Methodology and Precision Research of Wind Farm Power Prediction

Rimantas Pranas Deksnys[1], Aldas Stankevicius[1]
[1]*Department of Electrical Power systems, Kaunas University of Technology, Studentu St. 48–144, LT-51368 Kaunas, Lithuania*
rimantas.deksnys@ktu.lt

***Abstract*—The article presents a newly developed statistical regression wind farm power change prediction model. Results of the research and the data analysis performed show that the model is able to evaluate factors determining the wind farm gross output and to increase the prediction accuracy. The influence of the regression equation independent variables on the dependent one is determined by the means of the Students t-test, and the levels of varying equation coefficients are established. The transformation of weather density and wind speed product corresponding to the linear stochastic dependence of the variable power characteristic part for the wind power plant is estimated. Expression of the transformation is suitable to use for predicting the wind farm power in the range from the minimum values to the installed ones. The statistical regression model of the wind farm power prediction is presented on the basis of given technique of the linear regression analysis, the exponential regression equation, and variable coefficients of regression equation. Results of power prediction by the given model precision research show that the one-day relative average prediction error does not exceed 7.52 % of the installed value.**

***Index Terms*—Wind farms, wind power generation, power prediction, prediction precision.**

I. INTRODUCTION

The prediction of gross output of wind power plant is relevant to the selection of building place, identifying possible wind power generation volumes and determining the possibilities of wind power plant efficiency and cost-effectiveness options. Wind farm (WF) power prediction is necessary for the preliminary formation of operation schedules and for setting the system power balancing reserve. WFs power is predicted by the means of *eWindPredictor*, *WPPT*, *WPMS*, *Zephyr*, *Previento*, *LocalPred*, *Sipreolico*, *WEPROG* and other models [1]; also, new, versatile and possibly more precise models are being developed.

II. WIND FARM POWER PREDICTION SYSTEMS AND MODELS

According to the initial data used, WF power prediction models are classified as time series models and numerical weather prediction models (Fig. 1). Time series models use direct wind speed or power measurement data of wind farm (WF), by means of which the WF power is predicted a few hours ahead. Long-term predictions are based on the numerical weather prediction models that are more precise than time series models [2], [3].

Time series model is a simple model and the prediction error for short time periods (a few minutes or hours) might be relatively small, not exceeding a few percent. For longer periods, preciseness rapidly decreases and errors might reach even a few dozen percent. This situation is determined by the velocities of atmosphere processes [1], [2].

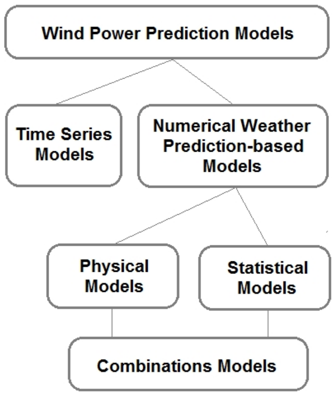

Fig. 1. Types of wind farm power prediction models.

Numerical weather prediction models are classified as physical and statistical models (Fig. 1). In physical models, local wind speed in wind farm territory is evaluated according to the data of numerical weather prediction model and is recalculated to the wind farm power prediction. Statistical models calculate WF power prediction according to the numerical weather prediction and WF power measurement data [1], [4].

WF power prediction using *physical models* is rather inaccurate even for one wind power plant. The accuracy of this model in the case of a wind farm decreases even more, since it is complicated to calculate the wind speed at least a little more precisely in the height of wind turbine axis or average wind speed for all plants of a wind farm. It has been identified that the wind speed might be completely different for every wind power plant. Therefore, aiming to decrease system errors of physical models and to increase preciseness of power prediction, wind direction is taken into account.

Statistical models are based on the inter-relation set between wind farm power measurements and relevant numeric weather predictions. Over time, parameters of such relationship change due to the changing weather conditions and wind speed fluctuations. Moreover, the WF

Manuscript received 23 March, 2016; accepted 11 December, 2016.

characteristics, as well as plants present in the territory of WF, and artificial obstacles are being changed. Therefore, it is necessary to recalculate the parameters of statistical relationship constantly. In statistical models, the relationship between numerical weather prediction and wind farm power measurement data can be set by the adaptive regression or artificial neural network methods [1], [4], [5]. Statistical models possess one evident advantage of the above physical models: the impact of relief, obstacles, atmosphere stability, and other factors are constantly checked and assessed by statistical relationship parameters. Predictions based on statistical models are linked to certain wind farm locality, thus, their system errors automatically decrease. Moreover, statistical models analyse past data and, thus, allow avoiding mistakes. Their main shortcoming is the requirement for the measurement data for a long enough period in order to set a statistical model and calculate regression coefficients. Therefore, such models cannot be used for newly constructed wind farms at the beginning of the performance. Mixed physical-statistical models are being constructed to eliminate this shortcoming.

Prediction models are assessed in accordance with their precision. Precision of physical and statistical wind prediction models is influenced by the accuracy of numerical weather prediction results. Relative average error of wind farm power prediction models depends on prediction period and reaches about 8 % of installed power for 24 hours, 8 %–12 % for 36 hours, and 16 % for 48 hours [1], [6], [7]. Since 2012, a prediction for 24 hours of power generated by all wind farms in Lithuania is performed using physical program model *AIOLOS* developed by Swedish company *VITEC* [8]. Its precision is 11 % with regard to the average power generated by WFs. It shows a high enough precision of physical model, since the error has been calculated with regard to the average power generated, but not the installed one. It is determined that lower power predicting errors of the WFs located in a larger region appear due to the so called spatial alignment effect and autocorrelation of errors [9].

