P767

An evaluation of different data mining methods for forecasting wind farm power

Gerry Wiener, J. M. Pearson, B. Lambi, W. Myers and K. Goodrich

National Center for Atmospheric Research, Boulder, CO

NCAR

In forecasting wind farm power output, it is important to obtain an accurate farm power output estimate based on given forecast winds. Generally, the manufacturer's turbine power curves are applied to obtain this estimate especially in cases when observed wind data at farms are not available. In this paper we will compare manufacturer power curve performance against the performance of a number of different data mining techniques including regression trees, KNN nearest neighbor and random forest regression based on using actual observed wind and power data. The modeling applied here differs from most traditional power curve applications since previous wind and power information are both utilized in order to forecast future power. Mean absolute error for the different techniques will be presented and the application of these techniques for forecasting winds and power will be discussed.

Overview

Introduction

- Wind power is now playing a significant role in world energy production
- Accurate 0-72 hour forecasts of wind/power are necessary
  - Supports effective power resource scheduling
    - Lessens greenhouse gas production
  - Supports effective power trading in the spot and day ahead energy markets
- Wind power forecasting can be thought of as having two phases:
  - Making an accurate wind speed forecast
  - Estimating wind power production using wind speed forecast and other pertinent information

NCAR Wind/Power System

- Utilizes forecast winds from multiple numerical weather models
  - WRF, GFS, NAM, RUC, GEM, MM5
- Applies statistical dynamic MOS technology DICast® to formulate a tuned wind forecast using the model forecast winds at each wind turbine
- Applies statistical model to evaluate power at each wind turbine
- Sums turbine powers to yield farm, connection node and regional power forecasts

Power Curve Estimation Approach

The simplest approach is to apply the manufacturer's power curve for a given turbine to a particular forecast wind to estimate turbine power.

Note however that observed wind turbine power output does not in practice resemble the idealized manufacturer's power curve. (See Figures 1 and 2)

Power curves representing the total power at a farm connection node tend to be better behaved owing to error cancellation but there are still problems. (See Figure 3)

Figure 1.

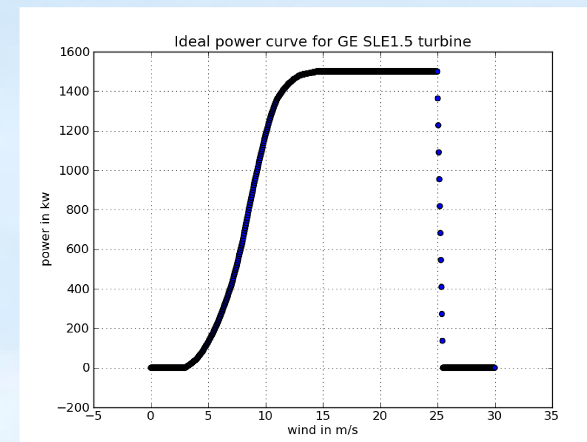


Figure 2.

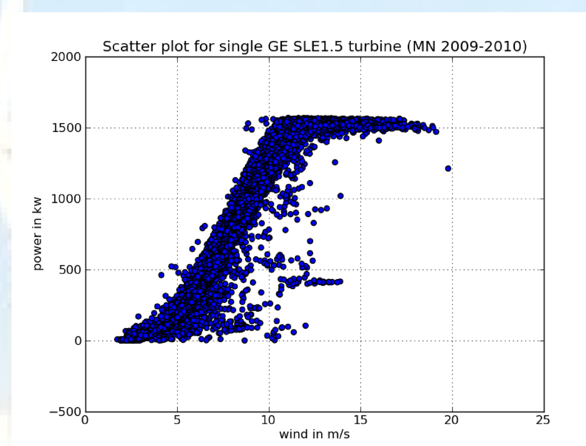
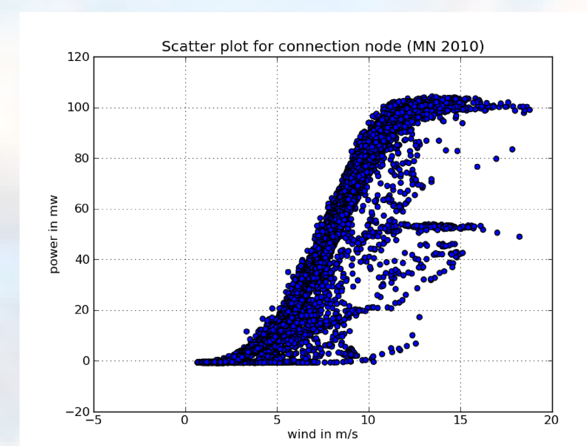


Figure 3.



