

One Day Ahead Wind Speed Prediction - Midterm Report

Team Name: Wind Predictor

Bo Yang (by57@cornell.edu)

Data preview

The wind speed history data were collected at each airport. Given 5 years of wind speed (Jan 1 2011 - Dec 31 2015) at the Ithaca Airport (ITH), one day ahead wind speed prediction models will be developed.

In order to avoid over-fitting and under-fitting, three years data were reviewed firstly (2011-2013). Like other measurement database, there are missing data. Fortunately, shown from the Figure 1 (a), there are only tiny amount of missing data hours with regard to the whole year. Data of other years have similar histograms. Therefore, the missing wind speed hours could be deleted. The histogram after cleaning was shown in Figure 1 (b).

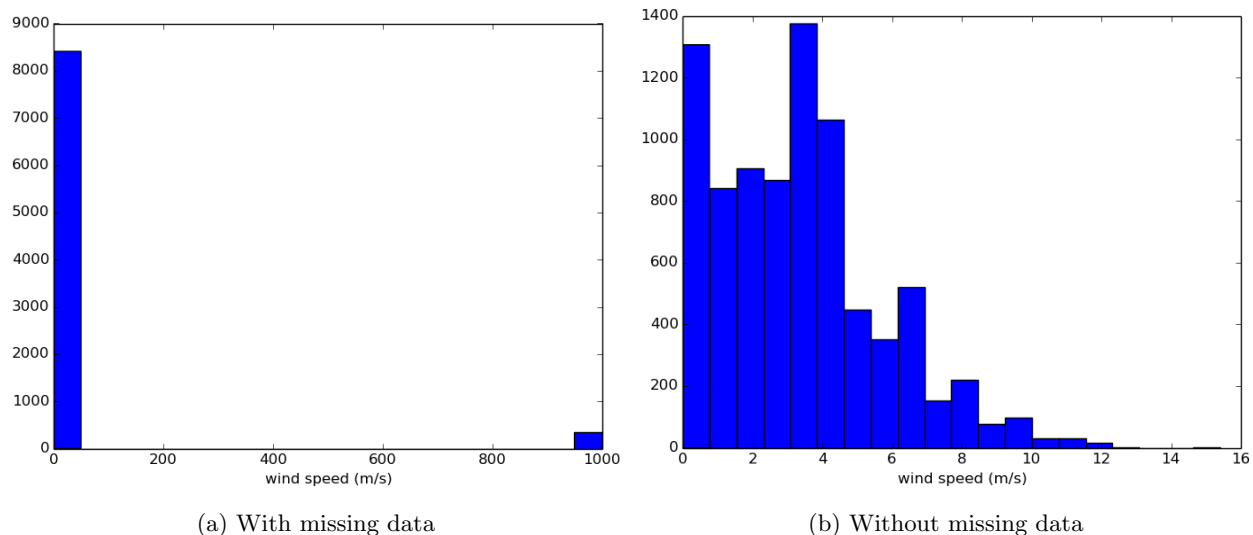


Figure 1: Histogram of wind speed of 2011 at ITH

Data Features

The hourly wind speed data are classic time series signals. And it is generally difficult to assume a certain pattern. Unfortunately, a simple Auto-Regression model using past few hours data cannot be directly applied for this one day ahead prediction because there are 24 hours wind speeds need be predicted. Wind speeds depend on many factors including temperature, pressure, cloud, and terrain, which means, for some specific locations, wind speeds variation may have some sort of common trend over specific periods of time. These periods might be much longer than one hour.

Figure 2 showed a continuous 72 hours wind speeds at ITH in 2011 and 2013. The general trends are similar from hour 150 to hour 180, illustrated in Figure 2 (a). In Figure 2 (b), the trends of the two years are also similar from hour 295 to hour 315. The wind speeds at the same hours of the past two years could be used as features for a linear model.

We may also look at the wind speeds at the same hour of continuous days. For example, the wind speeds at the same hour from Day 3 to Day 10 are similar, illustrated in Figure 3 (a). Also, the wind speed roughly oscillated around a certain value in Figure 3 (b). Then, the wind speeds at the same hours of several past

days could be applied as features for a linear model.