WF power prediction models have their advantages and disadvantages, however all of them are not precise enough. Therefore, the article suggests a more precise statistical regression model for wind farm power prediction.

III. Prediction Methodology of the Statistical Regression of Wind Farm Power

Aiming to predict wind farm power, it is necessary to identify the dependence of the farm gross output on the wind speed, air density, and wind direction, to evaluate the location of plants in the farm, wind and weather parameter prediction discrepancies, the impact of changing external factors, other quantitative and qualitative factors. The task will be solved by means of multiple regression model [10], [11]. Under multiple linear stochastic dependence, random change of independent variables defines the change of a dependent variable with a certain random error

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_m X_m + \varepsilon, \qquad (1)$$

where X_j is independent variables of regression function; β_j is unknown coefficients of regression function of data population researched; and ε refers to random error.

Having calculated the values of the coefficients β_j by the method of least squares, the regression equation can be used for setting the estimate of variable Y with certain precision. Assuming that variable Y_i was observed for n times $Y_1, Y_2,..., Y_n$ and the independent variable X_j was observed for n times as well $X_{1j}, X_{2j},..., X_{nj}$ (where j = 1, 2,..., m). Then the analysis model of multiple linear regression can be expressed by the following matrix equation

$$\boldsymbol{Y} = \boldsymbol{X} \times \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (2)$$

where $\boldsymbol{Y}$ is ($n \times 1$) dependent variable measurement vector; $\boldsymbol{X}$ refers to $n \times (m + 1)$ independent variable measurement matrix; $\boldsymbol{\beta}$ is ($m + 1$) measurement regression equation coefficient vector; and $\boldsymbol{\varepsilon}$ represents ($n \times 1$) measurement random error vector.

The expanded matrix can be expressed as follows

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1m} \\ 1 & X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nm} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \qquad (3)$$

In the analysis model of multiple linear regression, variables $Y, X_1, X_2 ,..., X_m$ are quantitative, measured in interval and relative scales, or dichotomous, of m values. Linear regression analysis model can be applied if data meet certain conditions. The main part of the assumptions for regression analysis are the requirements for random errors ε_i, which show the difference between the observational Y_i value and the value obtained from the regression equation created. These assumptions of linear regression analysis are as follows [12]–[14]:

1. Random errors ε_i are normally distributed.
2. Averages of all ε_i are equal to zero, $E\varepsilon_i = 0$.
3. Dispersions of all ε_i are equal (homoscedasticity assumption).
4. There are no data exclusions.

The first two assumptions are verified using a Kolmogorov-Smirnov criterion.

Supposedly, (3) expresses linear stochastic dependence on the population of variables with a population error ε, which is obtained by replacing actual form of dependence with a linear one. Components of vector β_j are the unknown population regression function coefficients with the index j = 0, 1,..., m. In an observation set, the major part of the realizations of this function Y_i must correspond to the fixed values of the independent variables $X_{i1}, X_{i2},..., X_{im}$, where i = 1, 2,..., n. Based on these data, formulas for identifying regression equation coefficients used for the calculation of regression equation coefficients (estimates) of a set and regression equation coefficients β_j estimates b_j (j = 0, 1,..., m) of a set researched are calculated by the least square method [12]. Applying this method estimates $\boldsymbol{b}$ of coefficients $\boldsymbol{\beta}$ minimizing functions are calculated as follows

$$\mathbf{L} = \sum_{i=1}^{n} \boldsymbol{\varepsilon}_i^2 = \boldsymbol{\varepsilon}^{\mathrm{T}} \boldsymbol{\varepsilon} = (\boldsymbol{Y} - \mathbf{X} \times \boldsymbol{\beta})^{\mathrm{T}} (\mathbf{Y} - \mathbf{X} \times \boldsymbol{\beta}) \rightarrow \min, \quad (4)$$

where T is matrix transpose sign.

The least square value is obtained by solving the system of equations $\partial \mathbf{L} / \partial \boldsymbol{\beta} = 0$

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}} \times \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}} \times \mathbf{Y}. \quad (5)$$

This estimate is unbiased E$\boldsymbol{b} = \boldsymbol{\beta}$, while diagonal elements of the covariant matrix cov$\boldsymbol{b}$ = = σ^2 × $(\mathbf{X}^{\mathrm{T}} \times \mathbf{X})^{-1}$ regression coefficient are the dispersions of estimates b_o, b_1, b_2, …b_m [12].

After setting the coefficient $\boldsymbol{\beta}$ estimates $\boldsymbol{b}$, multiple linear regression function suitable for the prediction of fixed estimates of independent variables of the average Y value is obtained

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_m X_m. \quad (6)$$

If the regression function is determined, it is necessary to check whether the regression function corresponds to the data. The most famous suitability criteria are the standard regression error and certainty factor [15]. Accordance of the sample regression equation to the observation results (X_{i1}, $X_{i2,...,}$ X_{im}, where $i = \overline{1,n}$) is checked. The better the regression function corresponds to the data, the lower is the difference between the estimates Y_i observed and the estimates $\hat{Y}_i$ calculated according to the set regression (6) function. This difference is called a residual error or simply a residual

$$e_i = Y_i - \hat{Y}_i, \quad (7)$$

where e_i is a residual, Y_i represents the estimate observed, and $\hat{Y}_i$ is the average estimate calculated according to the regression equation.

