Annals of Warsaw University of Life Sciences – SGGW
Agriculture No 70 (Agricultural and Forest Engineering) 2017: 71–78
(Ann. Warsaw Univ. Life Sci. – SGGW, Agricult. 70, 2017)
DOI 10.22630/AAFE.2017.70.19

Application of measure–correlate–predict method for wind speed prediction at Łódź Hills

RAFAŁ KORUPCZYŃSKI
Department of Fundamental Engineering, Warsaw University of Life Sciences – SGGW

Abstract: *Application of measure–correlate–predict method for wind speed prediction at Łódź Hills*. In this paper an application of measure–correlate–predict method has been presented. For particular location a statistical model of linear regression between two sets of data has been built. The real measurement from meteorological station and reanalysis data from ERA-Interim database have been used. The model has been applied to long-term data for accurate wind speed prediction. Results are useful for wind resource assessment at this location.

Key words: measure–correlate–predict, wind energy resources, wind farm, renewable energy

INTRODUCTION

In the European Union, the renewable energy sources have to be 20% of overall energy sources to year 2020 [Directive 2009/28/EC]. In 2015, about 13% of total primary energy production of renewable energy was from wind energy [Eurostat 2017]. Because of increased efficiency of using this energy source, in the central regions of USA wind power is the cheapest type of energy, about 30 USD per 1 MWh [Forbes 2017]. From the other side, the investments cost is quite high, for example, in Poland about 6,000,000 PLN, 1,417,000 EUR) for 1 MW of installed power [Pesta 2009]. Hence, economical success of wind farm project is very sensitive to initial estimation of future profits, which are based on energy gain. It requires to find the location with possibly the best wind energy resources. Moreover, a key problem is an assessment, as accurately as possible, wind energy potential at given location. Power of wind and electrical energy production of wind turbine is proportional to wind speed to the third power. Wind speed measurements are necessary limited to only one or at most, a few years [MEASNET Procedure 2016]. For a good determination of the multi-annual, typical wind conditions, there is a need of using a database of long-term wind measurements from another (but relatively close and similar) location. There is a possibility of linking above two locations by statistical methods called MCP (abbreviation measure–correlate–predict method) [Carta et al. 2013]. Hence, long-term data are the basis of prediction of wind resources at target site [Dinler et al. 2013]. Without this strategy, reliable assessment of wind performance at the concerned site from one-year measurements is practically impossible [Messac et al. 2012]. In this paper, MCP method was used for estimation of wind energy resources at Łódz Hills. Because of easy access to measurement data, a meteorological station at Agriculture Experimental Farm (AEF) of Warsaw University of Life Sciences – SGGW at Żelazna was used. This localization (agricultural fields of AEF) seems to be suitable for a wind farm construction. Application of MCP method is a first level of wind resource evaluations at this site.

MEASURE–CORRELATE–PREDICT METHOD

Measure–correlate–predict method helps to estimate of wind energy resources of particular location. There are required two data sets: one- to three-year measurements [Weekes et al. 2015] of wind speed from expected wind farm location: target site, and long-term measurements, i.e. 10 to 30 years or more [Velázquez et al. 2012], from the so-called reference site, i.e. relatively near meteorological station. At least, the same year of data can be present in both data sets [Rogers et al. 2005]. The terrain, vegetation, local obstructions and type of weather have to be almost the same in both locations. The data from the same year and from both location are used to determine correlation each other. This correlation is transferred to long-term reference data for creation of predicted data at target site [Perea et al. 2011]. There are many method of correlation determination: linear regression [Rogers et al. 2005, Thøgersen et al. 2007, Weekes et al. 2015], variance ratio [Rogers et al. 2005], artificial neural network [Velázquez et al. 2011, Zhang et al. 2013], support vector machines [Díaz et al. 2017]. In this paper the first method was used.