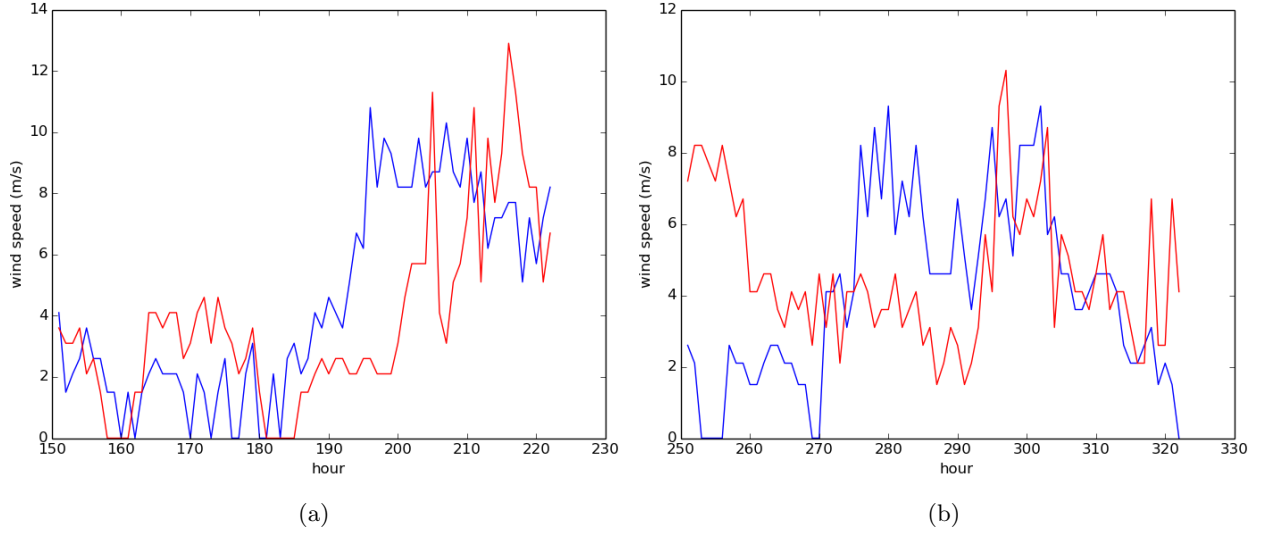


Figure 2: Wind speeds for 72 hours; Year 2011: blue lines; Year 2013: red lines

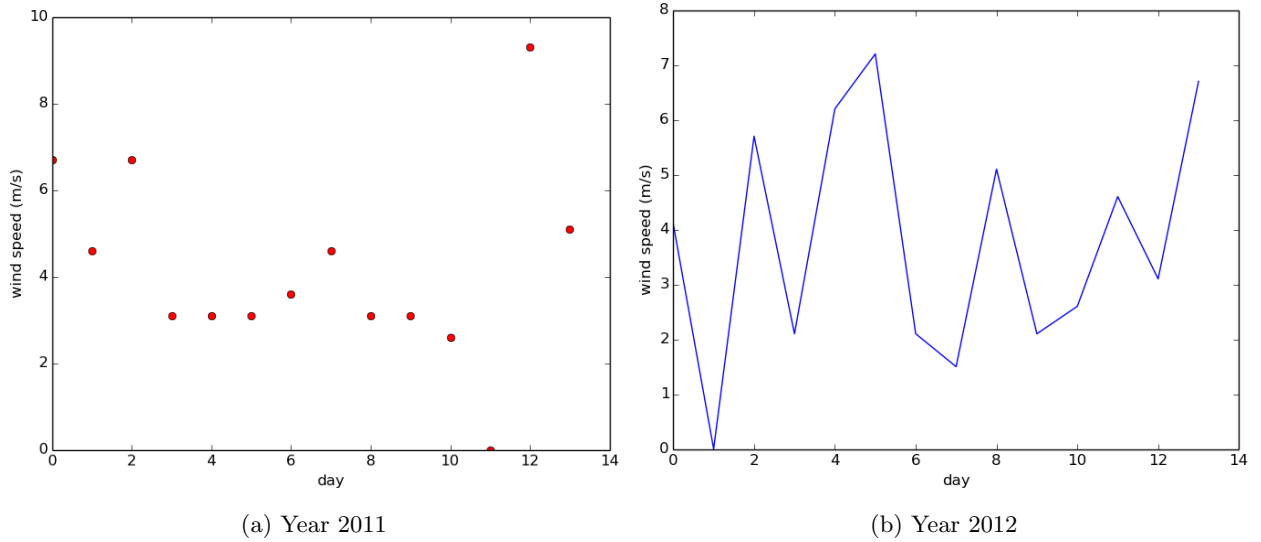


Figure 3: Wind speeds of the same hour on continuous days

Linear model: Two-Year-Based model

A simple Two-Year-Based model proposed by El-Fouly et al. (2006) uses data of n points from the current year and two previous years' series. The model is then used with the next n data points from the two previous year series to predict the next n points for the current wind speed series. The Two-Year-Based model can be represented by the following formula:

$$\hat{Y}_t(n+i) = w_1 X_{t-1}(n+i) + w_2 X_{t-2}(n+i) + w_3, (i = 1, 2, \dots, n)$$

X_{t-1}, X_{t-2} are the corresponding wind speed values of the two past years. We try to predict the one day ahead wind speed, then $n = 24$.