Estimates of errors can be found by using b_j estimates

$$\mathbf{e} = \mathbf{X} - \mathbf{X} \times \mathbf{b}. \quad (8)$$

The dispersion of errors σ^2 is also unknown, thus, its estimate s is also calculated as follows

$$s^2 = \frac{\mathbf{e}^{\mathrm{T}} \times \mathbf{e}}{n-m-1} = \frac{1}{n-m-1} \sum_{i=1}^{n} e_i^2. \quad (9)$$

Three sums of deviation squares are used for the assessment of regression function suitability, [16]:

1. Total sum of deviation squares SS_B, defining the distribution of Y estimates around the average $\overline{Y}$ and equal to $SS_B = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$;

2. The regression sum of deviation squares SS_R, defining a part of the distribution of Y estimates around the average $\overline{Y}$ and explained by linear regression Y in respect of variables X_j, i.e. by their linear dependence, and equal to $SS_R = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2$;

3. The error sum of squares SS_e, defining the distribution of Y estimates around the regression function, i.e. the distribution of Y estimates around the part of the average $\overline{Y}$ that is not explained by the linear regression and equal to $SS_e = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} e_i^2$.

Multiple regression certainty factor r^2 is obtained by dividing the sum of regression deviation squares SS_R to total sum of deviation squares SS_B

$$r^2 = \frac{SS_R}{SS_B}. \quad (10)$$

This coefficient shows what part of the estimate Y distribution around the average value $\overline{Y}$ corresponds to linear regression. The closer r^2 is to unit, the bigger part of the distribution is defined by linear regression, i.e. the regression function better defines the dependent variable Y. In multiple linear regression analysis, the adjusted multiple regression certainty factor, adjusted R Square, r^2_{adj} [10], [11], is used to set which number of the regression equation independent variables m and the size of the set n is necessary to know. The adjusted R Square shows which part of the distribution of the dependent variable Y around the average can be defined by Y linear regression in respect to independent variables X_1, X_2, …, X_m.

Looking for multiple linear regression function, the hypothesis on regression linear dependence is verified. The hypothesis on the estimates of the regression coefficients of all independent variables is verified using the Fisher F criterion with the right critical area and is identified by means of program package SPSS [17]:

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_m = 0,$

H_a : at least one $\beta \neq 0$.

Fisher F criterion with the right critical area is equal to

$$F = \frac{\overline{SS_R}}{\overline{SS_e}} \sim F(m, n-m-1), \quad (11)$$

where $\overline{SS_R}$ is regression average of deviation squares; $\overline{SS_e}$ is residual average of deviation squares.

After setting the rejection of hypothesis H_0 and the acceptance of the alternative hypothesis H_a, it can be stated that the regression is linear, there is at least one coefficient of an independent variable which is not equal to zero, and the regression model is at least partially suitable for predictions. If all regression function coefficients of independent variables are equal to zero, the regression model is not suitable for predictions and the equation is not formed correctly.

In solving the regression analysis equation, it is important to define the influence of the independent variables X_j on a dependent variable Y. Usually, the influence of X_j on the change of Y is verified by the zero hypothesis H_0 $\beta_j = 0$, i.e. whether the coefficient next to X_j in the population regression equation is equal to zero or not. The alternative hypothesis H_a $\beta_j \neq 0$ means the existence of linear dependence between X_j, where $j = 1, 2,\ldots, m$, and Y. The hypotheses on regression equation coefficients are verified using Student statistics

$$T_j = \frac{b_j}{s_{bj}} \sim \mathrm{St}\left(n-m-1\right),\ j = 0,\ 1,\ \ldots, m. \tag{12}$$

The values of Student statistics allow forming the observed significance level p around the rejection of zero hypothesis H_0. If the zero hypothesis is rejected, the coefficient β_j statistically considerably differs from zero, i.e. β_j confidence intervals (CI) with level $1 - \alpha$ are obtained using the formula

$$b_j - t_{1-\alpha/2;n-m-1} \times s_{bj} \le \beta_j \le \le b_j - t_{\alpha/2;n-m-1} \times s_{bj}, \tag{13}$$

where Student distribution with a freedom degree level $n - m - 1$ quintile $\alpha/2$; $j = 0, 1, \ldots, m$ is marked as $t_{\alpha/2;n-m-1}$; s_{bj} is a standard error of coefficient estimates b_j.

When the observational set is not large, even one completely different observation can considerably change the estimates of regression equation coefficients statistically. Therefore, it is important to identify, whether the data have no exclusions. There are various methods for identifying exclusions [10]–[12]. The simplest method is based on the standard residual obtained by contracting the arithmetical average of residual set from the residual e_i and dividing the result by the standard deviation. Standard residual average is equal to 0, while standard deviation is equal to 1. The observation is an exception, if the absolute size of standard residual exceeds three standard deviations [12]. Confidence intervals (CI) of average and individual values are calculated [10], [12] for prediction individual and average values of a dependent variable Y corresponding to given values of independent variables. The calculated upper and lower boundaries of variable Y prediction CI for the lower and upper boundaries prediction of Y individual value are recorded into the data matrix by a statistical applied program package SPSS [12], [13].