There are two types of linear regression models: simple linear regression and multiple linear regression [Larose 2008]. In the simple linear regression there is a function which describe dependency between independent variable (predictor) x, and response (predicted) variable y:

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{1}$$

Multiple regression has one difference: there are many independent variables $x_1, \ldots, x_n$:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \tag{2}$$

The last term of both equations is error (ε), which represent some indeterminacy of the model [Larose et al. 2015]. Coefficients β_0 and $\beta_1, \ldots, \beta_n$ can be calculated by least square fit method. In the case of wind resources, y is the wind speed at target site (v_t), and x is the wind speed at reference site (v_r). More dependent variables may be: day of the year (d), atmospheric pressure (p), temperature (t). Not all dataset is used for calculation of regression coefficients, but only 80% (train dataset). Remaining part (20%) – validation dataset – is used for validation of the obtained regression model.

The goodness of data fit by equation (1) is represented by coefficient of determination r^2:

$$r^2 = \frac{\sum_i (\hat{y}_t - y_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{3}$$

where:

$\hat{y}_t$ – predicted value of dependent variable y_i;

$\bar{y}$ – mean value of dependent variable y_i.

In the case of equation (2), r^2 is replaced by adjusted one (R^2_{adj}), which is so-called the penalized form of coefficient of determination [Larose et al. 2015]:

$$R^2_{adj} = 1 - \left(1 - R^2\right)\frac{n - 1}{n - m - 1} \tag{4}$$

where:
n – sample size;
m – total number of explanatory variables x_n in the model.

As noted in Carta et al. [2013], good quality of wind resources prediction requires R^2 above 0.70.

Second measure of the model accuracy is root mean square error (*RMSE*) [Díaz et al. 2017]:

$$RMSE = \sqrt{\frac{1}{n}\sum_i \left(y_i - \hat{y}_t\right)^2} \tag{5}$$

After model validation, long-term dataset is independent variable of obtained regression equation.

As the result of MCP method is time series of predicted wind speed at target site.

DISTRIBUTION OF WEIBULL

The description of wind speed time series is only possible by statistical methods. One of this methods is use of Weibull probability distribution [Carillo 2014]:

$$f(v) = \left(\frac{k}{c}\right)\left(\frac{v}{c}\right)^{k-1} e^{-\left(\frac{v}{c}\right)^k} \tag{6}$$

where:
$f(v)$ – probability of wind speed (v) [-];
k – shape parameter [-];
c – scale parameter [m·s^{-1}].

Cumulative Weibull distribution is [Bhattacharya et al. 2010]:

$$F(v) = 1 - e^{-\left(\frac{v}{c}\right)^k} \tag{7}$$

where
$F(v)$ – cumulative probability of wind speed (v) [-].

After the natural algorithm was taken from both sides of equation (7), it become:

$$\ln\left\{-\ln\left[1 - F(v)\right]\right\} = k\ln(v) - k\ln(c) \tag{8}$$

Then, the parameters of distribution may be calculated by simple regression method, as it is applied to equation (8), where [Bhattacharya et al. 2010]:

$$y = \ln\left\{-\ln\left[1 - F(v)\right]\right\} \tag{9}$$

$$\beta_0 = -k \ln(c) \tag{10}$$

$$\beta_1 = k \tag{11}$$

$$x = \ln(v) \tag{12}$$

Coefficients k and c are clearly characterize probability of wind speed at target location.

DATASETS

The used data set composes from two parts: real meteorological measurements from station which is located at agricultural fields of AEF (target site) and ERA – Interim reanalysis data acquired from European Centre for Medium-Range Weather Forecasts (reference site) [Dee et al. 2011]. This data source is attractive for MCP, because is easy accessible for free, and the time range is quite wide [Carta et al. 2013]. In this paper a selected time period was from 1979 to 2016 year. Among the wind speed at 10 m above ground level (hereafter referred to as a.g.l.), the wind direction, atmospheric pressure, temperature were used. The averaging time for ERA reanalysis is 6 h. The selection of reanalysis was based on the lack easy accessible reference meteorological station with long-term dataset.

The measuring station is at 51.875286 N, 20.112762 E. Wind speed was measured at 12 m a.g.l. Beside wind speed, some other parameters was measured [Bakoń et al. 2017]. Results from year 2008 (both measurements and reanalysis data) are used for calculation of regression model. Because of some measurement equipment failures the total data availability for 2008 year was about 90%.

