

Wind Speed Distribution Direct Approximation by Accumulative Statistics of Measurements and Root-Mean-Square Deviation Control

Vadim Romanuke* (Professor, Polish Naval Academy, Gdynia, Poland)

Abstract – In order to accurately estimate wind farm output and subsequently optimise it, a method of wind speed distribution approximation is suggested. The method is based on period-by-period accumulation of wind speed measurements, transforming them into empirical probabilities, and observing the moving approximation to the expected power produced by the wind turbine or entire wind farm. A year is a minimal term during which wind statistics are to be accumulated. The sufficient validity and reliability of the wind speed distribution approximation is supported by controlling root-mean-square deviations and maximal absolute deviations with respect to the moving average of the expected power. The approximation quality can be regulated by adjusting constants defining the requirements to the moving deviations.

Keywords – Accumulative statistics; Approximation; Empirical probabilities; Moving parameter; Root-mean-square deviation control; Wind power; Wind speed distribution.

I. INTRODUCTION TO WIND POWER OPTIMIZATION

Wind power is the most promising source of renewable and safe energy supply. Wind power optimization is a complex problem consisting of the three key components: modelling wind statistics (over the area where a wind farm is deployed) [1], [2], wind turbine construction optimization [1], [3], and the wind farm layout optimization problem [4], [5]. Each of these components is still an open problem influenced mainly by uncertainties [2], [6], [7]. The uncertainties are caused by wind speed volatility and seasonal fluctuations, as well as by variability in incorporating large amounts of wind power into a grid system [2], [8].

While the wind turbine construction has been attempted to be improved independently, the wind statistics of an area directly influences the wind farm layout optimization problem [6], [9]. Knowing the wind statistics of an area with an appropriate accuracy allows for grid balancing by matching the supply of energy to demand. Then the wind farm provides nearly the best efficiency of the investments in it and the electric power output. However, gathering wind statistics is itself an uneasy and long-lasting procedure. Moreover, even after gathering and processing, wind statistics have irremovable variations and fluctuations due to weather unpredictability and inconstancy

[1], [6]. Therefore, wind power optimization still remains a relevant field of scientific research.

II. BACKGROUND AND MOTIVATION

Wind statistics of an area are gathered to build a wind speed distribution of the area whereon a wind farm is projected. The wind speed distribution is commonly modelled as the Weibull distribution of wind speed s [1], [6], [9], [10]:

$$p(s; a, b) = \frac{b}{a} \cdot \left(\frac{s}{a}\right)^{b-1} \cdot e^{-\left(\frac{s}{a}\right)^b} \quad \text{by } s \in [0; \infty) \quad (1)$$

with a positive shape parameter b and a positive scale parameter a . The shape parameter relates to a factual range of wind speeds: the range narrows as the shape parameter increases [10]. The scale parameter corresponds to the mean wind speed, although it also influences the factual range of wind speeds: as the scale parameter increases, the range becomes wider (stretching out to the right) and the mean wind speed grows [10]. Obviously, these parameters of wind speed distribution (1) cannot be assessed as point estimates. They are assessed as interval estimates, although the intervals are relatively narrow [2], [6]. However, the information about winds and weather of the wind farm area is accumulated permanently, so the shape and scale parameters may be additionally re-assessed (and made more accurate) even when the wind farm has been deployed and incorporated into a grid system.

Accurate estimates of the shape and scale parameters are quite important. This is so because the expected power produced by the wind farm is very sensitive to small changes in wind speed distribution (1) defined by b and a . In particular, the expected power produced by a wind turbine of type k is [6], [9]

$$\rho_k = \int_0^{\infty} p(s; a, b) w(s; k) ds, \quad (2)$$

* Corresponding author.
E-mail: romanukevadimv@gmail.com

where function $w(s; k)$ shows a power in megawatts (MW) produced by this turbine at wind speed s . Function $w(s; k)$ is the power curve of the wind turbine of type k . Whereas a lot of scientific papers deal with wind power optimization based on the assumption that wind speed distribution (1) is known and fixed, the uncertainties in the shape and scale parameters are not taken into account. For instance, article [9] suggests a rapid method of maximizing the produced energy by simultaneously minimizing the costs dealing with uncertainties caused by the annual desired energy and the wind farm capacity assumptions. The computational core of the suggested method is that the expected power output is maximized via solving integer linear programming problems considering the shape and scale parameter constant. The method returns a row of optimal numbers of wind turbine types to be installed, and the respective annual energy along with costs spent on it. Thus, in an example with the five known and widespread wind turbines (2.3 MW Enercon E82 E2, Gamesa G128-4.5 MW, 2.5 MW Nordex N90/2500, 2 MW REpower MM82, Vestas V112-3.0 MW) by planning to produce maximum 50 000 MWh annually (see also [6], [9]), where $b = 2$ and $a = 5$ (which corresponds to the mean wind speed of 4.43 m/s), the optimal solution is to install 2 Enercon turbines and 21 REpower turbines. A wind farm of these 23 wind turbines will produce on average 45 480.32904427 MWh annually, and deployment of this wind farm costs 67.39445656 million EUR. However, the same problem at $a = 6$ (the mean wind speed of 5.32 m/s) has a different solution. In this case, it is optimal to install just a single Nordex turbine and 13 REpower turbines which cost 41.09368393 million EUR. Nevertheless, the expected annual energy then becomes even greater: it is 45 537.5008265 MWh, which exceeds the annual energy from the wind farm of the 23 turbines by 0.1257%. Here, almost the same energy costs 39% less. Moreover, the corresponding wind farm is of just 14 wind turbines occupying fewer locations of the area.