It has been identified that the gross output of the wind farm essentially depends on two quantitative independent variables: air density (ρ) and wind speed (v), as well as one qualitative variable, i.e., wind direction (K). Thus, for the solution of wind farm power prediction task formed, it is relevant to apply the regression analysis model for each wind direction: $K \in \{$N, NNE, NE, ENE, E, ESE , SE, SSE, S, SSW, SW, WSW, W, WNW, NW, NNW$\}$.

Therefore (1) can be simplified and worked out as follows

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon, \tag{14}$$

The dependence of the gross output generated by a certain wind farm and its coefficients upon wind speed and air density based on multiple linear regression analysis model and (14) obtained should be calculated. For this purpose, hourly prediction data of atmosphere pressure (p_o), temperature (ϑ), wind speed (v), and wind direction (K) were received from Sventoji station of the Lithuanian Hydrometeorological Service. The statistical hourly data of gross output (P) generated were received from Benaiciai WF. The activity data of four periods in 2009 were used for the establishment of wind farm power dependence and coefficients of (14). Hourly data of atmosphere air pressure and temperature were recalculated to the relative air density ρ according to the formula

$$\rho = \frac{p_o}{R\vartheta}, \tag{15}$$

where p_o is air density (hPa), ϑ is temperature (K), and R is a specific gas constant (287 JK^{-1}kg^{-1}).

Since the characteristics of power and the dependence of gross output P upon the product of air density ρ and wind speed v are not linear in changing a part of the characteristics, various transformations were applied aiming to obtain linear stochastic dependence. It was determined that the most suitable transformation of WF gross output P dependence upon $\rho \cdot v$ into linear dependence is as follows:

$$Y = \mathrm{Ln}(P), \tag{16}$$

$$X_1 = \frac{1}{\rho \times v}. \tag{17}$$

The expression (16) shows that two independent variables can be replaced by one variable of their product. After assessing two independent variables ρ and v as one variable ρv, multiple linear regression analysis expression applied for the solution of the task becomes a linear regression analysis equation of one variable

$$\mathrm{Ln}(P) = \beta_0 + \beta_1 \frac{1}{\rho v} + \varepsilon, \tag{18}$$

where β_0, β_1 are the unknown regression equation coefficients of identified wind directions.

The mentioned changes do not affect the dependence of WF gross output on the product of atmosphere air density and wind speed distribution points. Aiming to identify the values of unknown regression equation (18) coefficients β_o and β_1, linear regression analysis model was researched using an SPSS [13]. The correspondence of the set regression function obtained to the data was verified using regression model suitability indicators, i.e. the standard regression error and certainty index. It has been identified that the empirical function duly corresponds to theoretical function. WF power prediction linear regression function (18) regression certainty coefficients (*R. Squared)* calculated – $r^2 \in [0.961 \div 0.988]$ approaches to unit for all wind directions researched (Table I). Therefore, they are correct. The regression function obtained explains the average distribution of values from 96.1 % to 98.8 % ln *P* by a linear regression in respect of the independent variable $1/\rho v$.

TABLE I. SUITABILITY INDICATORS OF THE LINEAR REGRESSION EQUATION FOR PREVAILING WIND DIRECTION.

Wind Direction	R Squared, r^2	Std. Error off the Estimate	Kolmogorov-Smirnov criterion, *p*
WSW	0.978	0.134	0.442

Standard regression errors (*Std. error off the estimate*), changing in the interval from 0.084 to 0.166 and diminishing to zero were calculated. Linear regression equation suitability indicators set for the prevailing wind direction are delivered in Table II. Errors were compared to each other, and the best function with the minor errors was identified this way. It was identified that the regression function obtained meets the assumptions of the normal distribution, absence of exceptions, and data homoscedasticity of residual errors. After the assessment of the correspondence hypothesis and Kolmogorov-Smirnov criterion estimates, it is possible to conclude that standardized residual distributions are standard and normal, the hypothesis was not rejected for all wind directions, since ($p \in [0.442 \div 0.988]$ and meets the of Kolmogorov Smirnov condition $p > 0.05$.

The assumption of homoscedasticity, or the quality of conditional dispersions is the requirement that the distribution of residuals shall be the same with each fixed X_{1j} estimate or each fixed prediction estimate $\hat{Y}_i$. The closer the estimates distribute around X_1 or $\hat{Y}$ axis, the better regression function describes the data analysed. In the case researched, the distribution of standardized residuals is constant.

The influence of the independent variable $1/\rho v$ on changes of dependent variable *P* was verified using zero hypothesis H_0: $\beta_1 = 0$, which was rejected in favour of the alternative hypothesis H_a: $\beta_1 \neq 0$. It confirms the existence of linear dependence between $1/\rho v$ and *P* and the significance of all sample and population regression equation coefficients, since $p < 0.001$ (Table II).

Estimates *b* of all regression equation coefficients β, standard errors, Student statistics, significance levels, and confidence intervals were identified for all wind directions. The above-mentioned coefficients and parameters are provided in Table II and Table III only for the prevailing wind direction.