RESULTS AND DISCUSSION

In the first case simple linear regression was applied. The independent variable was wind speed (v_r) from reanalysis (reference) and dependent variable was measured wind speed (v_t) at 12 m a.g.l. at Żelazna (target). Assumed significance level α was equal 0.005. The regression coefficients and standard errors were calculated by least square fit method (training dataset, $n = 1,032$) in R Statistical Software [R Development + Core Team 2008]. Results indicated that model is statistically significant (F-statistic value: 4,108; $p < 2.2 \cdot 10^{-16}$). Coefficient of determination (r^2) and root-mean-square error (*RMSE*) was calculated with validation dataset ($n = 259$). Values are presented in Table 1. On the Figure 1 the comparison of real and predicted wind speed was presented, as an illustration of model accuracy based on validation data.

TABLE 1. Parameters of simple linear regression model

Parameter	Value	*SE*
β_0 [-]	0.49242	0.06337
β_1 [-]	0.96725	0.01409
r^2 [-]	0.8206	–
RMSE [m·s^{-1}]	1.006	–

As result the model equation is:

$$v_t = \beta_0 + \beta_1 v_r = 0.4924 + 0.9673\ v_r \quad (13)$$

Predicted long-term time series of wind speed (v_t) was calculated by equation (13). As the independent variable, wind speed in time range 1979–2016 ($n = 55,520$; from reanalysis) was used. On the Figure 2 predicted wind speed versus time is shown. Long-term mean wind speed (v_{lt1}) is equal 4.46 m·s^{-1}.

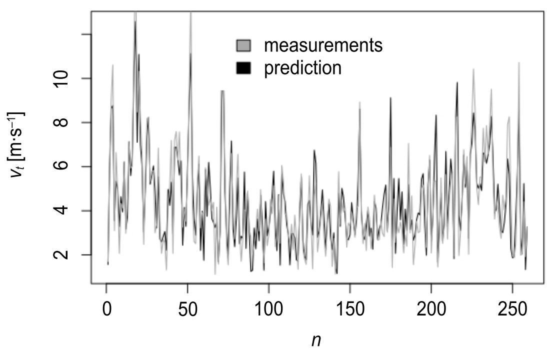

FIGURE 1. Comparison of real and predicted wind speed (12 m a.g.l.) versus time for simple model (validation data)

The next stage, for the training dataset multiple linear regression was applied. From many algorithms [Larose 2008] backward stepwise linear regression was selected. At first time the model has the full set of independent variables: day (d), hour (h), wind speed (v_r), wind direction (γ), atmospheric pressure (p), temperature (t). For each step, one the most statistically insignificant variable is removed from equation of model, based on t-test and maximum p-value [Koronacki et al. 2006]. Results indicated that model is statistically significant (F-statistic value: 783.2; $p < 2.2 \cdot 10^{-16}$), at significance level α = 0.005. Values derived from the final step are in Table 2.

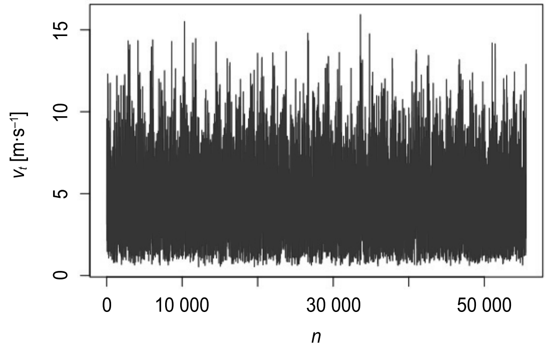

FIGURE 2. Predicted wind speed (12 m a.g.l.) at Żelazna versus time (time range: 1979–2016) for simple model

