Effect of Outliers in Wind Speed Assessment

Takvor H. Soukissian, Flora E. Karathanasi, Evangelos G. Voukouvalas
Hellenic Centre for Marine Research, Institute of Oceanography
Anavysos, Greece

ABSTRACT

In offshore wind energy assessment, the probability of a stochastic model failure for the wind speed should always be examined, since wind data are prone to the occurrence of outliers. Their presence suggests the use of more efficient and less sensitive methods. In this work, robust regression methods are introduced, described and applied in order to model the linear relationship between wind speeds obtained from two different data sources in the Aegean Sea. Moreover, an evaluation procedure, which is very frequently used in the measure–correlate–predict family of methods, is also applied for comparing the efficiency of the robust and the most commonly adopted ordinary least squares estimators for different time frames. The obtained results clearly suggest that outliers in wind data samples should not be ignored or diminished.

KEY WORDS: wind speed; wind energy; robust methods; outliers; calibration; Aegean Sea.

ABBREVIATIONS

BDP: breakdown point
CV: coefficient of variation
LMS: least median of squares
LTS: least trimmed of squares
MAE: mean absolute error
MCP: measure–correlate–predict
MLM: maximum likelihood method
NWP: numerical weather prediction
OLS: ordinary least squares
OWF: offshore wind farm
RE: relative error
RMSE: root mean square error
SI: scatter index

INTRODUCTION

For the appropriate site selection of an offshore wind farm (OWF), it is necessary to quantify the wind speed fields in the area of interest with the highest possible accuracy. The most frequently used wind data sources are in situ measurements from oceanographic buoys, coastal or offshore meteorological stations (masts) and gridded data obtained either from satellite sensors or from numerical weather prediction (NWP) models. Nowadays, more quantitative and accurate remote sensing measurement techniques for offshore wind energy applications include SoDAR and LiDAR.

For a preliminary assessment of the wind climate in an area, installation of offshore meteorological masts is a very costly procedure; therefore, wind data from buoys, satellites, coastal meteorological stations or NWP models are frequently used. However, each data source is characterized by its shortcomings and limitations. In order to obtain concurrent long-term wind data sets and model effectively the wind climate in an offshore area, the combined assessment of wind data obtained from various sources is preferred. In this respect, a validation and calibration procedure should be implemented through linear regression analysis, which is the primary theoretical background for such applications (e.g. Hwang et al., 1998; Caires, Sterl, 2003; Soukissian, Kechris, 2007). Among the numerous possible regression techniques, the most common is ordinary least squares (OLS) because of its ease of computation and statistical efficiency. This method belongs to the family of classical regression parameter estimation methods along with other optimal statistical methods such as maximum likelihood method (MLM), since they rely on the assumption that normality holds exactly. However, as is stated in Maronna, Martin (2006): “theoretical and computational convenience does not always deliver an adequate tool for the practice of statistics and data analysis”.

Moreover, in many cases, it is noticed that some observations do not fit the general pattern of the data by appearing “peculiar” relatively to the bulk of the data. Such cases are called outliers. If these atypical observations are not detected, examined and finally removed from a data sample, then the linear regression model can result in misleading interpretations. However, in practice, outliers may be unnoticed or even ignored in a regression analysis or other mathematical procedures (e.g. principal component analysis). In this respect, robust methods have been developed to produce reliable regression estimators that are less sensitive to the presence of outliers or model misspecifications while preventing efficiency loss. Some fundamental references for robust methods are: Tukey (1960), Huber (1973, 1981), Hampel (1971) and Rousseeuw, Leroy (1987).

Robust statistical techniques have been particularly implemented in wind data filtering, with promising results; see Sainz et al. (2009). Moreover, Pinson et al. (2007) introduced robust estimators in conditional parametric models with time-varying coefficient functions regarding wind power forecasting. The results of the application on real wind data showed that robust estimators will be beneficial for