TABLE II. COEFFICIENTS AND ASSUMPTION PARAMETERS OF LINEAR REGRESSION EQUATION FOR THE PREVAILING WIND DIRECTION.

Wind Direction	Coefficient of regression equation β_j			Student's statistics	Significance level *p*
	Estimate b_j	Value	Std. Error		
WSW	b_0	4.282	0.039	108.602	0.000
	b_1	-24.051	0.321	-74.928	0.000

The results of analysis performed show that the estimates of linear regression equation coefficients $b_0 = 4.282$ and $b_1 = –24.051$ are determined with the confidence intervals $PI_{0.95}\ \beta_0 = (4.204; 4.361)$ and $PI_{0.95}\ \beta_1 = (–24.686; –23.416)$ for the prevailing West-South-West (WSW) wind direction (Table III). It can be stated on the basis of these data that the average estimate of WF gross output will fall into the confidence interval set. Analogous conditions are valid for the other wind directions.

TABLE III. CONFIDENCE INTERVAL COEFFICIENTS OF LINEAR REGRESSION EQUATION FOR THE PREVAILING WIND DIRECTION.

Wind Direction	Coefficient	95 % Confidence interval for β_j	
		Lower boundary	Upper boundary
WSW	b_0	4.204	4.361
	b_1	–24.686	–23.416

It can be stated on the basis of the analytical results assessed that the linear regression assumptions are met and the linear regression analysis function can be used in practice. Therefore, the WF power of wind direction established can be identified by using the following linear regression equation

$$\mathrm{Ln}(\hat{P}) = b_0 + b_1 \frac{1}{\rho v}. \tag{19}$$

Having applied the inverse transformation, the following exponential regression equation for the prediction of WF power is obtained

$$\hat{P} = e^{b_0 + b_1 \frac{1}{\rho v}}. \tag{20}$$

The results of linear regression function analysis show that the linear regression equation (20) meets the assumptions of linear regression and can be used for the prediction of WF power from minimum to the installed power. Expression (20) can be transformed to the simpler and more comfortable form to use

$$P = C \times e^{-\frac{b}{\rho v}}, \tag{21}$$

where *C* is a coefficient equal to the dimension e^{b_0}, *b* is a coefficient equal to b_1.

Values of WF prediction formula coefficient *C* and *b* for prevailing wind directions are presented in Table IV.

TABLE IV. POWER PREDICTION EXPRESSION COEFFICIENTS.

Coeffici-ent	Wind Direction					
	WSW	W	SW	WNW	S	Not taken into account
C	72.385	71.023	75.189	61.621	52.353	61.744
b	24.051	24.124	24.859	23.143	22.258	23.347

It is determined by the research that the dependence of the average WF power on wind direction is rather significant and can reach quite high differences from the average power under increasing estimates of the product of air density and wind speed. If the product ρv reaches the estimate of 17.5 kg/m^3 m/s, the difference from the average power can reach about 5 MW and be equal to about 30 % of the installed power of the farm. It means that the assessment of wind direction and the prediction of WF power is not only relevant, but necessary.

The dependence of power prediction on the product of air density and wind speed, in the standard atmosphere (p_o = 1013,25 hPa; ϑ = 15 °C; density, ρ = 1,225 kg/m^3; g = 9,81 m/s^2) in Benaiciai WF with generators of V100-2,75 type curves are provided in Fig. 2 according to the expression (21). Benaiciai WF power characteristics (Fig. 2) visually show the influence of wind direction and allow the approximate prediction of gross output generated by the farm under the atmosphere parameters predicted.

Rather significant differences of Benaiciai WF power for various directions are influenced by the location of plants and natural or artificial obstacles causing turbulent air movements.

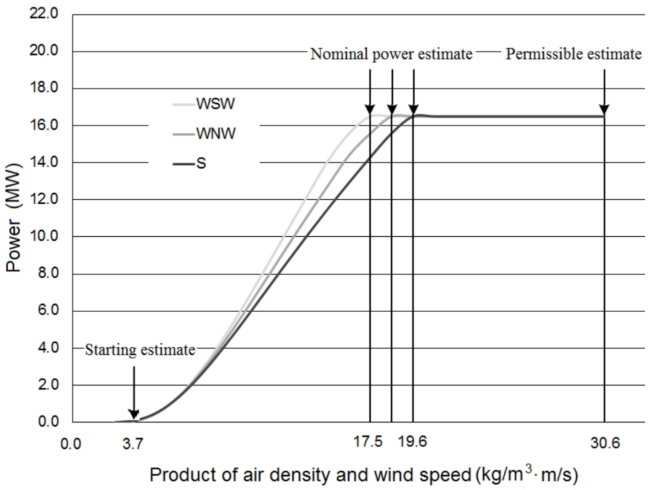

Fig. 2. Dependencies of Benaiciai wind farm power on the values of ρv under standard atmosphere conditions.

Overall, it can be stated, that the statistical regression of WF power prediction model evaluates wind speed and direction, atmosphere parameters, their prediction system errors, plant location within the farm, natural and artificial obstacles, ground roughness, permanent parameter variation and allows a more precise prediction of wind farm power.