After four steps of calculations, variables: day (d), hour (h), wind direction (γ), was removed from the model (Table 2). Hence, the equation of multiple linear model is:

$$v_t = \beta_0 + \beta_3 v_r + \beta_5 p + \beta_6 t \tag{14}$$

or

$$v_t = 11.8656 + 0.9365\, v_r - 0.01106\, p - 0.02295\, t \tag{15}$$

After the equation (15) was applied to validation data, *RMSE* was calculated. This value is very close to *RMSE* value of simple linear model, as stated in Table 1. On the Figure 3 the comparison of real and predicted wind speed, based on validation dataset, was presented. The accuracy of prediction is satisfactory. Multiple linear model was also used for prediction of wind speed (v_r) at target site, on basis of 1979–2016 reanalysis data. Long-term mean value (v_{lt2}) is equal 4.48 m·s^{-1}. The value is very close for that derived from simple model. In the report [Eko-Efekt 2004] is stated that annual mean wind speed at this area is between 3 and 5 m·s^{-1}. Both values calculated by two statistical models are found to be in

TABLE 2. Parameters of final step of multiple linear regression model

Variable	Parameter	Value	*SE*	t-test value	*p*
×	β_0 [-]	11.865632	3.195121	3.714	0.000215
v_r	β_3 [-]	0.936447	0.014767	63.414	$<2 \cdot 10^{-16}$
p	β_5 [-]	–0.011061	0.003179	–3.479	0.000524
t	β_6 [-]	–0.022952	0.003779	–6.074	$1.76 \cdot 10^{-9}$
R^2_{adj} [-]		0.819	×		
RMSE [m·s^{-1}]		1.000	×		

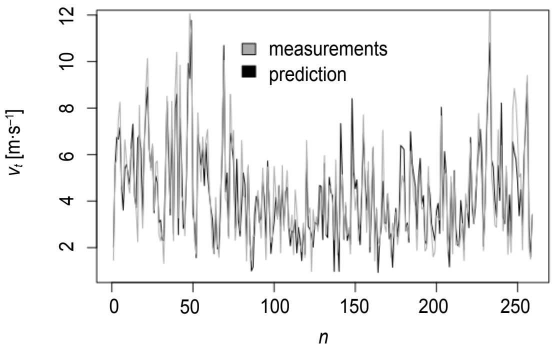

FIGURE 3. Comparison of real and predicted wind speed (12 m a.g.l.) versus time for multiple linear model (validation data)

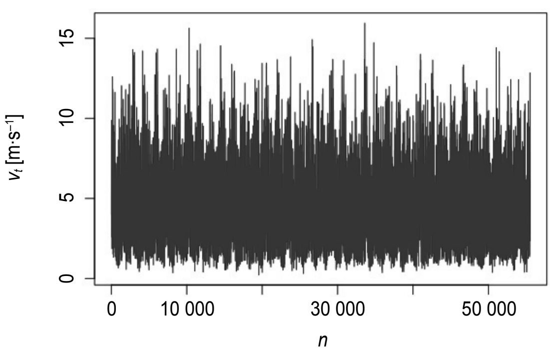

FIGURE 4. Predicted wind speed (12 m a.g.l.) at Żelazna versus time (time range: 1979–2016) for multiple linear model

this range. At Figure 4 predicted wind speed versus time for this model is presented. The nature of wind speed changes is similar for that which was obtained from simple model (Fig. 2).

Both wind speed time series from simple and multiple regression MCP methods are used to calculation of Weibull distribution function. At Figure 5, relevant diagrams are presented as well as histograms and k and c parameters. Histograms as well as curves are very similar. The distinction between values of k and c coefficients obtained by different regression models is very small.

CONCLUSIONS

In this paper the application of measure–correlate–predict method was presented. On the basis of the obtained results, it was found that linear regression models are adequate for

