 2007-2013, and for the levels of 10, 20, 30, 40, 50, 60, 70, 80, and 90 percent of the total installed wind plant capacity

Level [%]	2007	2008	2009	2010	2011	2012	2013	Avg.	Std.	Avg./std. [%]
10	445	390	333	338	329	355	276	352.29	53.17	15.09
20	369	334	321	323	346	304	288	326.43	26.68	8.18
30	372	353	269	295	341	350	237	316.71	50.29	15.89
40	392	363	282	255	359	342	267	322.86	53.94	16.71
50	452	357	275	254	391	315	272	330.86	72.71	21.98
60	434	391	285	271	336	286	277	325.71	64.16	19.70
70	405	404	294	229	303	248	253	305.14	72.63	23.80
80	355	337	242	199	209	165	173	240.00	76.80	32.00
90	153	130	146	107	51	27	97	101.57	47.60	46.87
Avg.	375.22	339.89	271.89	252.33	296.11	265.78	237.78			
Std.	90.30	82.41	54.39	69.84	104.79	107.61	63.04			
Avg./std. [%]	24.06	24.25	20.01	27.68	35.39	40.49	26.51			

Table 7. Parameters A of probability distributions (Table 2, row No. 6) of duration of intervals with total generated power, expressed as percentage of total installed capacity, lower than a pre-defined level, for the years 2007-2013, and for each of the ten pre-specified levels

Level [%]	2007	2008	2009	2010	2011	2012	2013	Avg.	Std.	Avg./std. [%]
1	0.15024	0.23608	0.07311	0.24253	0.27735	0.25489	0.17896	0.20188	0.07200	35.6636
2	0.14262	0.26497	0.10398	0.22668	0.26401	0.27912	0.18737	0.20982	0.06743	32.1403
3	0.12890	0.24079	0.12182	0.25223	0.27271	0.26103	0.24816	0.21795	0.06408	29.4003
4	0.13929	0.23842	0.16550	0.29074	0.29499	0.26345	0.17132	0.22339	0.06407	28.6804
5	0.12576	0.23511	0.18801	0.29557	0.28862	0.28004	0.18180	0.22784	0.06479	28.4359
6	0.13250	0.24455	0.21960	0.32280	0.23489	0.27490	0.18558	0.23069	0.06117	26.5142
7	0.13478	0.20121	0.25547	0.30684	0.25327	0.24183	0.20059	0.22771	0.05471	24.0260
8	0.14042	0.21822	0.23984	0.32049	0.25547	0.27041	0.16152	0.22943	0.06247	27.2229
9	0.15218	0.23959	0.23967	0.29102	0.23398	0.25126	0.17678	0.22635	0.04685	20.6995
10	0.14856	0.23207	0.23104	0.29020	0.25761	0.25492	0.16522	0.22566	0.05117	22.6747
Avg.	0.13952	0.23510	0.18380	0.28391	0.26329	0.26318	0.18573			
Std.	0.00908	0.01663	0.06475	0.03273	0.02053	0.01277	0.02467			
Avg./std. [%]	6.50434	7.07287	35.2260	11.5277	7.79679	4.85244	13.2813			

Table 8. Parameters B of probability distributions (Table 2, row No. 6) of duration of intervals with total generated power, expressed as percentage of total installed capacity, lower than a pre-defined level, for the years 2007-2013, and for each of the ten pre-specified levels

