Applied Energy 156 (2015) 321–330

Contents lists available at ScienceDirect

Applied Energy

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

Short-term wind speed prediction based on robust Kalman filtering: An experimental comparison

Carlos D. Zuluaga *, Mauricio A. Álvarez, Eduardo Giraldo

Department of Electrical Engineering, Faculty of Engineering, Universidad Tecnológica de Pereira, 660003, Colombia

HIGHLIGHTS

- Three robust Kalman filters are applied to one-step-ahead forecast of wind speed.
- We provide a detailed description of differences among robust Kalman filters.
- Two wind speed datasets with outliers are used to validate all methods.
- Results shown that proposed methods yield better performances than the Kalman filter.

ARTICLE INFO

Article history:
Received 16 January 2015
Received in revised form 26 June 2015
Accepted 15 July 2015
Available online 25 July 2015

Keywords:
Outliers
Robust Kalman filtering
Wind speed prediction
Wind power generation

ABSTRACT

The use of wind energy for power electric systems attempts to reduce the dependence on fuel-based energy. With the aim of generating electrical power based on wind energy, it becomes necessary to model and predict wind speed. Wind speed observations are packed with outliers what makes it difficult to propose accurate predictors. This paper presents an experimental comparison of three different methods for making a Kalman filter robust to outliers in the context of one-step-ahead wind speed prediction. Two wind speed databases were used to test the predictive performance of the algorithms. The performance for all the methods is measured in terms of skewness and kurtosis for the predicted signal. The algorithms discussed worked efficiently in a sequential approach, and outperformed the standard Kalman filter.

1. Introduction

The use of wind energy for power electric systems provides an alternative to reduce the dependence of fuel-based energy, stimulating clean production and reducing gas emissions.

Putting in action a power system based on wind energy implies modeling and predicting wind speed. Knowledge about the behavior of this variable is needed for estimating the amount of electrical power obtained from a wind energy conversion system. In turn, determining ahead the amount of electrical power would give electricity providers the flexibility to plan and manage the offer of this resource to end users.

Wind speed prediction or wind speed forecasting methods are important tools to guarantee the best operation and performance of wind turbines (turbines active control) or for real-time power dispatch, since hours-ahead or days-ahead forecasts can be used for power scheduling as well as for the optimization of maintenance operations.

Different algorithms for predicting wind speed have been proposed in the past. According to [1], these algorithms can broadly be divided into five groups, namely, persistence methods, physical-based methods, statistical-based methods, methods based on intelligent approaches, and hybrid methods. Persistence methods are based on the hypothesis that wind speed at time step $k+1$ will not change significantly with respect to wind speed at time k. This approach is used in very short term applications [2], and usually as a complement to numerical weather prediction. Physical-based methods use parametric models based on a physical description of the atmosphere, and are intended mainly for long-term forecasts [3]. Statistical-based methods include artificial neural networks [4–8], and time series models [9–11]. These techniques use data for wind speed forecasting, without resorting to a model of the real system [12]. Methods based on an intelligent approach [13–16], use a set of rules specified in advance by an expert. A typical example is a neuro-fuzzy system. These methods are employed for short-term wind speed forecasts. Finally, hybrid methods are combinations of different techniques such as physical

* Corresponding author.
E-mail addresses: cardazu@utp.edu.co (C.D. Zuluaga), malvarez@utp.edu.co (M.A. Álvarez), egiraldos@utp.edu.co (E. Giraldo).

http://dx.doi.org/10.1016/j.apenergy.2015.07.043
0306-2619/

and statistical approaches. Alternatively, they look for the integration of models performing short-term and medium-term predictions [17].

According to [18], wind speed data can be divided in six categories: valid, missing, constant, exceeding, irrational, and unnatural data. From these ones, irrational data (wind speed data that are not physically possible), and unnatural data (wind speed data that produce low wind output power or limited power production[1]) are considered as outliers or spurious data. These types of outliers may arise due to data entry problems, defective data collection instruments, or faulty instruments in the wind speed measuring system. If not taken seriously, outliers can lead to false conclusions about the real activity of the wind speed variable that the forecasting method is trying to predict.

Recent studies attempt to improve the forecasting accuracy of a prediction method by including a pre-processing stage for outlier detection. For example, in [19], the authors use an outlier detector based on robust regression by reweighted least squares. They compute a weighted residual error between the robust regression value and the forecasted value. If this error is below a certain threshold, the observation is considered as an outlier. In [20], the authors use a support vector regression (SVR) approach to preprocess abnormal data in the original wind speed series. After applying SVR to the original data, the authors compute an absolute error between the observed value and the corresponding regression value. If the error is below a certain threshold, the observation is considered as outlier and then it is replaced by the regression value, otherwise the observation is accepted.

