Energy Conversion and Management 140 (2017) 334–354

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Comparison of several measure-correlate-predict models using support vector regression techniques to estimate wind power densities. A case study

Santiago Díaz [a], José A. Carta [a,*], José M. Matías [b]

[a] Department of Mechanical Engineering, University of Las Palmas de Gran Canaria, Campus de Tafira s/n, 35017 Las Palmas de Gran Canaria, Canary Islands, Spain
[b] Department of Statistics, University of Vigo, Lagoas Marcosende, 36200 Vigo, Spain

ARTICLE INFO

Article history:
Received 16 November 2016
Received in revised form 22 February 2017
Accepted 23 February 2017

Keywords:
Wind power density
Measure-correlate-predict
Support vector regression
Feature selection
Statistical significance

ABSTRACT

The long-term annual mean wind power density (WPD) is an important indicator of wind as a power source which is usually included in regional wind resource maps as useful prior information to identify potentially attractive sites for the installation of wind projects. In this paper, a comparison is made of eight proposed Measure-Correlate-Predict (MCP) models to estimate the WPDs at a target site. Seven of these models use the Support Vector Regression (SVR) and the eighth the Multiple Linear Regression (MLR) technique, which serves as a basis to compare the performance of the other models. In addition, a wrapper technique with 10-fold cross-validation has been used to select the optimal set of input features for the SVR and MLR models. Some of the eight models were trained to directly estimate the mean hourly WPDs at a target site. Others, however, were firstly trained to estimate the parameters on which the WPD depends (i.e. wind speed and air density) and then, using these parameters, the target site mean hourly WPDs. The explanatory features considered are different combinations of the mean hourly wind speeds, wind directions and air densities recorded in 2014 at ten weather stations in the Canary Archipelago (Spain).

The conclusions that can be drawn from the study undertaken include the argument that the most accurate method for the long-term estimation of WPDs requires the execution of a specially trained model which considers the variability of the wind speeds of the reference stations, as well as of the wind directions and air densities, and in addition the functional manner in which these variables participate in the proposed MCP models. It is also concluded that it is important to consider the annual variation of air density even in regions at sea level. It is further concluded that, of the eight MCP models under comparison, the one that predicts the WPDs based on two sub-models (which estimate the wind speeds and air densities in an unlinked manner) always provides the best MAE (Mean Absolute Error), MARE (Mean Absolute Relative Error) and R^2 (Coefficient of determination) metrics, with the differences being statistically significant (5% significance) for most of the cases assessed. Additionally, the regulatory capacity of the SVR technique was sufficient to manage most of the overfitting problems, and hence the contribution of the wrapper method was not relevant in our study.

1. Introduction

In this section, a background is firstly provided to the problem related to long-term estimation of Wind Power Densities (WPDs) when Measure-Correlate-Predict (MCP) methods are used which are based on information provided by multiple reference weather stations (WSs). Subsequently a description is given of the aim and original contribution of this paper.

1.1. Background

In the scientific literature, an extensive collection of MCP methods [1] have been proposed for hindcasting of the long-term wind characteristics at sites for which only measurements recorded over a short time period are available.

The most commonly proposed and used methods to date in the wind industry have been based on information obtained from a single reference station. However, in the scientific literature concerned with renewable energies a growing number of proposals can be seen for methods which are based on the use of several ref-

* Corresponding author.
E-mail address: jcarta@dim.ulpgc.es (J.A. Carta).

