Predicting Daily Wind Power Forecast for Optimal Thermal Power Backup

Group A4

Rahul Gupta (61710574)
Souvik Sen (61710577)
Srikanth Akkapeddi (61710122)
Avinash Dhanuka (61710696)
Mihir Kedia (61710267)
Animesh Biyani (61710155)

Problem Description

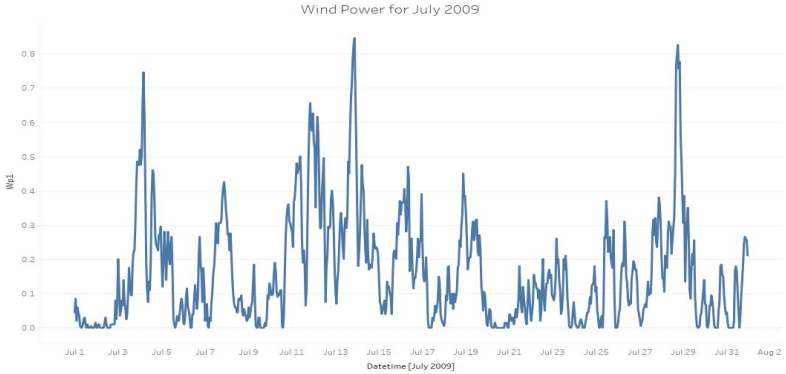
- Utility companies that rely heavily on wind power need to maintain backup power level because of the intermittency of wind power.

- The backup power is generated using non-renewable sources like coal, which makes it expensive and dirty.

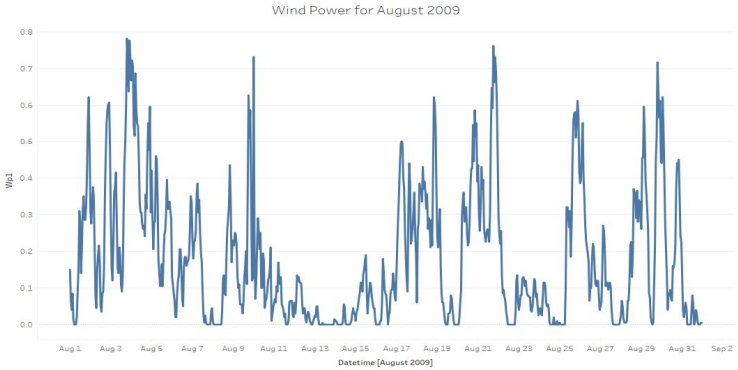
- With accurate wind power forecasts, the utility companies can optimize the level of backup power required.

Data Description & Preparation

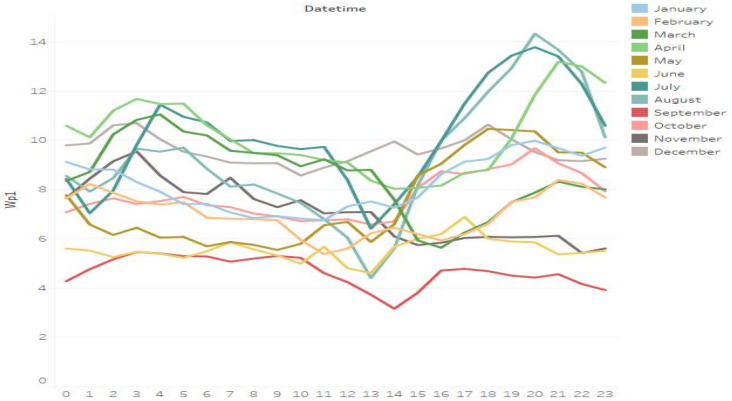
- Historical hourly wind power (normalized) from July 1, 2009 – July 1, 2012 (more than 1,00,000 rows)
- Reduced the dataset to 10,000 rows spanning a period from July 1, 2009 – Aug 20, 2010
- Hourly forecasts of wind speed and wind direction provided by meteorological department (for each hour, 4 forecasts are made, we took the average)
- Next step was to identify the time series components i.e. seasonality and trend.