Level [%]	2007	2008	2009	2010	2011	2012	2013	Avg.	Std.	Avg/std. [%]
1	0.1115	0.1520	0.0850	0.1175	0.1385	0.1215	0.0710	0.11386	0.02831	24.8648
2	0.0825	0.1130	0.0750	0.0845	0.0910	0.0835	0.0590	0.08407	0.01632	19.4161
3	0.0605	0.0845	0.0690	0.0775	0.0830	0.0635	0.0565	0.07064	0.01117	15.8092
4	0.0545	0.0800	0.0685	0.0695	0.0700	0.0535	0.0425	0.06264	0.01286	20.5299
5	0.0530	0.0695	0.0580	0.0590	0.0640	0.0520	0.0405	0.05657	0.00933	16.4912
6	0.0430	0.0625	0.0570	0.0585	0.0565	0.0475	0.0370	0.05171	0.00934	18.0610
7	0.0410	0.0550	0.0545	0.0520	0.0520	0.0440	0.0385	0.04814	0.00681	14.1480
8	0.0430	0.0535	0.0515	0.0505	0.0470	0.0435	0.0345	0.04621	0.00652	14.1045
9	0.0420	0.0525	0.0485	0.0440	0.0485	0.0400	0.0340	0.04421	0.00622	14.0769
10	0.0400	0.0460	0.0425	0.0410	0.0475	0.0375	0.0300	0.04064	0.00582	14.3066
Avg.	0.05710	0.07685	0.06095	0.06540	0.0698	0.05865	0.04435			
Std.	0.02314	0.03308	0.01312	0.02306	0.02856	0.02588	0.01329			
Avg./std. [%]	40.5160	43.0481	21.5190	35.2600	40.9184	44.1172	29.9606			

Nomenclature of the presented results

In subsequent sections of this article, to keep the number of different mathematical symbols reasonably low for reader’s convenience, the following simple notation rules will be used for all the quantities dealt with:

- The random variable dealt with will be marked with X ;
- The probability density function of that variable will be marked with $y(x)$, and the cumulative probability distribution function will be marked with $z(x)$. It always applies: $y(x) = dz(x)/dx$;
- The distribution parameters will be marked with capital Latin letters, e.g. A, B, C , etc.;
- For all other mathematical symbols, lower case Latin or Greek lettering will be used;
- Thus, all the different quantities will be presented using the same basic x - y notation.

DISCUSSION

In this work historical data on wind electricity generation in the BPA region were analyzed, in order to devise simple, yet accurate, statistical models for the following quantities:

- Distributions of total generated power expressed as percentage of total installed capacity;
- Distributions of short-term temporal changes in total generated power expressed as percentage of total installed capacity;
- Level-crossing rates and distributions of duration of intervals with total generated power, expressed as percentage of total installed capacity, lower than a certain pre-specified level.

Additionally, VaR functions for what can be called “intrinsic capacity” of the wind plant system (the generation power, expressed as percentage of total installed wind plant capacity, that can be counted on with certain default risk) were derived, as well as for “regulation multiplier”, a quantity that shows how many times, given certain acceptable risk of fast regulation reserve getting short, it can be exceeded by total installed wind plant capacity, provided nothing else in the system is variable.

In each of the above mentioned cases we obtained very simple statistical models and proved in a rigorous manner that they model the associated quantities correctly.

An approach that was used in this research was to treat a relatively large wind plant (sub)system as a “black box” and to analyze statistical properties of large, very finely quantized, time series of measured generating power. The main research goal was to devise simple closed-form statistical models of several stochastic variables of interest, in order to get as simple descriptions of “the nature” as possible, in terms of quantities needed for economic and/or technical analyses.

The “regulation multiplier” concept (see Table 2, row No. 5) was devised in the first place to explore some very basic properties of “the nature” regarding necessary fast regulation reserves that the system operators must keep aside the energy markets in order to balance inherently variable generation. As already mentioned, the M_{Reg} quantity, as defined here, is not absolutely accurate. Yet, it gives a good insight into the nature of the problem. By “fast reserve” we understood any spinning generation capacity reserve that can be fully engaged in 15 minutes, regardless of flexible generation technology employed. In real electricity systems, capacities engaged in such way are being gradually released by engaging slower flexible generation (e.g. tertiary regulation reserve), so to make the fastest flexible generators available again.

The idea of our research was that 15 minute changes in total generated power should be matched with the reserve up to a certain probability of default. This risk comes from the fact that 15 minute changes have certain statistical properties, which can be, as in the presented research, normalized to total installed power. Therefore, if the available 15 minute reserve was $M_{\text{Reg}}(r)$ times smaller than total installed power, it would be sufficient to balance the generation variability, except during r percent of total time. A practical problem with such an approach is that only the variability of wind generation was taken into account. In reality, load also varies and other generators, although being more predictable, may experience e.g. technical faults, so that they are also generally variable. As variances of mutually independent stochastic variables (such as wind generation and load, especially in short time) add, obviously, the regulation multiplier would in reality be smaller.