IV. Research of the Precision of Statistical Regression Prediction Model of Wind Farm Power

Statistical regression model for the prediction of wind farm power, consisting of the prediction and statistical parts, has been proposed on the basis of linear regression analysis methodology formed in Section II, exponential regression equation obtained (21) and regression equation coefficients for wind directions set (Fig. 3).

The statistical module accumulates prediction data from Hydrometeorological Service of wind speed, wind direction, atmosphere pressure and air temperature for the same period from the closest station. WFs SCADA system directly delivers the information about the gross output and the number of plants operating. These data are grouped according to wind directions, the data of the time when not all plants were operating are rejected. Data of wind farm gross output are analysed and compared to wind parameters using linear regression analysis model, exclusions are identified and rejected. Based on the finally selected data, regression analysis model calculates regression equation coefficients for each wind direction. Coefficients set are periodically renewed and sent to the prediction module.

For the prediction of WF power, the numerical weather prediction block of the prediction module is uploaded with the prediction period data on wind speed and direction as well as other meteorological data grouped according to wind directions. Then air density estimates are calculated for the predicted wind directions, and all necessary regression equation coefficients are chosen from the block of regression equation coefficients according to wind directions. WF power prediction block identifies farm power for the prediction period according to the regression equation (21) and wind directions.

One of the most comprehensive periods, namely 15–22 July 2009, has been chosen for the prediction precision research. Numerical weather prediction data on wind and atmosphere in this period were regenerated by *Harmonie* model. The wind speed and direction of Benaiciai WF were regenerated using the data base of *HIRLAM* model and recalculated to the height of 100 m vertically, while horizontally it were recalculated to ensure that the wind prognosis for the location and WF coordinates matched as much as possible. For the period set, numerical weather prediction on hourly wind and atmosphere data were delivered 24 hours forward and renewed four times per day (12 AM, 6 AM, 12 PM, and 6 PM).

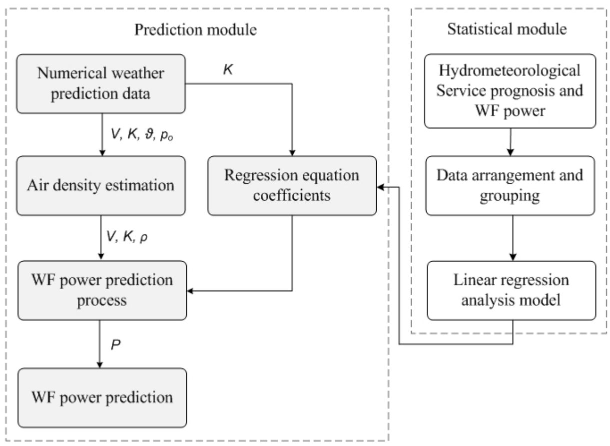

Fig. 3. Wind farm power prediction statistical regression model.

Data of Benaiciai WF actual gross output as well as wind speed and direction values measured in the farm for the same period were used for the research of statistical regression model precision. Estimates of relative average errors (SVP) were calculated hourly for the comparison and precision evaluation of prediction models

$$SVP = \frac{1}{nP_{\text{inst}}} \sum_{i=1}^{n} \left| P_{i\text{fact}} - P_{i\text{pr}} \right|, \qquad (21)$$

as well as relative average square errors (SVKP) in respect of the installed power of the WF

$$SVKP = \frac{1}{P_{\text{inst}}} \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_{i\ \text{fact}} - P_{i\ \text{pr}})^2}, \quad (22)$$

where n is a number of data estimates, $P_{i\ \text{fact}}$ is actual power, $P_{i\ \text{pr}}$ is predicted power, and P_{inst} is installed power of wind farms.

The precision analysis of WF power prediction using the statistical regression model created and the results obtained show that the hourly estimate of one-week relative average error identified daily, taking into account the wind direction, changes from 7.51 % to 9.72 %, and does not exceed 8.66 % of WF installed power. It increases up to 8.16 %–10.91 % of WF installed power without taking into account the wind direction (Fig. 4). In this case, the estimates of relative average errors of weekly duration increase to 9.24 % of WF installed power. For the assessment of model developed, the same power predictions were made by a comparative time series model; and it was identified that hourly relative average error fluctuates within the range of 8.73 %–11.39 %, while the weekly relative average error reaches 9.82 % of WF installed power.

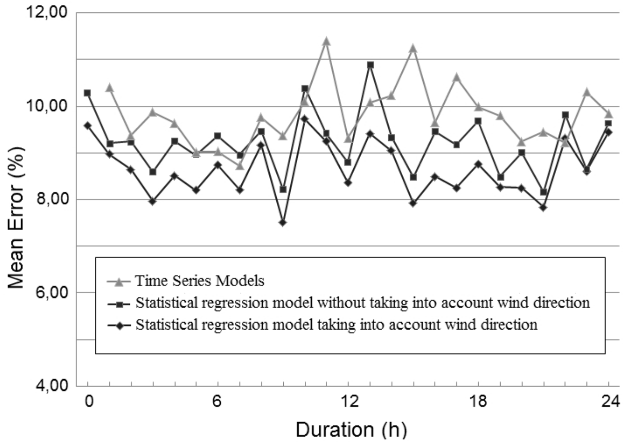

Fig. 4. Relative average errors of the wind farm power prediction.