The described example might seem rather exaggerated because the inaccuracy in estimating the scale parameter from 5 to 6 is uncommonly rough. However, the solution under a more real uncertainty, when $a = 5.1$ (which corresponds to the mean wind speed 4.52 m/s), badly differs from the solution for $a = 5$. The expected annual energy in 45 671.7216983 MWh produced by 22 REpower turbines will cost 63.52656992 million EUR. Once again, almost the same energy will cost 5.74% less than in the case with $a = 5$.

Thus, estimation of wind speed distribution (1) via its parameters b and a should be extremely accurate and reliable to determine the stable expected power by (2). On the other hand, the wind speed distribution may be approximated directly, without tying to the Weibull distribution. In this case, no distribution parameters are required. The expected power is calculated numerically without knowing a mathematical description of the wind speed distribution curve. The only requirement is high reliability and validity of the approximated distribution.

III. GOAL AND OBJECTIVES

Issuing from practical unreliability of the shape and scale parameters of wind speed distribution (1), the goal is to suggest a method to approximate the wind speed distribution directly, without determining the distribution type and fitting it. To achieve the goal, a procedure of processing wind speed measurements is to be described first. The procedure is believed to give at least a rough discrete distribution of the wind speed. Second, the wind speed distribution approximation should be developed based on the procedure. The approximation will be expected to meet the requirement of its reliable validity. Finally, the method is to be discussed by comparing to the existing approach and the corresponding conclusions are to be made.

IV. WIND SPEED MEASUREMENTS AND STATISTICS

Wind speeds are usually measured with anemometers. The wind speed measurement accuracy denoted by μ can achieve 0.1 m/s. Therefore, let

$$s = \mu h \quad \text{by} \quad h = \overline{1, h_{\max}} \quad (3)$$

be a discrete wind speed magnitude, where $s_{\text{cut-out}} = \mu h_{\max}$ is a general cut-out speed (an ultimate wind speed, at which any of the wind turbine types used for deploying the wind farm is turned off). The set of wind speed magnitudes given by (3) starts from the lowest speed $s = h$ because $s = 0$ does not contribute to integral (2). Number h_{\max} of these magnitudes is defined by the cut-out speed.

The power curve of the wind turbine of type k is sampled into a set

$$w_{hk} = w(\mu h; k) \quad \text{by} \quad h = \overline{1, h_{\max}}. \quad (4)$$

Then the probability of that wind speed $s = \mu h$ registered at some time period is

$$P_h = P(\mu h) \quad \text{by} \quad h = \overline{1, h_{\max}} \quad (5)$$

and the expected power produced by the wind turbine of type k is

$$\rho_k = \sum_{h=1}^{h_{\max}} P_h w_{hk} \quad (6)$$

instead of integral (2).

Let us suppose that wind speed $s = \mu h$ is registered $u_h^{(1)}$ times during the period of wind speed measurements. Obviously, $u_h^{(1)} \in \mathbb{N} \cup \{0\}$. In practice, it is convenient to consider the period as one day (although a longer period as a week would not be rare). Then the very first rough distribution approximation is a set of relative frequencies

$$P_h^{(1)} = \frac{u_h^{(1)}}{\sum_{i=1}^{h_{\max}} u_i^{(1)}} \text{ by } h = \overline{1, h_{\max}}. \quad (7)$$

In fact, relative frequencies (7) are empirical probabilities (e.g., see [10]).