http://dx.doi.org/10.1016/j.enconman.2017.02.064
0196-8904/

Nomenclature

AEMET Spanish initial: state meteorological agency of the Spanish Government
aa_i, $aw1_i$, $aw2_i$ parameters that define the second molar virial coefficients of the mixture. Eqs. (13) and (14). Table 3
$\mathrm{aaa_i}$, $\mathrm{aaw_i}$ parameters that define the third molar virial coefficients of the mixture. Eqs. (16) and (17). Table 4
ANN Artificial Neural Network
$\mathrm{aww_i}$ parameters that define the third molar virial coefficients of the mixture. Eq. (18) and Table 4
b bias parameter in support vector regression. Eqs. (28), (34) and (36).
B_{aa} second virial air–air coefficient ($m^3\ mol^{-1}$). Eqs. (11) and (13)
B_{aw} second cross virial coefficient ($m^3\ mol^{-1}$). Eqs. (11) and (14)
BH Benjamini and Hochberg step-up procedure [56]
B_m second molar virial coefficient of the mixture. Eq. (11)
B_{ww} second virial water–water coefficient ($m^3\ mol^{-1}$). Eqs. (11) and (15)
C constant that determines the trade-off between the flatness of $f(x)$ and the amount up to which deviations larger than ϵ are tolerated in Support Vector Machine. Eqs. (29), (30) and (33)
C_{aaa} third virial air coefficient ($m^6\ mol^{-2}$). Eqs. (12) and (16)
C_{aaw} third air–air–water virial coefficient ($m^6\ mol^{-2}$). Eqs. (12) and (17)
C_{aww} third air–water–water virial coefficient ($m^6\ mol^{-2}$). Eqs. (12) and (18)
C_{www} third virial water coefficient ($m^6\ mol^{-2}$). Eqs. (12) and (19)
CIPM International Committee for Weight and Measures
C_m third molar viral coefficient of the mixture. Eq. (12)
D variable that represents the wind direction in degrees
$\hat{e}_i$ variable that represents the estimated values. Eqs. (39)–(42)
$E[\bullet]$ population mean of a random variable. Eqs. (1) and (3)
$E\ [\hat{\bullet}]$ mean of an estimated variable. Eqs. (4) and (6)
$\widehat{E[\bullet]}$ estimated mean of a variable. Eq. (5)
ef enhancement factor (non-dimensional). Eqs. (8) and (20)
$f_A(B_1,\ldots,B_p)$ regression method that uses the features $B_1,\ldots,B_p$ to obtain a forecast of variable A
FDR False Discovery Rate
$f(\boldsymbol{x})$ regression function
$f_\rho(\rho)$ air density probability density function. Eq. (3)
$f_v(v)$ wind speed probability density function. Eq. (3)
$f_{\rho v}(\rho, v)$ joint probability density function of ρ and v. Eq. (44)
$f_{w,b}(x)$ regression function in Support Vector Machine that depends on w and b. Eq. (34)
$f_{WPD}[\rho, v]$ wind power density probability density function. Eq. (44)
$f_\beta(x)$ regression function in Support Vector Machine that depends on β. Eq. (25)
$k(x_i, x_j)$ kernel function in Support Vector Machine. Eqs. (32), (34), (35) and (36)
g_i parameters of saturation vapour pressure. Eq. (9). Table 2
H relative air humidity (%)
H_0 null hypothesis. Eq. (43)
H_1 alternative hypothesis. Eq. (43)
$\mathrm{ic}_1,\ldots,\mathrm{ic}_7$ parameters of isothermal compressibility. Eq. (21). Table 2
ISA International Standard Atmosphere
ITC Spanish initials: Technological Institute of the Canary Islands
$\ell_{p,\epsilon}$ loss functions in Support Vector Machines. Eqs. (30) and (31)
M_a molar mass of dry air ($kg\ mol^{-1}$). Eq. (7)
MAE Mean Absolute Error ($W\ m^{-2}$). Eq. (40)
MARE Mean Absolute Relative Error. Eq. (41)
MCP Measure–Correlate–Predict
ML Machine Learning
MLR Multiple Linear Regression. Eq. (24)
M_v molar mass of water vapour ($kg\ mol^{-1}$). Eq. (7)
M1,..., M8 MCP models to estimate the Wind Power Density that are compared in this study
$\mathrm{mv_i}$ parameters of molar volume of saturated liquid water. Eq. (22). Table 2
n number of data. Eqs. (2)–(6), (24), (27), (29), (30), (33), (39)–(42)
o_i variable that represents the observed values. Eqs. (39–42)
$\bar{o}$ variable that represents the mean of observed values. Eq.(42)
S support vectors in Support Vector Machine
p barometric pressure (hPa). Eqs. (7), (8), (10) and (20)
p loss function parameter. Eq. (31)
p_{sv} saturation vapour pressure (Pa). Eqs. (8), (9) and (20)
p-value estimated probability of rejecting the null hypothesis (H_0) when that null hypothesis is true
R gas constant of dry air ($J\ K^{-1}\ mol^{-1}$)
R^2 coefficient of determination (%). Eq. (42)
$\mathbb{R}^d, \mathbb{R}^h$ feature space where "h" is usually bigger than "d"
RFE Recursive Feature Selection
RMSE Root Mean Square Error. Eq. (39)
SVM Support Vector Machine
SVR Support Vector Regression
t_a ambient temperature in degrees celsius (°C). Eqs. (7), (9), (10), (13–20) and (22).
v variable which represents the wind speed of weather stations (ms^{-1}). Eqs. (3) and (44)
v_i variable which represents the mean hourly wind speed (ms^{-1}). Eqs. (4), (5), (6) and (38)
$\bar{v}$ variable which represents the mean wind speed (ms^{-1}). Eq. (10)
v_m monthly mean wind speed (ms^{-1}). Eq. (6)
w characteristic parameter in Support Vector Regression. Eqs. (28–30), (33), (34) and (36)
WPD Wind Power Density ($W\ m^{-2}$). Eq. (38)
WPD_i variable that represents the mean hourly Wind Power Density ($W\ m^{-2}$). Eq. (38)
$\overline{WPD}$ Mean Wind Power Density ($W\ m^{-2}$). Eq. (2)
WS-1...WS-10. weather stations
$\mathrm{ww_i}$ parameters that define the second molar virial coefficient of the mixture. Eq. (15). Table 3
$\mathrm{www_i}$ parameters that define the third molar virial coefficient of the mixture. Eq. (19). Table 4
Y set of target sites
y_i vector which contains the observed wind power values at the target site. Eqs. (28–31) and (33)
X set of references sites
$x_i,\ x_j$ vector which contains the observed wind power values at the references site. Eqs. (24), (28–36)
x_{O2} molar fraction of oxygen in air. Eq. (23)
x_{N2} molar fraction of nitrogen in air. Eq. (23)
X_v molar fraction of water vapour in air
Z compressibility factor of air (non-dimensional). Eq. (10)