Moreover, the reserved 15 minute capacity may not always be fully available as tertiary regulation cannot act instantly, so that realistic regulation multiplier may even be smaller than that. Next, the 15 minute changes in generation power are merely a proxy for “regulation demand”, used here to illustrate the basic relation between acceptable default risk and regulation multiplier. The physical balancing of the system is of a more complex nature. Next, in this calculation we ignored ever-improving possibilities of forecasting of wind generation, which can release regulation demand to a certain degree. However, it is still questionable whether the changes in wind speed can be forecasted across wide territories in a time-coherent manner with usable precision. If not, as regards impact on demand for regulation, imprecise and incoherent forecasts of fast changes are as good as none. As this particular point was not studied here, definitive conclusions on it are omitted.

Finally, cross-border interconnections may help a little in case the domestic power system runs short of fast secondary regulation for a short period of time, as neighboring systems may provide missing regulation if they have some free at the moment. Yet, should every country pursue rapid deployment of renewable sources integration, the possibilities for such assistance with natural energy exchanges will inevitably diminish with time, not to mention system operation rules, requiring each power system operator to keep its control area balanced all the time.

One of the other notable features of the presented research is the connection between distribution parameters of changes in generation power over different (short) time periods. This is an interesting point because measurements with fine time resolution are

generally not easily available on public web sites. However, data with one-hour resolution can be found much more easily. Therefore, a possibility to credibly estimate, say, 15 minute changes from the 60 minute data can lead to considerable growth of experimental data usable for analyses of demand for regulation generated by intermittent wind plants.

In addition, in this research note we introduced new statistics that may prove important for analyses concerned with time variability of wind power production on all time horizons: level-crossing rate, and distribution of intervals with momentary generation power lower than a certain level. Naturally, in order to make use of the statistical distributions and other quantities discussed in this article, it would be good to have more systems from different parts of the world analyzed, to compare the results.

CONCLUSIONS

The models established in this research, which emulate processes of natural production of electricity in a wind plant system, are important for sound policy making as they enable modelers and planners to simulate economic effects that intermittent wind generation exerts on electricity markets, both in the short and long run, but also on electricity system in technical terms.

The data gathered from wind plants, only, were analyzed. One could have taken into account other parts of the power system, too, but it was decided to go with the “pure” wind generation data, to study the wind plant system as a separate component. In more distant future, once the renewables assume dominating shares in total energy generation, this kind of approach may prove quite important, as time variability of power production will become one of the biggest economic and technical challenges utilities will have to be dealing with. The problem of compensation of time variability of power sources is, at least in the authors’ minds, somewhat underestimated these days. The policy making regarding renewables is sometimes heavily influenced by both industry interests, and ideology. Therefore, it is important to construct easy-to-use and easy-to-understand modeling tools, to bring more rationality into the debate. As much as a rapid greenification of the power system is needed, one of the biggest threats to this process is to keep leading it in an economically non-sustainable way.

FUTURE RESEARCH

The analysis presented here can be completed and improved in several ways. First, it would be useful to analyze generation from other bigger wind plant systems, perhaps those with greater spatial diversity, to see whether it significantly influences the shape of statistical distributions. (Naturally, the parameters of distributions are expected to change, but one should observe what, if anything, happens to the very types of them.) Then, analyses which take account of other variable parts of the power system, as well, would complement the results displayed here. As the wind generation is not the only variable part of the system, it would be useful to determine how a statistical interaction with other variable parts leads to capacity credit at various levels of wind penetration. The concept of M_{Reg} should also be deepened, i.e. modeled with more detail, because, as concluded in the Discussion section, under more realistic conditions M_{Reg} would turn out to be somewhat smaller.