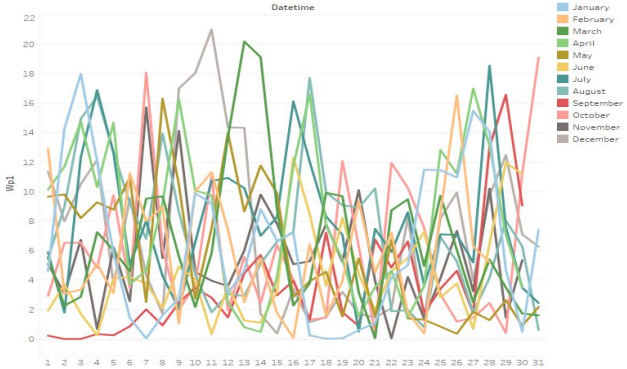
The trend of sum of Wp1 for Datetime. The view is filtered on Datetime, which ranges from 01-Jul-09 12:00:00 AM to 31-Jul-09 11:59:59 PM.



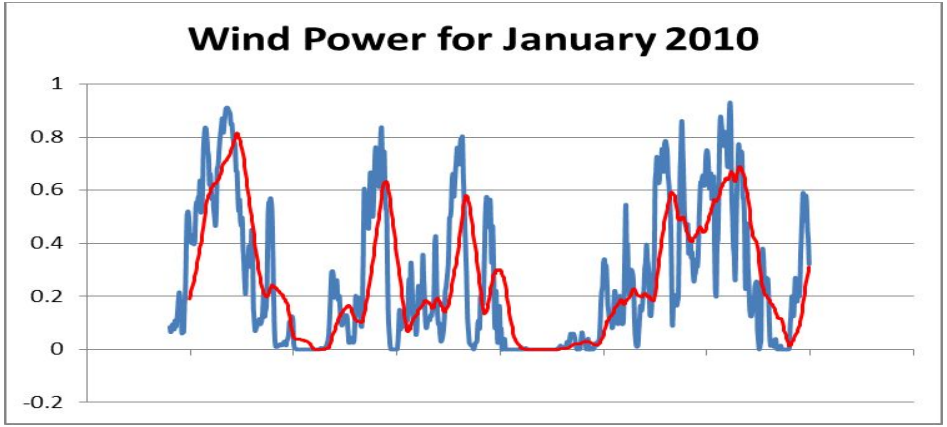
The trend of sum of Wp1 for Datetime. The view is filtered on Datetime, which ranges from 01-Aug-09 12:00:00 AM to 31-Aug-09 11:59:59 PM.

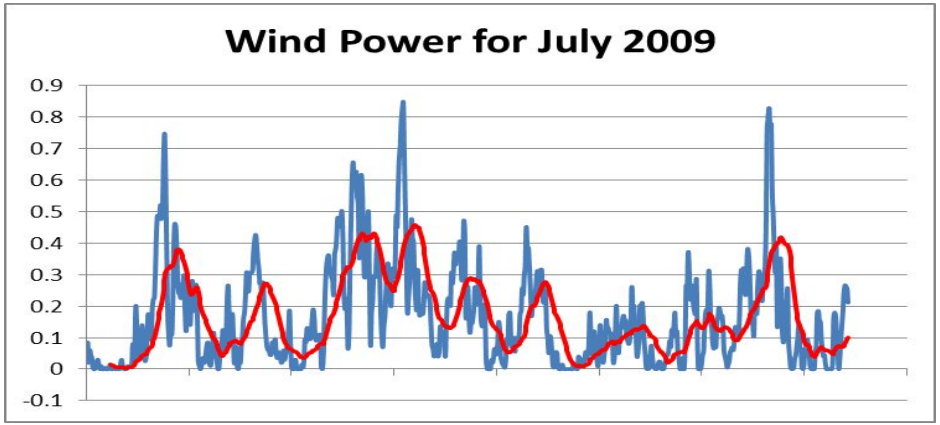


The trend of sum of Wp1 for Datetime Hour. Color shows details about Date Month. The data is filtered on Datetime, which ranges from 01-Jul-09 12:00:00 AM to 20-Aug-10 11:00:00 PM.