The next day (or another short period) the wind speed is continued to be measured, and wind speed $s = \mu h$ is registered $u_h^{(2)}$ times, $u_h^{(2)} \in \mathbb{N} \cup \{0\}$. The second wind speed distribution approximation can use now both counts $\{u_h^{(2)}\}_{h=1}^{h_{\max}}$ and $\{u_h^{(1)}\}_{h=1}^{h_{\max}}$. Therefore, a set of empirical probabilities

$$P_h^{(2)} = \frac{u_h^{(1)} + u_h^{(2)}}{\sum_{i=1}^{h_{\max}} (u_i^{(1)} + u_i^{(2)})} \text{ by } h = \overline{1, h_{\max}} \quad (8)$$

becomes the following distribution approximation. On the third day (or another short period identical to the previous one) wind speed $s = \mu h$ is registered $u_h^{(3)}$ times, and the following set of empirical probabilities

$$P_h^{(3)} = \frac{u_h^{(1)} + u_h^{(2)} + u_h^{(3)}}{\sum_{i=1}^{h_{\max}} (u_i^{(1)} + u_i^{(2)} + u_i^{(3)})} \text{ by } h = \overline{1, h_{\max}} \quad (9)$$

is determined, and so forth. The procedure of processing wind speed measurements should be repeated for a year. Without losing generality, it can be stated that after wind speed $s = \mu h$ is registered $u_h^{(m)}$ times on day m , the m -th approximation of the wind speed distribution is found by empirical probabilities

$$P_h^{(m)} = \frac{\sum_{j=1}^m u_h^{(j)}}{\sum_{i=1}^{h_{\max}} \sum_{j=1}^m u_i^{(j)}} \text{ by } h = \overline{1, h_{\max}}, \quad (10)$$

where $m = 1, 2, 3, \dots, 365$.

The described routine is an accumulation of wind statistics, which allows controlling the consistency of the measurements using previous data. This is possible owing to the law of large numbers. If the wind speeds are measured daily at a high measurement rate, then the reliability and validity of empirical probabilities (10) will grow as m increases.

V. WIND SPEED DISTRIBUTION APPROXIMATION

As the consistency of empirical probabilities (10) is expected to be tending to improve period by period (literally, day by day, in this case), there may be such day m_* after which the empirical probabilities are sufficient to approximate the wind speed distribution at an acceptable level. The sufficiency will be interpreted by using approximations to expected power (6). Thus,

$$\rho_k^{(m)} = \sum_{h=1}^{h_{\max}} P_h^{(m)} w_{hk} \quad (11)$$

is an approximation to expected power (6) after the m -th period. As m increases, approximation (11) must be more stable. This implies that a series of expected powers approximated by (6) for a few subsequent m 's (i.e., for a few recent periods) should have less variance than a series of the same length for the respective preceding m 's (i.e., for a few earlier periods). This principle is visually explained in Fig. 1.

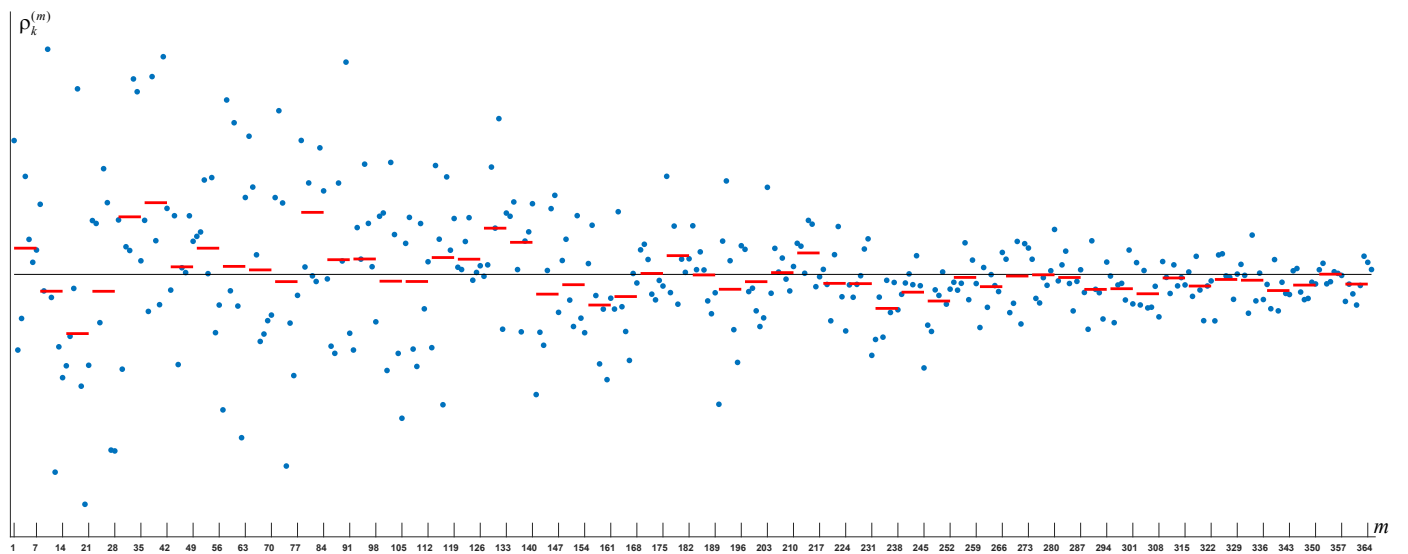


Fig. 1. A visual explanation of the principle of considering the expected power approximation to progress more stable. Each dot representing a day of measurements is an expected power by (11). The length of the measurements is a year, so there are 365 dots overall. The series of seven subsequent m 's (week by week) is used here. The unknown ground truth mathematical expectation is shown as a horizontal line. The week-length expected values are shown as short horizontal lines. The variance is apparently decreasing, and the short lines are "sticking" to the expectation. The sufficiency acceptable level is not shown, though.