Analogous relative average square error research shows that one-week identified daily relative average square error estimate changes from 10.3 % to 13.7 %. The weekly estimate of the daily-identified relative average square error does not exceed 11.64 %. Without taking into account the wind direction, it increases to 11.02 %–14.72 %. The weekly estimate of identified daily relative average square error increases to 12.45 % of WF installed power (Fig. 5). The comparative calculations using the time series method show that weekly relative average square error of this model fluctuates within the range of 12.24 %–15.47 %, while weekly relative average square error reaches 14.09 % of WF installed power.

Potentially more accurate prediction of WF power changes during the following 24 hours is necessary for composing the operation schedules of WF power balancing equipment. The prediction is made for 21 July 2009. The initial data of this time interval are given in Fig. 6. The diagrams also show the predicted average wind speed per day, WF actual wind speed of 10 minutes, and the actual average gross output of the farm per 1 minute. The predicted wind speeds and actual wind speeds per day are quite similar, however the precision of prediction is not the same. This may have a quite significant influence on the precision of statistical regression power prediction model.

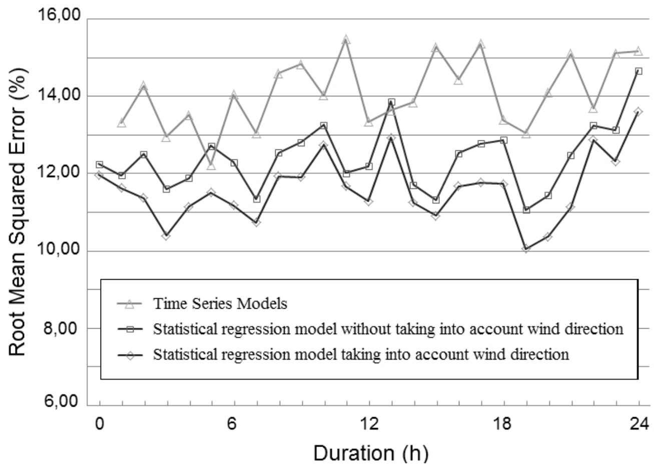

Fig. 5. Relative average square errors of the wind farm power prediction.

Hourly WF power of one day (Fig. 7) was predicted on the basis of initial prediction data and WF power prediction statistical regression model. The initial data were updated every 6 hours (6 AM, 12 PM, and 6 PM). The figure also shows WF average gross output per minute and per hour.

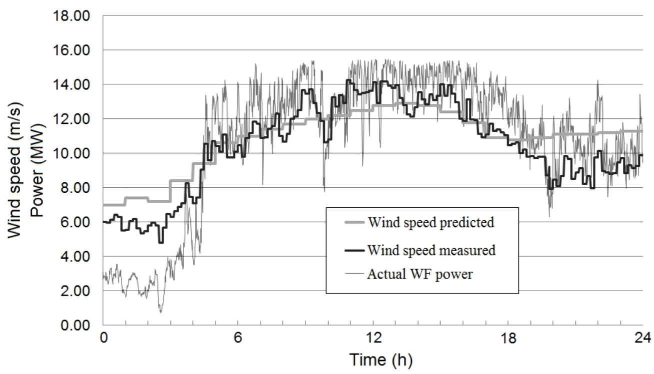

Fig. 6. Wind farm day time predicted and measured wind speed, also the gross output according to the data of 21/07/2009 (the scale of wind speed and power are coincided).

The analysis of predicted and measured data shows (Fig. 7) that the discrepancies in predicted and actual power during the first three hours of operation may be explained by the insufficient wind speed at some wind plants. It is determined that at WF average 5.43 m/s wind speed in 10 minutes interval the wind speed at certain wind plants can change from 3.3 m/s to 5.9 m/s at the same time. Wind plants of V100–2.75 type in Benaiciai farm start operating when the average wind speed of 10 minutes reaches 4.0 m/s and stop when the wind speed is 3.0 m/s. Such WF operation settings are used aiming to reduce the number of unnecessary connections to the electrical network. It shows that the wind plants cannot start operating at the same time if the wind speed is close to the initial WF operation speed depending on the plants layout in the farm and different wind speeds. Also, at the initial wind speed prognosis, it was impossible to have a more precise evaluation of the reduction of wind speed at 10–11 AM and after 6 PM.

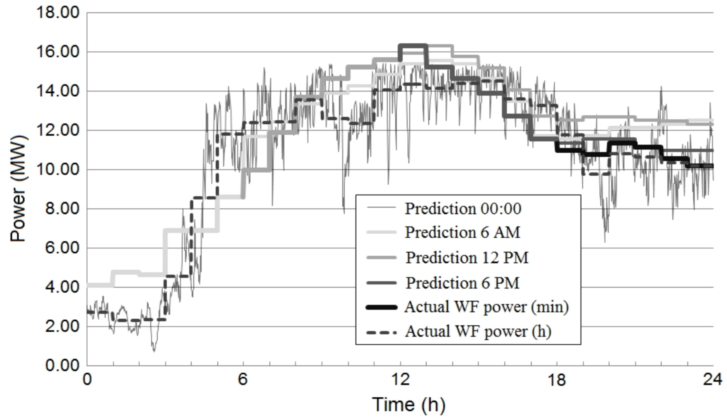

Fig. 7. Wind farm predicted and actual power per day (21/07/2009).