REFERENCES

1. Morel, J., Obara, S. and Morizane, Y., Stability Enhancement of a Power System Containing High-penetration Intermittent Renewable Generation, *J. Sustain. Dev. Energy Water Environ. Syst.*, Vol. 3, No. 2, pp 151-162, 2015, <https://doi.org/10.13044/j.sdewes.2015.03.0012>

2. Ji, B., Yuan, X., Chen, Z. and Tian, H., Improved Gravitational Search Algorithm for Unit Commitment Considering Uncertainty of Wind Power, *Energy*, Vol. 67, pp 52-62, 2014, <https://doi.org/10.1016/j.energy.2014.02.014>
3. Rotich, N. K., Backman, J., Linnanen, L. and Daniil, P., Wind Resource Assessment and Forecast Planning with Neural Networks, *J. Sustain. Dev. Energy Water Environ. Syst.*, Vol. 2, No. 2, pp 174-190, 2014, <https://doi.org/10.13044/j.sdewes.2014.02.0015>
4. Song, Z., Jiang, Y. and Zijun, Z., Short Term Wind Speed Forecasting with Markov-switching Model, *Applied Energy*, Vol. 130, pp 103-112, 2014, <https://doi.org/10.1016/j.apenergy.2014.05.026>
5. Wan, C., Xu, Z., Pinson, P., Dong, Z. Y. and Wong, K. P., Probabilistic Forecasting of Wind Power Generation using Extreme Learning Machine, *IEEE Transactions on Power Systems*, Vol. 29, No. 3, pp 1033-1044, 2014, <https://doi.org/10.1109/TPWRS.2013.2287871>
6. Zhang, Y., Wang, J. and Wang, X., Review on Probabilistic Forecasting of Wind Power Generation, *Renewable and Sustainable Energy Reviews*, Vol. 32, pp 255-270, 2014, <https://doi.org/10.1016/j.rser.2014.01.033>
7. Jung, J. and Broadwater, R. P., Current Status and Future Advances for Wind Speed and Power Forecasting, *Renewable and Sustainable Energy Reviews*, Vol. 31, pp 762-777, 2014, <https://doi.org/10.1016/j.rser.2013.12.054>
8. Cannon, D. J., Brayshaw, D. J., Methven, J., Coker, P. J. and Lenaghan, D., Using Reanalysis Data to Quantify Extreme Wind Power Generation Statistics: A 33 Year Case Study in Great Britain, *Renewable Energy*, Vol. 75, pp 767-778, 2015, <https://doi.org/10.1016/j.renene.2014.10.024>
9. Wen, S., Lan, H., Fu, Q., Yu, D. C. and Zhang, L., Economic Allocation of Energy Storage System Considering Wind Power Distribution, *IEEE Transactions on Power Systems*, Vol. 30, No. 2, pp 644-652, 2015, <https://doi.org/10.1109/TPWRS.2014.2337936>
10. Zakariazadeh, A., Jadid, S. and Siano, P., Economic-environmental Energy and Reserve Scheduling of Smart Distribution Systems: A Multiobjective Mathematical Programming Approach, *Energy Conversion and Management*, Vol. 78, pp 151-164, 2014, <https://doi.org/10.1016/j.enconman.2013.10.051>
11. Mostafaeipour, A., Jadidi, M., Mohammadi, K. and Sedaghat, A., An Analysis of Wind Energy Potential and Economic Evaluation in Zahedan, Iran, *Renewable and Sustainable Energy Reviews*, Vol. 30, pp 641-650, 2014, <https://doi.org/10.1016/j.rser.2013.11.016>
12. Sun, C., Bie, Z., Xie, M. and Ning, G., Effects of Wind Speed Probabilistic and Possibilistic Uncertainties on Generation System Adequacy, *IET Generation, Transmission & Distribution*, Vol. 9, No. 4, pp 339-347, 2015, <https://doi.org/10.1049/iet-gtd.2014.0708>
13. Chauhan, A. and Saini, R. P., A Review on Integrated Renewable Energy System Based Power Generation for Stand-alone Applications: Configurations, Storage Options, Sizing Methodologies and Control, *Renewable and Sustainable Energy Reviews*, Vol. 38, pp 99-120, 2014, <https://doi.org/10.1016/j.rser.2014.05.079>
14. Mazidi, M., Zakariazadeh, A., Jadid, S. and Siano, P., Integrated Scheduling of Renewable Generation and demand Response Programs in a Microgrid, *Energy Conversion and Management*, Vol. 86, pp 1118-1127, 2014, <https://doi.org/10.1016/j.enconman.2014.06.078>
15. Adaramola, M. S., Agelin-Chaab, M. and Paul, S. S., Assessment of Wind Power Generation along the Coast of Ghana, *Energy Conversion and Management*, Vol. 77, pp 61-69, 2014, <https://doi.org/10.1016/j.enconman.2013.09.005>
16. Wang, J., Qin, S., Jin, S. and Wu, J., Estimation Methods Review and Analysis of Offshore Extreme Wind Speeds and Wind Energy Resources, *Renewable and*