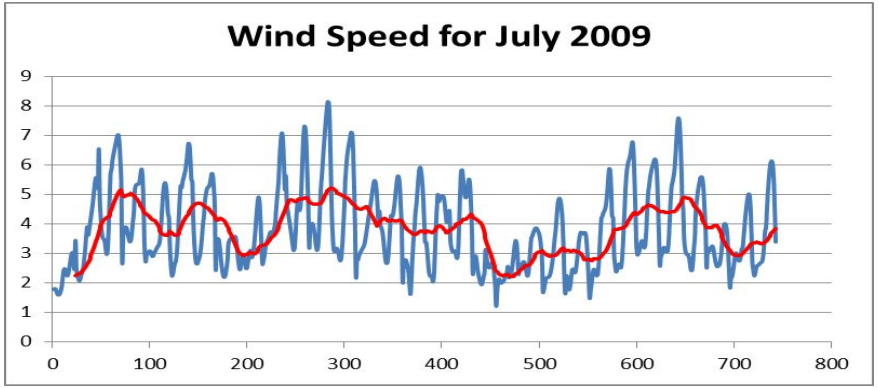
The trend of sum of Wp1 for Datetime Day. Color shows details about Date Month. The data is filtered on Datetime, which ranges from 01-Jul-09 12:00:00 AM to 20-Aug-10 11:00:00 PM. The view is filtered on Date Month, which keeps 12 of 12 members.

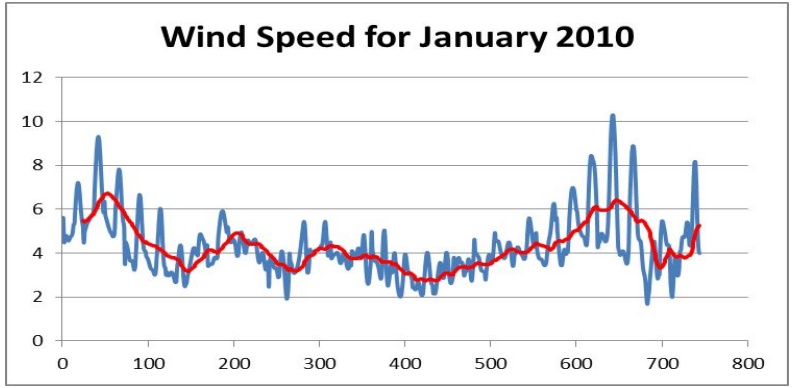

Wind Power for January 2010
1
0.8
0.6
0.4
0.2
0
-0.2

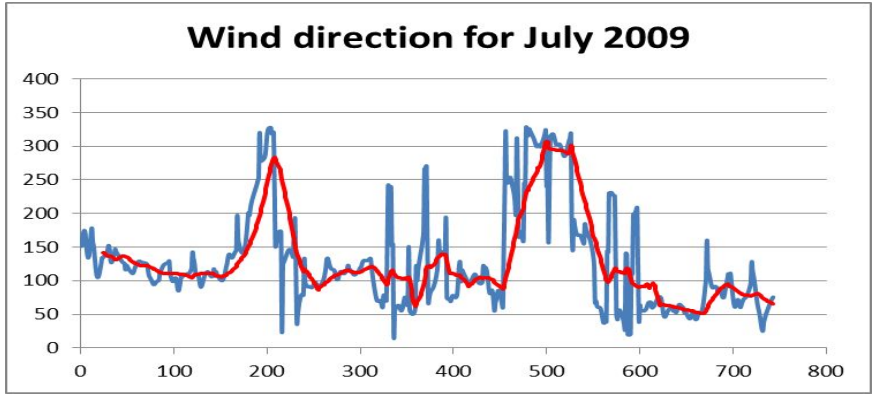

Wind Power for July 2009
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1

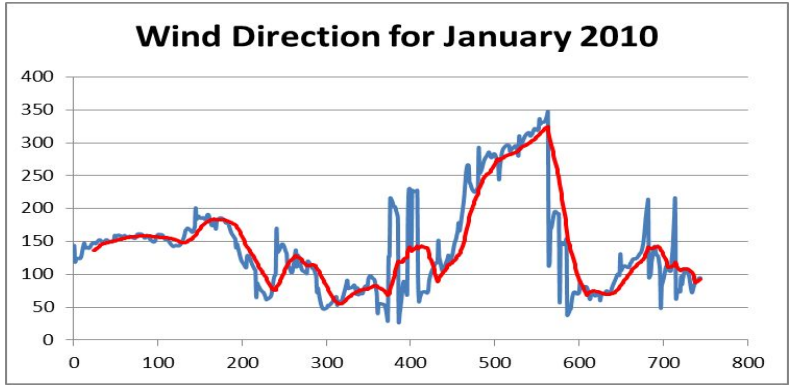
Models used

- Since there is no seasonality in the data, and the trend is not a typical one, we cannot apply any of the exponential smoothing methods directly.
- We adopted a two-step approach. In the first step we de-trended the data. After that we used methods like simple exponential smoothing and AR models to extract more information from the residuals.
- To de-trend the data we had two options – either use a centered MA or regression. It turns out that the explanatory variables (avg wind speed, wind direction) have a similar trend, and we can use these in the regression to de-trend the wind power.