Thus, let Δt be a time window between measurement periods

$$(l-1)\Delta t + 1 \text{ and } l\Delta t, \quad l=1, 2, 3, \dots$$

Then the average of the expected power approximations across this window is

$$\tilde{\rho}((l-1)\Delta t + 1, l\Delta t) = \frac{1}{\Delta t} \sum_{m=(l-1)\Delta t+1}^{l\Delta t} \rho_k^{(m)}. \quad (12)$$

An example of how this average changes, being calculated week by week, is shown in Fig. 1 as short horizontal lines. Further, the root-mean-square deviation of the expected power approximation across window Δt is the square root of its variance:

$$\begin{aligned} \sigma_{\tilde{\rho}}((l-1)\Delta t + 1, l\Delta t) &= \\ &= \sqrt{\frac{1}{\Delta t} \sum_{m=(l-1)\Delta t+1}^{l\Delta t} [\rho_k^{(m)} - \tilde{\rho}((l-1)\Delta t + 1, l\Delta t)]^2}. \end{aligned} \quad (13)$$

Here, the most preferable case is

$$\sigma_{\tilde{\rho}}((l-1)\Delta t + 1, l\Delta t) > \sigma_{\tilde{\rho}}(l\Delta t + 1, (l+1)\Delta t) \quad (14)$$

for every $l=1, 2, 3, \dots$. Nevertheless, inequalities (14) are not sufficient to consider expected power approximation reliably valid. The matter is that if the root-mean-square deviations are too great, an acceptable stability of the expected power approximation may not be achieved. This is why the following two inequalities should additionally hold starting at some l_* :

$$\eta(l_*) = \frac{\sigma_{\tilde{\rho}}((l_*-1)\Delta t + 1, l_*\Delta t)}{\tilde{\rho}((l_*-1)\Delta t + 1, l_*\Delta t)} < \varepsilon \quad (15)$$

and

$$\mu(l_*) = \frac{\max_{m=(l_*-1)\Delta t+1, l_*\Delta t} |\rho_k^{(m)} - \tilde{\rho}((l_*-1)\Delta t + 1, l_*\Delta t)|}{\tilde{\rho}((l_*-1)\Delta t + 1, l_*\Delta t)} < \lambda \varepsilon \quad (16)$$

for some $\varepsilon > 0$ and $\lambda \geq 1$. In practice, it is relevant to set $\varepsilon = 0.01$, $\varepsilon = 0.005$, or less.

VI. SIMULATION

It is quite obvious that the direct approximation of the wind speed distribution by accumulating statistics of measurements using statement (10) relies on the fact that expected power approximations (11) and their moving average

(12) with its root-mean-square deviation (13) will obey conditions (14)–(16). At least, these conditions are desired to be true, although some violations for a few l are not excluded. Thus, root-mean-square deviation (13) is not necessary to be a set of strictly decreasing values. A violation in inequalities (15) and (16) is not considered a crucial fail of the method likewise.

For simulating the wind speed distribution approximation method, observations (measurements) of wind speed are conventionally made for a year because different seasons have their strong impact on the resulting wind speed distribution. For example, the Enercon E82 E2 (2.3 MW) wind turbine is taken. It is also assumed that the wind speed distribution is Weibull distribution (1) whose parameters are $b=2$ and $a=5$ (which nonetheless are pretended to be unknown for the observer). Therefore, the theoretical expected power is

$$0.24448531 \text{ MW}$$

(surely, unknown for the observer also).

First, a series of 365 instances with values

$$0.24448531 + 0.1 \cdot 9 \cdot (\zeta + 0.5)$$

$$\text{by } \zeta \in (0; 1) \quad (17)$$

is generated by random values ϑ and ζ drawn from the standard normal and uniform distributions, respectively. Their cumulative sum is a model of approximations (11). Figure 2 shows how it progresses along with its moving average (12), root-mean-square deviation (13), and functions $\eta(l)$ and $\mu(l)$.

Second, a similar example by modelling counts

$$\left\{ \left\{ u_h^{(m)} \right\}_{h=1}^{h_{\max}} \right\}_{m=1}^{365}$$

is presented in Fig. 3 by Fig. 4 and Fig. 5 for inequalities (15) and (16).