- Sustainable Energy Reviews*, Vol. 42, pp 26-42, 2015, <https://doi.org/10.1016/j.rser.2014.09.042>
17. Khahro, S. F., Tabbasumm, K., Soomro, A. M., Dong, L. and Liao, X., Evaluation of Wind Power Production Prospective and Weibull Parameter Estimation Methods for Babaurband, Sindh Pakistan, *Energy Conversion and Management*, Vol. 78, pp 956-967, 2014, <https://doi.org/10.1016/j.enconman.2013.06.062>
 18. Pishgar-Komleh, S. H., Keyhani, A. and Sefeedpari, P., Wind Speed and Power Density Analysis Based on Weibull and Rayleigh Distributions (A Case Study: Firouzkooh County of Iran), *Renewable and Sustainable Energy Reviews*, Vol. 42, pp 313-332, 2015, <https://doi.org/10.1016/j.rser.2014.10.028>
 19. Shu, Z. R., Li, Q. S. and Chan, P. W., Statistical Analysis of Wind Characteristics and Wind Energy Potential in Hong Kong, *Energy Conversion and Management*, Vol. 101, pp 644-657, 2015, <https://doi.org/10.1016/j.enconman.2015.05.070>
 20. Ouarda, T. B. M. J., Charron, C., Shin, J.-Y., Marpu, P. R., Al-Mandoos, A. H., Al-Tamimi, L. H., Ghedira, H. and Al Hosary, T. N., Probability Distributions of Wind Speed in UAR, *Energy Conversion and Management*, Vol. 93, pp 414-434, 2015, <https://doi.org/10.1016/j.enconman.2015.01.036>
 21. Fazelpour, F., Soltani, N., Soltani, S. and Rosen, M. A., Assessment of Wind Energy Potential and Economics in the North-Western Iranian Cities of Tabriz and Ardabil, *Renewable and Sustainable Energy Reviews*, Vol. 45, pp 87-99, 2015, <https://doi.org/10.1016/j.rser.2015.01.045>
 22. Amelin, M., Comparison of Capacity Credit Calculation Methods for Conventional Power Plants and Wind Power, *IEEE Trans. Power Syst.*, Vol. 24, No. 2, pp 685-691, 2009, <https://doi.org/10.1109/TPWRS.2009.2016493>
 23. Moore, D. S., Chi-Square Tests (Hogg, R. W., ed.), *Studies in Statistics*, pp 66-106, Mathematical Association of America, Washington DC, USA, 1978.
 24. Noether, G. E., A Brief Survey of Nonparametric Statistics (Hogg, R. V., ed.), *Studies in Statistics*, pp 39-95, Mathematical Association of America, Washington DC, USA, 1978.
 25. Sabolić, D., Statistical modeling of Large Wind Plant System's Generation – A Case Study, *Proceedings of the 23rd Forum of Croatian Energy Society*, Zagreb, Croatia, 2014.
 26. Sachs, L., *Angewandte Statistik*, pp 427-431, Springer-Verlag, Berlin, Heidelberg, Germany, 1997, <https://doi.org/10.1007/978-3-662-05746-9>
 27. Lee, W. C. Y., *Mobile Communications Engineering* (2nd ed.), pp 203-240, McGraw Hill, New York, USA, 1998.

Paper submitted: 13.10.2016
Paper revised: 21.11.2016
Paper accepted: 05.12.2016