Power Curve Approximation Techniques

The power curves depicted in the previous slide demonstrate that estimating power using wind speed alone can lead to significant errors.

Using Wind and Power to Forecast Power

Owing to the inaccuracy mentioned above, we decided to investigate incorporating previous wind speed and power data in addition to current wind speed data in order to reduce wind to power conversion error.

Techniques Explored:

Power Curve
Regression Tree (Cubist)
Random Forest Regression (R package)
KNN Nearest Neighbor
Persistence

Power Curve:
Utilize industrial power curve.

Regression Tree:
Recursively partition the domain of the training variables into sub-domains. Perform a linear regression in each sub-domain to estimate the target variable. Evaluate target based on the sub-domain of target's predictors choosing the appropriate regression.

Random Forest Regression:
Similar to regression tree but use multiple trees and average final result.

KNN Nearest Neighbor:
Given a target's predictors, find k nearest neighbors in training set. Find average or median of their target values.

Persistence:
Persist previous observed power.

Data Sets

This particular study consists of two parts: a turbine study and a connection node study.

Turbine Data

The turbine study utilizes Nacelle wind speed and turbine power data gathered for three GE SLE 1.5 megawatt turbines in Minnesota over a period of approximately one year starting mid-2009. The data are broken up into 15 minute time intervals starting at the top of each hour. For each turbine 2/3 of the data was used for the training set and 1/3 of the data was used for the test set.

Connection Node Data

The connection node study utilizes approximately one year's worth of average Nacelle wind speed and farm connection power gathered from a single wind farm in Minnesota. The farm is the same farm from where the turbine data was gathered. The data are broken up into 15 minute time intervals starting at the top of each hour. Again 2/3 of the data was used for the training set and 1/3 of the data was used for the test set.

Results

Table Description

The tables present mean absolute error (MAE) in kilowatts for the turbine study and megawatts for the connection node study.

The Different Approximation Techniques

Curve represents results derived from using the standard power curve in going from wind to power.

Tree represents results derived from using a regression tree. The training variables used were previous wind, previous power, observed wind and the target variable was observed power.

Forest represents results derived from using a regression forest with 200 trees. The training and target variables utilized are identical to those of the regression tree application.

KNN represents results derived from using KNN nearest neighbor with k=20. Again the training and target variables are identical to those of the regression tree application.

Persist represents persisting the previous observed power.

Turbine Power MAE in KW

	Curve	Tree	Forest	KNN	Persist
Turbine 1	44.4	17.8	19.1	89.4	78.3
Turbine 2	48.4	16.7	18.6	83.2	74
Turbine 3	35.2	17.2	18.7	93.1	73.8

Turbine Power Normalized MAE

	Curve	Tree	Forest	KNN	Persist
Turbine 1	0.030	0.012	0.013	0.060	0.052
Turbine 2	0.032	0.011	0.012	0.055	0.049
Turbine 3	0.023	0.011	0.012	0.062	0.049

Connection Node Power MAE in MW at a 102 MW Farm

	Curve	Tree	Forest	KNN	Persist
Connection Node	2.99	0.685	1.28	3.71	3.7

Summary of Findings

Turbine Results

The turbine results illustrate that the regression tree and regression forest produce the lowest errors for the set of techniques chosen here and for the data set under consideration. Applying the industrial power curve directly to the turbine wind increased the error by a factor of approximately 2 to 3 times. The KNN algorithm was in line with persistence and trailed in performance. Note that the Cubist implementation of the regression tree is significantly more efficient than the R implementation of random forest outperforming it by more than a factor of 100.

Connection Node Results

The connection node results illustrate again that both the regression tree and regression forest are yielding the best error results out of the techniques evaluated. Note that the tuning parameters for the random forest were not evaluated exhaustively so there is potential for improved performance. Still the application of random forest generally requires some additional tuning and significantly longer run times.