It is likely that the simple simulation (Fig. 2) reflects fewer statistical peculiarities than the “deeper” simulation (Fig. 3), where the process is modelled from the very start, having just counts of wind speed registration. This may be the reason by which expected power approximations are “sticking” tighter to the expectation in the case of the “deeper” model. The moving root-mean-square deviations (13) decrease faster for Fig. 3 (see Fig. 4), although the decrement is not strict. Besides, it is worth noticing that the difference between functions $\eta(l)$ and $\mu(l)$ is very small. Function $\mu(l)$, which shows maximal absolute deviations with respect to the moving average (12), only slightly dominates function $\eta(l)$, which shows the ratio of the root-mean-square deviations to the moving average (12). Moreover, the domination sometimes is even violated (it can be clearly seen in Fig. 5). The sufficiency acceptable level is achieved quickly for both types of the simulation.

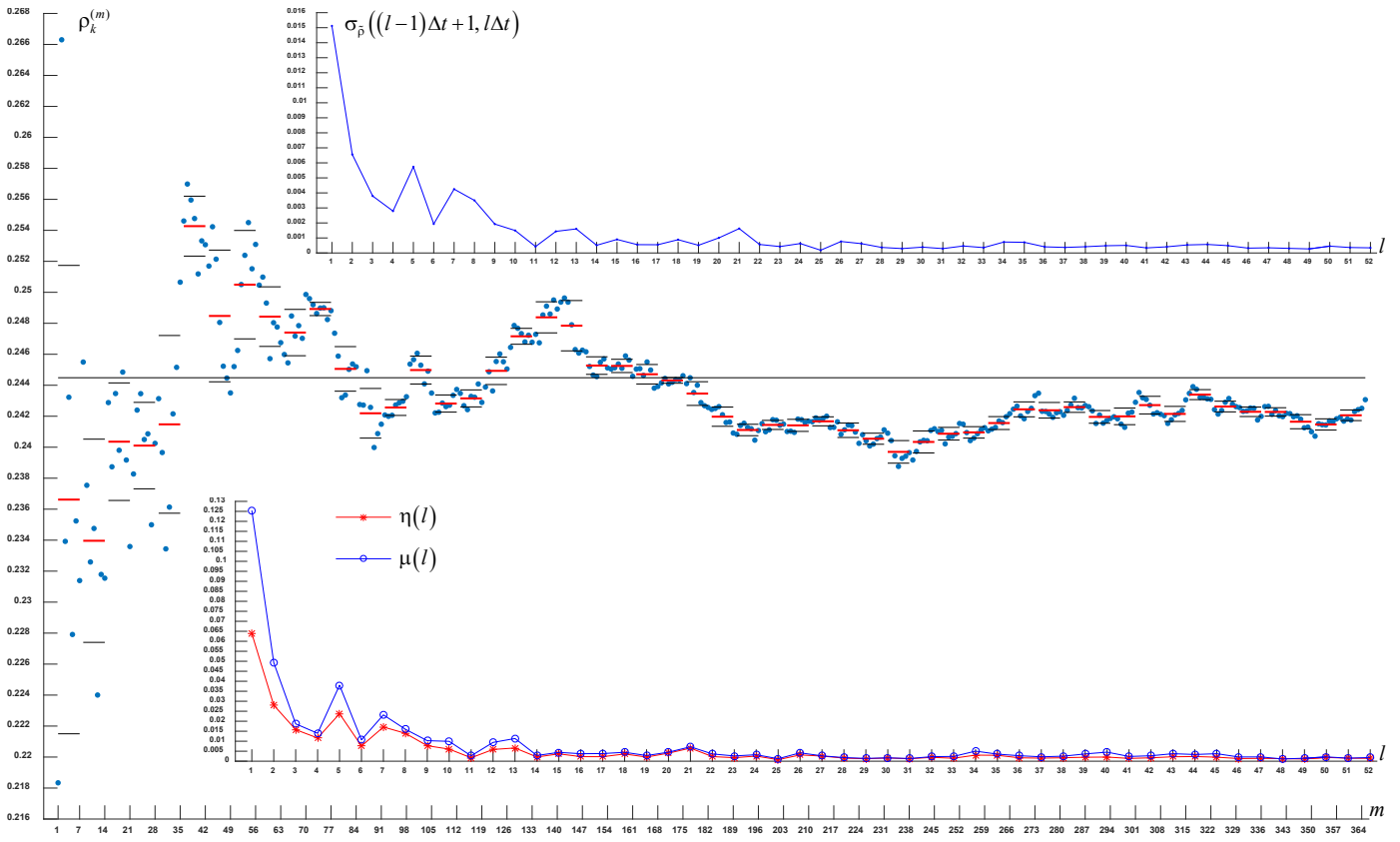


Fig. 2. An example of approximations (11) obtained by the cumulative sum of 365 values (17). The bounds defined by deviations (13) are shown as short lines.