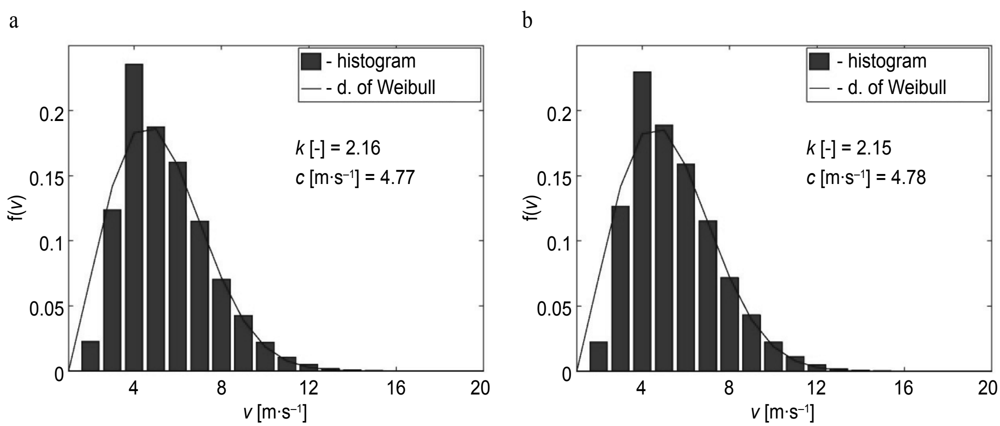

FIGURE 5. Weibull probability distribution (12 m a.g.l.) for simple (a) and multiple linear model (b)

calculations of long-time wind speed time series (12 m a.g.l.) for the given site. The relatively high values of coefficients of determination mean that reanalysis data are sufficient as the input data of regression models. There is no need of use the real long-term measurements from reference meteorological station. The obtained equations, by using of the least squares fit methods, allow for obtaining practically the same values of long-term mean wind speeds and parameters of Weibull wind speed distribution at target site. Hence, simple regression model with only one independent variable, i.e. wind speed from reanalysis, is better, just because of its simplicity. The calculated, long-time wind speed time series at Łódź Hills can be used in future for wind energy resources prediction at this place. If the predicted resources will be at least sufficient, there will be a premise of wind farm construction. One must be careful for every generalization of obtained models for other, even close, target sites, because of significant influence of local terrain and vegetation for wind resources.

REFERENCES

BAKOŃ T., KORUPCZYŃSKI R. 2017: Wind resource assessment for the potential wind farm location at field of Warsaw University of Life Sciences. WINERCOST'17: International Conference on Wind Energy Harvesting, 20–21 April 2017 University of Coimbra, Portugal: 116–120.

BHATTACHARYA P., BHATTACHARJEE R. 2010: A study on Weibull distribution for estimating the parameters. J. App. Quant. Meth. 5 (2): 234–241.

CARRILLO C., CIDRÁS J., DÍAZ-DORADO E., OBANDO-MONTAÑO A.F. 2014: An Approach to Determine the Weibull Parameters for Wind Energy Analysis: The Case of Galicia (Spain). Energies 7: 2676–2700.

CARTA J.A., VELÁZQUEZ S., CABRERA P. 2013: A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site. Ren. Sus. Energ. Rev. 27: 362–400.

DEE D.P., UPPALA S.M., SIMMONS A.J., BERRISFORD P., POLI P., KOBAYASHI S., ANDRAE U., BALMASEDA M.A., BALSAMO G., BAUER P., BECHTOLD P., BELJAARS A.C.M., Van De BERG L., BIDLOT J., BORMANN N., DELSOL C., DRAGANI R., FUENTES M., GEER A.J., HAIMBERGER L., HEALY S.B., HERSBACH H., HÓLM E.V., ISAKSEN L., KÅLLBERG P., KÖHLER M., MATRICARDI M., McNALLY A.P., MONGE-SANZ B.M., MORCRETTE J-J., PARK B-K., PEUBEY C., De ROSNAY P., TAVOLATO C., THÉPAUT J-N., VITART F. 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quart. J. Royal Meteorol. Soc. 137: 553–597.

DÍAZ S., CARTA J.A., MATÍAS J.M. 2017: Comparison of several measure-correlate-predict models using support vector regression techniques to estimate wind power densities. A case study. Energ. Conver. Manag. 140: 334–354.

DINLER A. 2013: A new low-correlation MCP (measure-correlate-predict) method for wind energy forecasting. Energy 63: 152–160.

Directive 2009/28/EC of the European Parliament and the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC. L 140/16.

Eko-Efekt 2004: Program Ochrony Środowiska dla Gminy Skierniewice. Eko-Efekt Sp. z.o.o., Warszawa.