The diagram (Fig. 7) shows that the initial prognosis of 0.00 PM and 6 PM were not precise enough for the last four hours of the day. It also affected the total precision of WF power prediction.

The estimated one-day relative average error for the non-regenerated prediction is 7.52 %, while the estimate of a relative average error for regenerated prediction every 6 hours decreases to 7.42 %.

It is established on the basis of the comparison of statistical regression model power prediction precision with the studies of other authors given in literature that the prediction accuracy of the sequence model of comparative data is 10 %, precision of the physical program model exceeds 11 % and accuracy of other physical and statistical models reaches approximately 8 %.

The results obtained allow to state that the statistical regression prediction model developed is suitable for WF power prediction and can be a versatile prediction model, taking into account not only the wind parameter errors, but also the effect of changing WF environmental factors.

V. CONCLUSIONS

The statistical regression of wind farm power prediction model allows a more precise prediction of wind farm power, taking into account the dependence of gross output of the farm on wind speed, wind direction, atmosphere parameters, their system errors, plant location in a farm, natural and artificial obstacles, ground roughness, permanent changes of the parameters, and rejecting data unsuitable for prediction.

A one-day relative average error of wind farm power prediction does not exceed 7.52 % of the installed power.

It has been identified that the accuracy of the power prediction model of the wind farm essentially depends on the precision of wind speed and atmosphere parameter predictions, as well as the evaluation of the number of operating plants.

The regeneration of wind and atmosphere parameter prognosis during a day allows achieving more accurate predictions of wind farm power.

REFERENCES

[1] X. Wang, P. Guo, X. Huang, "A review of wind power forecasting models", *Energy Procedia*, vol. 12, pp. 770–778, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.egypro.2011.10.103

[2] A. Carpinone, M. Giorgio, R. Langella, A. Testa, "Markov chain modeling for very-short-term wind power forecasting", *Electric Power Systems Research*, vol. 122, pp. 152–158, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.epsr.2014.12.025

[3] W. Chang, "A literature review of wind forecasting methods", *Journal of Power and Energy Engineering*, vol. 2, pp. 161–168, 2014. [Online]. Available: http://dx.doi.org/10.4236/jpee.2014.24023

[4] A. Foley, P. G. Leahy, A. Marvuglia, E. J. McKeogh, "Current methods and advances in forecasting of wind power generation", *Renewable Energy*, vol. 37, no. 1, pp. 1–8, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.renene.2011.05.033

[5] P. Gomes, R. Castro, "Wind speed and wind power forecasting using statistical models: autoregressive moving average (ARMA) and artificial neural networks (ANN)", *International Journal of Sustainable Energy Development (IJSED)*, vol. 1, no. 1/2, 2012. [Online]. Available: https://doi.org/10.20533/ijsed.2046.3707.2012.0007

[6] J. Junga, R. Broadwater, "Current status and future advances for wind speed and power forecasting", *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 762–777, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.rser.2013.12.054

[7] A. Yamaguchi, T. Ishihara, "Wind power forecasting by using physical and statistical approach", *European Wind Energy Conf. & Exhibition (EWEC 2010)*, Warsaw, Poland, 2010, p. 3432. [Online]. Available: http://windeng.t.u-tokyo.ac.jp/ishihara/proceedings/2010-2_poster.pdf

[8] A. Baranauskas, M. Azubalis, "The evaluation of the load and wind power forecasting errors", in *Proc. 9th Int. Conf. Electrical and control technologies (ECT 2014)*, Kaunas, Lithuania, 2014, pp. 162–167.

[9] N. Gasset, M. Landry, Y. Gagnon, "A comparison of wind flow models for wind resource assessment in wind energy applications", *Energies*, vol. 5, pp. 4288–4322, 2012. [Online]. Available: http://dx.doi.org/10.3390/en5114288

[10] S. Chatterjee, J. S. Simonoff, *Wiley Handbooks in Applied Statistics: Handbook of Regression Analysis*. NJ, USA: Wiley, 2013, p. 252.

[11] J. Marija Norusis, *IBM SPSS 19.0 Guide to Data Analysis*. USA: Pearson, 2012, pp.672.

[12] V. Cekanavicius, G. Murauskas, *Statistika ir jos taikymai*. Vilnius: TEV, 2004, p. 272. (in Lithuanian)

[13] J. T. McClave, *Statistics for business and economics*. Upper Saddle River: Pearson Prentice Hall, 2006.

[14] A. S. Gaur, *Statistical methods for practice and research: a guide to data analysis using SPSS*. New Delhi: Response Books, 2007.

[15] V. Janilionis, *Daugialypes regresines analizes taikymas socialiniuose tyrimuose*. Lietuvos HSM duomenu archyvas. [Online]. Available: http://www.lidata.eu/index.php?file=files/mokymai/Janilionis_IV/jan_IV.html&course_file=jan_IV_1.html (in Lithuanian).

[16] V. Janilionis, *Koreliacines ir regresines analizes pagrindai*. Lietuvos HSM duomenu archyvas. [Online]. Available: http://www.lidata.eu/index.php?file=files/mokymai/Janilionis_III/jan_III.html&course_file=jan_III_2.html (in Lithuanian).

[17] *SPSS software*. USA: IBM Corporation.: [Online]. Available: http://www.spss.com/