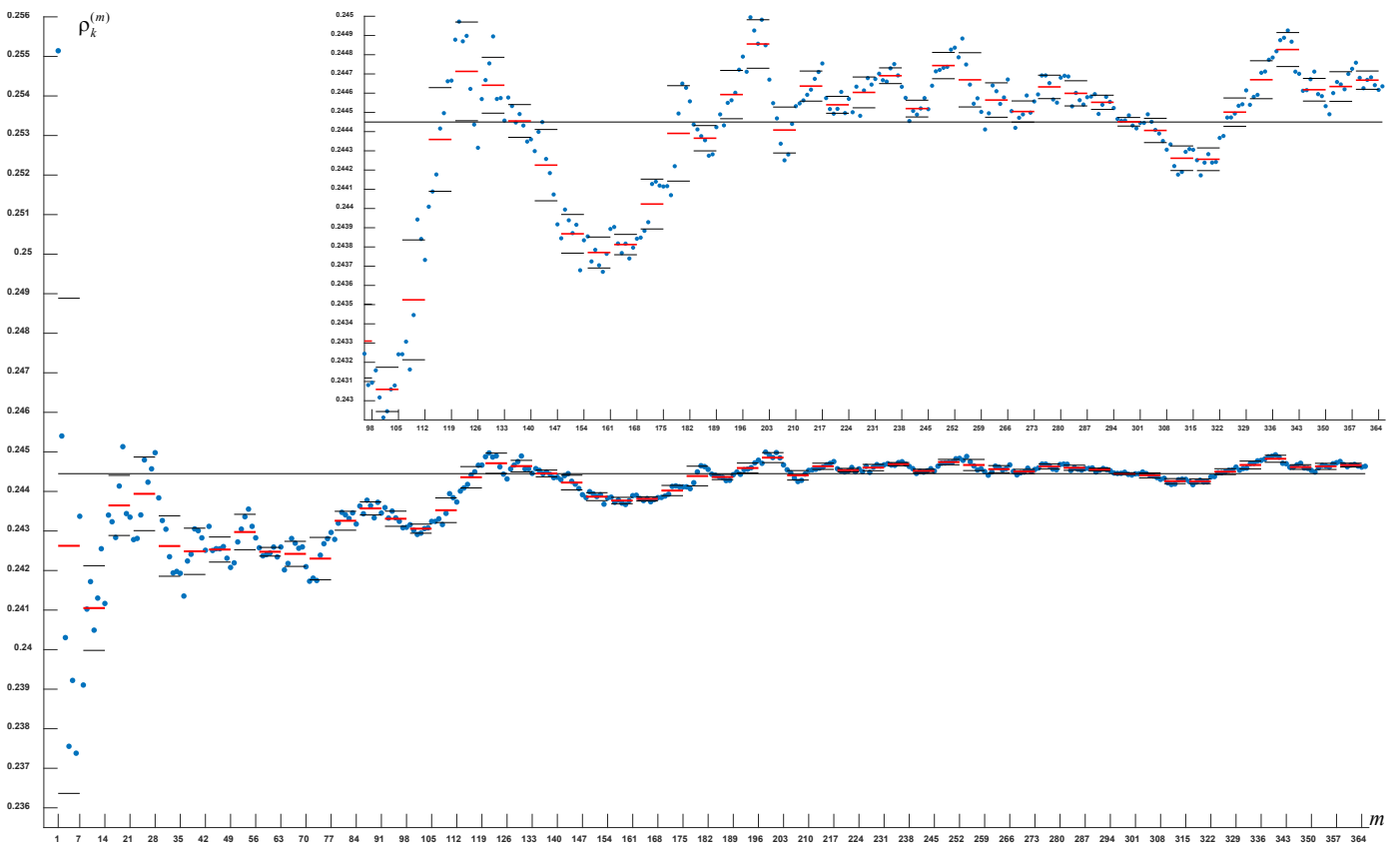


Fig. 3. Another example of approximations (11) obtained by generating wind speed registration counts. This model is far “deeper” compared to the previous one.