The training set contains hourly wind speeds of 259 days, which were randomly sampled from one year. The wind speeds of the same 259 days were taken from the other two years. The training data were around 70% of one year. The training data set is slightly conservative, but we don't want to over-fit the model. Then, 53 days data (15% of one year) were randomly sampled from the rest of data to form a validation set, which could be used to decide good model(s). The rest of 53 days data form the test set.

Preliminary results of the Two-Year-Based model

From the training set, 259 linear models were obtained by using the least-square method. The root-mean-square error (RMSE) was calculated for each model by using the validation set. The best model, $w = [0.212, 0.103719, 2.09972]$, has the least average error (around 2.3 m/s) for the whole validation set (53 days). The standard deviation of this model's error of the whole validation set is around 0.9 m/s. Figure 4 showed the model performance for two days from the test set. For one day ahead wind speed prediction problem, this model is not too bad, but not good enough, either.

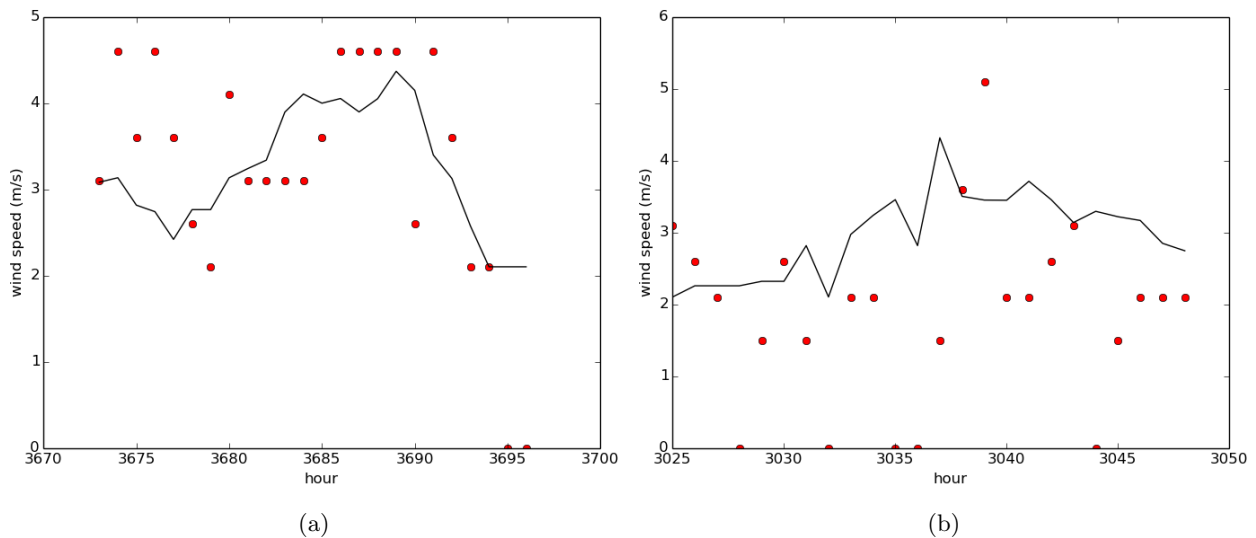


Figure 4: Two-Year-Based model performance

Ongoing work

1. The other linear model using data at same hours of previous days (no pages to show in this report);
2. Model performance with regularizers;
3. The Auto-Regression Integrated Moving Average (ARIMA) model development;
4. Model performance with one more important feature: temperature.

Since I am the only one working on this project, a rough plan is to finish one of the missions mentioned above in one week. Then summarize the whole project before the last week.

Reference

El-Fouly, T. H. M., El-Saadany, E. F., & Salama, M. M. A. (2006, June). One day ahead prediction of wind speed using annual trends. In 2006 IEEE Power Engineering Society General Meeting (pp. 7-pp). IEEE.