Eurostat 2015: Environmental Data Centre on Natural Resources, Natural Resources, Energy Resources, Wind Energy. Retrieved from http://ec.europa.eu/eurostat/web/environmental-data-centre-on-natural-resources/natural-resources/energy-resources/wind-energy.

Forbes 2017: Renewable Energy Hits Global Tipping Point. Retrieved from https://www.forbes.com/sites/morganstanley/2017/08/02/renewable-energy-hits-global-tipping-point/#5d218dd61f85.

KORONACKI J., MIELNICZUK J. 2006: Statystyka dla studentów kierunków technicznych i przyrodniczych. WNT, Warszawa.

LAROSE D.T. 2008: Metody i modele eksploracji danych. PWN, Warszawa.

LAROSE D.T., LAROSE C. 2015: Data Mining and Predictive Analytics. John Wiley & Sons, Hoboken, NJ.

MEASNET Procedure 2016: Evaluation of site-specific Wind Conditions. Version 2.

MESSAC A., CHOWDHURY S., ZHANG J. 2012: Characterizing and Mitigating the Wind Resource-based Uncertainty in Farm Performance. J. Turbul. 13 (13): 1–26.

PEREA A.R., AMEZCUA J., PROBST O. 2011: Validation of three new measure-correlate-predict models for the long-term prospection of the wind resource. J. Ren. Sus. Energ. 3: 023105. https://doi.org/10.1063/1.3574447.

PESTA R. 2009: Analiza opłacalności budowy farmy wiatrowej o mocy 40 MW. Rynek Energii 1: 22–25.

R Development Core Team 2008: A language and environment for statistical computing. R Fundation for Statistical Computing, Vienna, Austria.

ROGERS A.L., ROGERS J.W., MANWELL J.F. 2005: Comparison of the performance of four measure–correlate–predict algorithms. J. Wind Eng. Ind. Aerodyn. 93 (3): 243–264.

THØGERSEN M.L., MOTTA M., SØRENSEN T., NIELSEN P. 2007: Measure-correlate-predict methods: Case studies and software implementation. European Wind Energy Conference & Exhibition, European Wind Energy Association.

VELÁZQUEZ S., CARTA J.A., MATIAS J. 2011: Comparison between ANNs and linear MCP algorithms in the long-term estimation of the cost per kWh produced by a wind turbine at a candidate site: A case study in the Canary Islands. App. Energ. 88: 3869–3881.

WEEKES S.M., TOMLIN A.S., VOSPER S.B., SKEA A.K., GALLANI M.L., STANDEN J.J. 2015: Long-term wind resource assessment for small and medium-scale turbines using operational forecast data and measure–correlate–predict. Ren. Energ. 81: 760–769.

ZHANG J., CHOWDHURY S., MESSAC A., HODGE B. 2013: Assessing Long-Term Wind Conditions by Combining Different Measure-Correlate-Predict Algorithms, ASME. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference 3A: 39th Design Automation Conference.

Streszczenie: *Zastosowanie metody measure––correlate–predict do predykcji prędkości wiatru na Wzniesieniach Łódzkich.* W artykule podano przykład zastosowania metody measure–correlate–predict (MCP). Badania przeprowadzono na terenie Rolniczego Zakładu Doświadczalnego SGGW w Żelaznej (powiat skierniewicki). Dwa zestawy danych: rzeczywiste pomiary meteorologiczne ze stacji pomiarowej oraz dane pozyskane z tzw. reanalizy meteorologicznej ERA-Interim zostały użyte do budowy modelu statystycznego. Model został wykorzystany do długoterminowej predykcji prędkości wiatru (1979–2016). Uzyskane wyniki są przydatne do oceny zasobów energii wiatru na Wzniesieniach Łódzkich.

MS received November 2017

Author's address:
Rafał Korupczyński
Wydział Inżynierii Produkcji SGGW
Katedra Podstaw Inżynierii
02-787 Warszawa, ul. Nowoursynowska 164
Poland
e-mail: rafal_korupczynski@sggw.pl