In this paper, we will focus on the Kalman filter for short-term wind speed prediction, considering the potential presence of outliers in the data. Particularly, we will focus on one-step-ahead prediction of wind speed data. As a byproduct, and in contrast to [19,20], the models that we use merge the preprocessing step for outlier detection together with the prediction step. Kalman filtering has a long history of successful applications in which unknown variables, corrupted by Gaussian noise, can be approximately determined from observations. Applications include tracking [21–23], and short-term predictions [24]. For wind speed prediction, the Kalman filter has been used in different applications. The authors of [25–27] use the Kalman filter to improve the initial predictions performed by a limited-area numerical weather prediction model. Poncela et al. [28] present a recursive wind power forecasting system based on standard Kalman filtering, where the parameters of the filter are tuned using an Expectatio n–Maximization algorithm. In [29], the authors use an autoregressive integrated moving average (ARIMA) model as a state space model of the wind generation process, and the Kalman filter is used for wind speed prediction. The authors of [30] use support vector regression to develop a non-linear state space model (that describes the wind speed samples), which is then employed by the unscented Kalman filter for short-term wind speed prediction.

The applications of the Kalman filter mentioned above have used the standard formulation of the Kalman filter for which the wind speed observations are assumed to be contaminated by Gaussian noise. As we have seen, this is not always the case for wind speed observations. In real-time environments, such observations usually contain outliers, which disobey the Gaussian noise assumption. Hence, a proper forecasting method for wind speed should be robust to the presence of outliers.

In this paper, we present an experimental comparison of three one-step-ahead forecasting methods for wind speed data based on robust Kalman filtering. The methods can be divided in two groups according to whether they include all observations for the inference process or not. In the first group, we consider the weighted robust Kalman filter (wrKF) proposed by Ting et al. [31], and the robust statistics Kalman filter (rsKF) proposed by Cipra and Romera [32]. Both methods incorporate a mechanism that gives a weight to each observation. Each weight might be interpreted as the probability of that observation being an inlier. In the second group, we consider the thresholded Kalman filter (tKF) proposed by Ting et al. [31] and Schick and Mitter [33]. For each observation, a particular criteria is computed. If this value is below a certain threshold, the observation is discarded.

Experimental results obtained include the application of the different methods described above over a synthetic dataset and a real dataset. The real dataset was taken from the "Database of Wind Characteristics". [2]

This paper is organized as follows. Section 2 presents the material and methods that were used for performing robust wind speed estimation based on the Kalman filter. In this section, we also introduce the synthetic and the real data sets used for experimentation. In Section 3, we present and discuss the results obtained when applying the methods outlined before. Finally, conclusions and future work are described in Section 4.

2. Material and methods

The first subsection describes the synthetic dataset and the real dataset used for our experiments. The next subsection briefly exposes the standard Kalman filter. In the third subsection, we introduce the theory behind the different alternatives for robust Kalman filtering. Then we present some statistical tests and procedures for detecting outliers in a dataset, and two performance metrics used to evaluate the influence of the outliers in the predictions based on the different Kalman filters.

2.1. Datasets

For our first experiment, we generated a synthetic dataset with observations containing different percentages of contamination by outliers. Our purpose here was to test the robustness of the prediction for the different methods, when the "ground truth" data, this is, the clean data, is known. The clean data was generated using a wind model developed by the RISØDTU National Laboratory for Sustainable Energy at the Technical University of Denmark. The model computes the wind speed as the weighted average of the values of the wind speed at different locations in the blade of several wind turbines. It also takes into account the tower shadow and turbulence over the rotor of every turbine. We used a rotor diameter of 15 m [m]; an average wind speed of 12 m per second [m/s]; a turbulence level of 4%; and sampling time was set at 0.1 s [s]. For details see [34]. Gaussian noise, with mean zero and variance of $\sigma^2 = 0.001$, was added to the output data. Outliers were introduced randomly along the first-half of the time series. The number of outliers was equal to a 5% of the total amount of datapoints in the time series.

We then perform a second experiment using a real dataset. The real data set was taken from the "Database of Wind

[1] For unnatural data, it is necessary to distinguish between two cases. The first case refers to valid wind speed data that should result in power production, but that for some undefined reasons results in limited power production recorded. The second case is related to extreme or high values for wind speed data that lead to low or limited power production due to wind generator limitations. Moreover, when the wind speed is higher than a cut-off wind speed value, the wind turbines are shut down for safety reasons [18,19].

[2] The "Database of Wind Characteristics" is compiled and maintained by Kurt S. Hansen, Department of Mechanical Engineering, Technical University of Denmark (DTU). The dataset was collected from http://www.winddata.com.