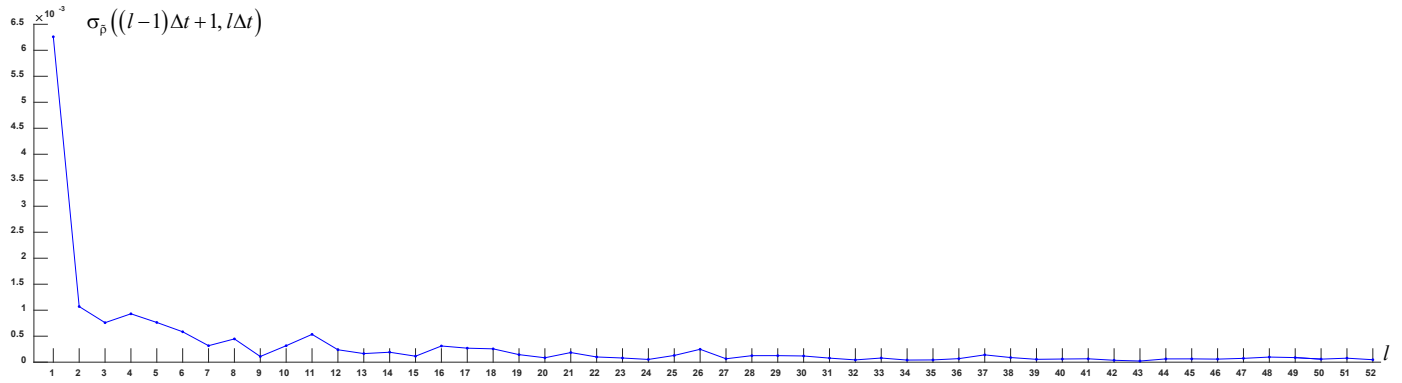


Fig. 4. Root-mean-square deviations (13) for the example in Fig. 3. The moving deviations quickly decrease similarly to those in Fig. 2, although not strictly.

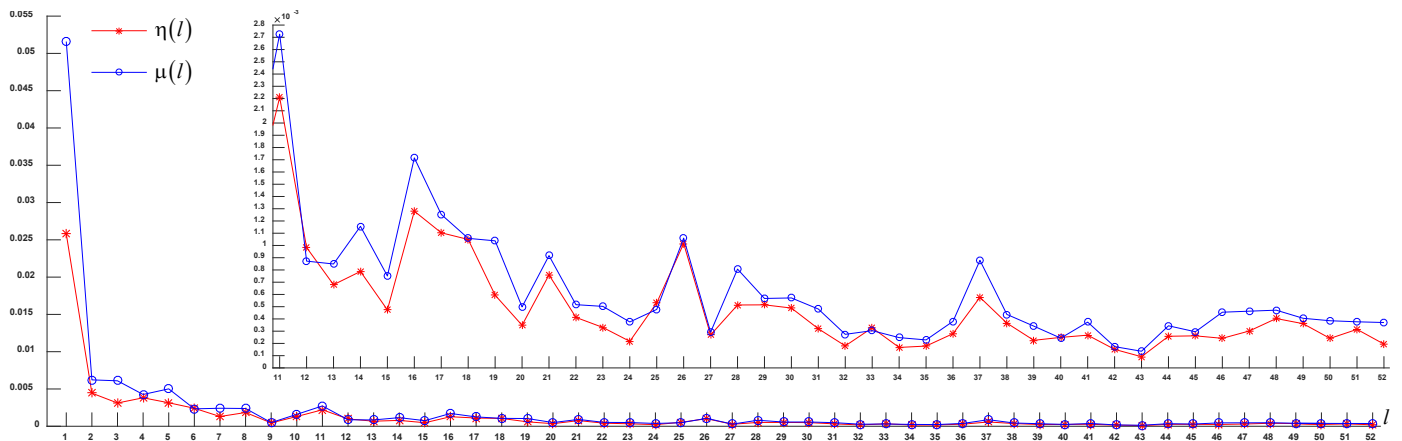


Fig. 5. Functions $\eta(l)$ and $\mu(l)$ showing the progress of ratios in inequalities (15) and (16) for the example in Fig. 3. Both polylines decrease but not strictly.

VII. DISCUSSION

The simulation results seem to be very promising but they do not consider possible seasonal fluctuations of the average wind speed, gusts, unpredictable weather anomalies, etc. Therefore, a year is a minimal term during which wind statistics are to be accumulated. Even when inequalities (14)–(16) hold without violations, the measurements should last no less than a year, whichever ε and λ are.

An important specificity issues from the trace of expected power approximations: while they generally satisfy conditions (14)–(16), the theoretical expected power unknown for the observer may be not so close. This effect is easily seen in Fig. 2, although the expected power approximations in Fig. 3 appear close enough to the unknown expectation (they are “sticking” tighter to the expectation as m increases). This is a subtle drawback of the suggested method, which cannot remove a possible bias with respect to the unknown expectation.

One should distinguish between the wind speed distribution approximation and expected power approximations. The final practical goal is to ascertain a reliably valid distribution that allows calculating the expected power accurately, whichever the configuration of the wind farm is. Nevertheless, the expected power approximations help process the wind statistics with a comprehensive outlook for the electric power output. They can relate as to a certain type of the wind turbine, as well as to an entire wind farm.

