ARTICLE IN PRESS

Applied Energy xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Applied Energy

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

Performance assessment of five MCP models proposed for the estimation of long-term wind turbine power outputs at a target site using three machine learning techniques

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

HIGHLIGHTS

- Five models based on MCP methods to estimate WTPOs at a target site are assessed.
- Models are proposed which take into account air density variability.
- Three machine learning techniques implemented in the models are analysed.
- The models use wind turbines with blade pitch control and stall-regulated wind turbines.
- Statistical hypothesis tests are used to compare the ML techniques in the best model and to compare some of the models.

ARTICLE INFO

Keywords:
Support vector machine
Artificial neural network
Random forest
Wind turbine power curve
Wind turbine power output
Air density

ABSTRACT

Various models based on measure-correlate-predict (MCP) methods have been used to estimate the long-term wind turbine power output (WTPO) at target sites for which only short-term meteorological data are available. The MCP models used to date share the postulate that the influence of air density variation is of little importance, assume the standard value of 1.225 kg m^{-3} and only consider wind turbines (WTs) with blade pitch control.

A performance assessment is undertaken in this paper of the models used to date and of newly proposed models. Our models incorporate air density in the MCP model as an additional covariable in long-term WTPO estimation and consider both WTs with blade pitch control and stall-regulated WTs. The advantages of including this covariable are assessed using different functional forms and different machine learning algorithms for their implementation (Artificial Neural Network, Support Vector Machine for regression and Random Forest).

The models and the regression techniques used in them were applied to the mean hourly wind speeds and directions and air densities recorded in 2014 at ten weather stations in the Canary Archipelago (Spain). Several conclusions were drawn from the results, including most notably: (a) to clearly show the notable effect of air density variability when estimating WTPOs, it is important to consider the functional ways in which the features air density and wind speed and direction intervene, (b) of the five MCP models under comparison, the one that separately estimates wind speeds and air densities to later predict the WTPOs always provided the best mean absolute error, mean absolute relative error and coefficient of determination metrics, independently of the target station and type of WT under consideration, (c) the models which used Support Vector Machines (SVMs) for regression or random forests (RFs) always provided better metrics than those that used artificial neural networks, with the differences being statistically significant (5% significance) for most of the cases assessed, (d) no statistically significant differences were found between the SVM- and RF-based models.

1. Introduction

When making a decision as to whether a particular wind turbine (WT) should be installed at a target site it is of interest to know the wind turbine power outputs (WTPOs). The WTPOs are estimated using the power curve of the WT and the characteristics of the wind regime and air density at the site where the WT is to be installed (or not). The estimated long-term gross annual mean WTPO at a target site allows a

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

https://doi.org/10.1016/j.apenergy.2017.11.007
Received 8 August 2017; Received in revised form 3 October 2017; Accepted 2 November 2017