VIII. CONCLUSION

The suggested method of wind speed distribution approximation is based on period-by-period accumulation of wind speed measurements, transforming them into empirical probabilities, and observing the moving approximation to the expected power produced by the wind turbine or wind farm. The method does not involve any assumptions about the type of the distribution or its parameters, so it is named the direct approximation. As the wind turbine power curve is presumed to be accurate, then an accurate approximation of the expected power implies that the wind speed distribution is approximated accurately enough. The sufficient validity and reliability of the wind speed distribution approximation are supported by controlling root-mean-square deviations by (14) and (15) and maximal absolute deviations by (16) with respect to the moving average (12). The approximation quality can be regulated by adjusting constants ε and λ defining the requirements to the moving deviations. If the approximation accuracy is desired to be higher, these constants are to be decreased. Lesser Lower magnitudes of the constants better fit longer terms of measurements, and vice versa.

REFERENCES

- [1] T. M. Letcher, *Wind Energy Engineering. A Handbook for Onshore and Offshore Wind Turbines*. Elsevier, Academic Press, 2017.

- [2] P. Breeze, "Chapter 11. Wind Power," in: *Power Generation Technologies (Third Edition)*, Breeze P. (ed.). Newnes, 2019, pp. 251–273. <https://doi.org/10.1016/B978-0-08-102631-1.00011-0>
- [3] A. Chehouri, R. Younes, A. Ilinca, and J. Perron, "Review of performance optimization techniques applied to wind turbines," *Applied Energy*, vol. 142, pp. 361–388, 2015. <https://doi.org/10.1016/j.apenergy.2014.12.043>
- [4] G. Gualtieri, "A novel method for wind farm layout optimization based on wind turbine selection," *Energy Conversion and Management*, vol. 193, pp. 106–123, 2019. <https://doi.org/10.1016/j.enconman.2019.04.059>
- [5] N. Charhouni, M. Sallaou, and K. Mansouri, "Realistic wind farm design layout optimization with different wind turbines types," *International Journal of Energy and Environmental Engineering*, vol. 10, pp. 307–318, 2019. <https://doi.org/10.1007/s40095-019-0303-2>
- [6] V. V. Romanuke, "Wind farm energy and costs optimization algorithm under uncertain parameters of wind speed distribution," *Studies in Informatics and Control*, vol. 27, no. 2, pp. 155–164, 2018. <https://doi.org/10.24846/v27i2y201803>
- [7] M. Santos and M. González, "Factors that influence the performance of wind farms," *Renewable Energy*, vol. 135, pp. 643–651, 2019. <https://doi.org/10.1016/j.renene.2018.12.033>
- [8] W. Du, Y. Wang, H. F. Wang, and Q. Fu, "Open-loop modal analysis to identify the SSO instability risk caused by grid-connected wind farms," *International Journal of Electrical Power & Energy Systems*, vol. 107, pp. 352–362, 2019. <https://doi.org/10.1016/j.ijepes.2018.12.013>
- [9] V. V. Romanuke, "Iterative power maximization by one-half cost dichotomy for optimizing wind farm deployment," *KPI Science News*, vol. 4, pp. 44–53, 2019. <https://doi.org/10.20535/kpi-sn.2019.4.177315>
- [10] S. Campisi-Pinto, K. Gianchandani, and Y. Ashkenazy, "Statistical tests for the distribution of surface wind and current speeds across the globe," *Renewable Energy*, vol. 149, pp. 861–876, 2020. <https://doi.org/10.1016/j.renene.2019.12.041>

Vadim V. Romanuke was born in 1979. Higher education was received in 2001. In 2006, he received the Degree of Candidate of Technical Sciences in Mathematical Modelling and Computational Methods. The Candidate dissertation suggested a way of increasing interference noise immunity of data transferred over radio systems. The degree of Doctor of Technical Sciences in Mathematical Modelling and Computational Methods was received in 2014. The Doctor-of-Science dissertation solved a problem of increasing efficiency of identification of models for multistage technical control and run-in under multivariate uncertainties of their parameters and relationships. In 2016, he received the status of Full Professor.

He is a Professor of the Faculty of Mechanical and Electrical Engineering at the Polish Naval Academy. His research interests concern decision making, game theory, statistical approximation, job scheduling, and control engineering based on statistical correspondence. Vadim Romanuke has good programming skills in Matlab. For practical implementations, Vadim Romanuke uses Python. Since 2019, Vadim Romanuke has been participating as a scientific collaborator in two budget grant studies concerning automation in navigation.

Address for correspondence: 69 Śmidowicza Street, Gdynia, Poland, 81-127.

E-mail: romanukevadimv@gmail.com

ORCID iD: <https://orcid.org/0000-0003-3543-3087>