Nomenclature

AEMET Spanish initials: State Meteorological Agency of the Ministry of the Environment and Rural and Marine Environs of the Spanish Government
ANN Artificial Neural Network
b Bias parameter in Support Vector Regression. Eqs. (3) and (4)
BH Benjamini and Hochberg step-up procedure [61]
BMA Bayesian Model Averaging
$\boldsymbol{C}$ 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 Machines. Eqs. (3) and (4)
c_i weights which analyse the error functions in the case of model M5. Eqs. (14) and (15)
C(v(z)) electrical power coefficient of a wind turbine in function of the wind speed v(z) measured at the height z of the hub of said wind turbine
d number of neurons in the input layer of an Artificial Neural Network. Eq. (2)
D variable that represents the wind direction in degrees
e_i variable that represents the estimated values. Eqs. (14), (22), (23) and (24)
Epochs maximum number of iterations permitted in the Artificial Neural Network algorithm as "Early Stopping" criterion
FEDER Spanish initials: European Regional Development Fund
$f(\mathbf{x})$ regression function
$f_k(x)$ Random Forest function applied over L regression trees. Eq. (9)
$f_v(v,D,\rho_0)$ function which estimates the target site wind speed using the wind speed and direction data of the reference stations and standard air density (1.225 kg m^{-3}). Eq. (12)
$f_v(v,D,\rho)$ function which estimates the target site wind speed using the wind speed and direction data and mean air densities of the reference stations. Eqs. (10), (11) and (13)
$f_\rho(\rho,D)$ function which estimates the target site air densities using the air densities and wind directions of the reference stations (kg m^{-3}). Eq. (16)
f_{RF} regression calculated with the mean of the results obtained in the L estimated regression trees. Eq. (19)
GIS Geographical Information System [45]
$k(x_i,x_j)$ Kernel function in Support Vector Machine. Eqs. (4) and (5)
$h_j(\cdot)$, $g(\cdot)$ Activation function in the Artificial Neural Network algorithm. Eq. (2)
H_0 null hypothesis. Eq. (25)
H_1 alternative hypothesis. Eq. (25)
IEC International Electrotechnical Commission
ISO International Standard Organization
ITC Spanish initials: Technological Institute of the Canary Islands
$I(\cdot)$ function pertaining to the Random Forest method which returns a value equal to 1 if the argument is true and 0 otherwise. Eq. (6)
L Pseudo-training set from the sample **T** comprised of n data with which L regression trees are separately fitted. Eq. (19)
LTS long-term support
m neurons in the hidden layer of the Artificial Neural Network. Eq. (2)
M number of trees in the Random Forest technique. Eq. (6)
MAC space of cooperation formed by the outermost regions of Madeira, Azores and Canary Islands
MAE Mean Absolute Error (kW). Eq. (22)
MARE Mean Absolute Relative Error. Eq. (23)
MCP Measure-Correlate-Predict
MLP Multilayer perceptron
ML Machine Learning
Mtry number of variables randomly organised as candidates in each division made in the nodes of the regression trees
M1, … , M5 MCP models to estimate the WTPOs and which are assessed in this paper
n Number of data. Eqs. (10), (11), (12), (13) and (16)
n_i total number of observations of the target variable Y pertaining to the leaf node, R_j, in Random Forest. Eq. (7)
Neurons number of neurons of the hidden layer in the Artificial Neural Network technique
Nodesize minimum size of the terminal nodes in the Random Forest technique
o_i variable that represents the observed values. Eqs. (14), (22), (23) and (24)
$\bar{o}$ variable that represents the mean of observed values. Eq. (24)
S Support vectors in Support Vector Machine
PDF Probability Density Function
Pr Rated power. Eq. (15)
p-value the estimated probability of rejecting the null hypothesis (H_0) when that null hypothesis is true
$P(v(z), \rho_0)$ power curve of a wind turbine which provides, for a specific air density, (ρ_0), the relationship between the wind speed, v(z), at hub height, z, and the electrical power generated
R gas constant of dry air (J K^{-1}mol^{-1}). Eq. (18)
R_j nodes which form each one of the tress in the calculation of the Random Forest technique. Eqs. (6), (7) and (8)
$\mathbb{R}^d$,$\mathbb{R}^h$ feature space where "h" is usually bigger than "d"
RF Random Forest
RFE Recursive Feature Selection
RMSE Root Mean Square Error. Eq. (14)
RSS Residual Sum of Squares. Eq. (8)
R^2 Coefficient of determination (%). Eq. (24)
SIOSE Spanish initials: Information System of Land Occupation in Spain
SVM Support Vector Machine for regression
SVR Support Vector Regression
$\boldsymbol{T}$ data sample comprised of variables $\mathbf{x_i}$ and y_i
$T(z_r)$ reference height air temperature. Eq. (18)
V variable which represents the weather station wind speed (ms^{-1})
$\bar{v}$ variable which represents the mean wind speed (m s^{-1}). Eqs. (10), (11), (12) and (13)
v_i variable which represents the mean hourly wind speed (ms^{-1}). Eqs. (10), (11), (12) and (13)
$v(z)$ variable which represents the mean hourly wind speed at hub height (m s^{-1}). Eqs. (19), (20) and (21)
w_{ij} weights in Artificial Neural Network technique. Eq. (2)
w_j weights of the connections of the m neurons of the hidden layer with bias β_j. Eq. (2)
WS Weather Station
WS-1,…,WS-10. codes assigned to the weather stations used in this work
WT-1 Wind Turbine with active power control considered in this study
WT-2 stall-regulated wind turbine considered in this study
$\boldsymbol{Y}$ set of targets sites
y_i vector which contains the predicted values at the target site. Eq. (1)
$\boldsymbol{X}$ set of references sites
$\boldsymbol{x_i}$ vector which contains each observation of the variables **X** at the reference sites