Wind Speed for July 2009
9
8
7
6
5
4
3
2
1
0
0
100
200
300
400
500
600
700
800

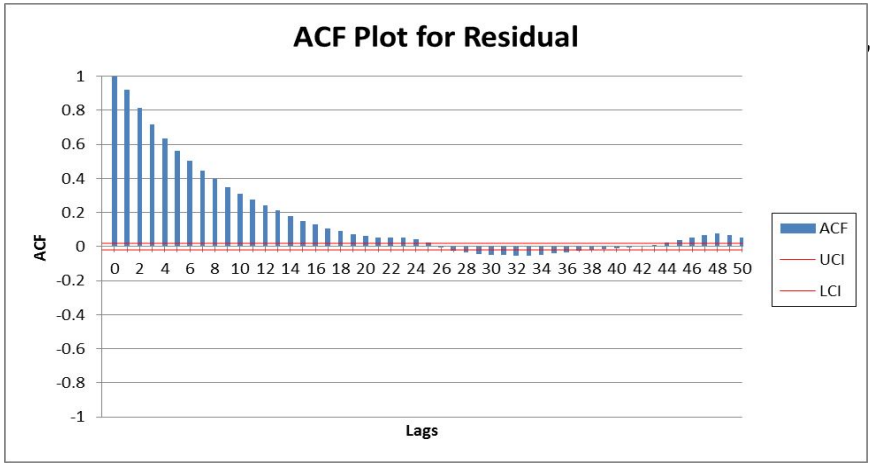

Wind Speed for January 2010
12
10
8
6
4
2
0
0
100
200
300
400
500
600
700
800

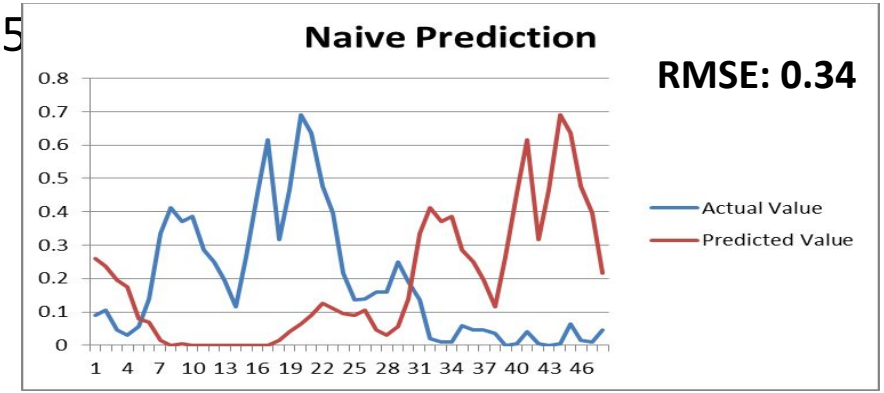

Wind direction for July 2009
400
350
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700
800

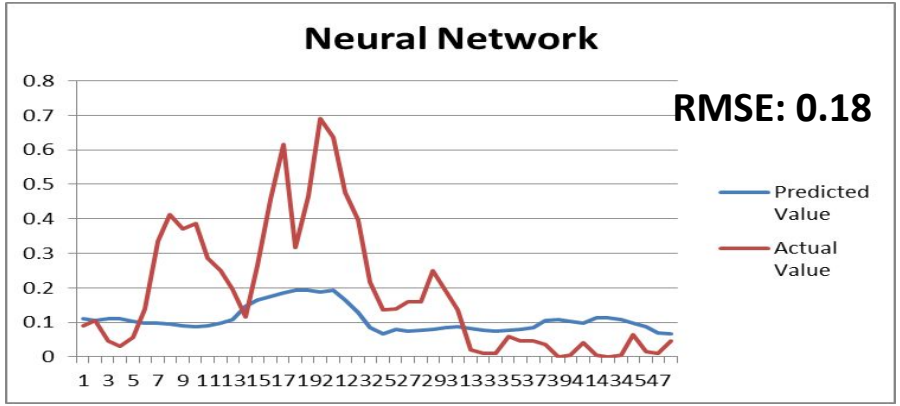

Wind Direction for January 2010
400
350
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700
800

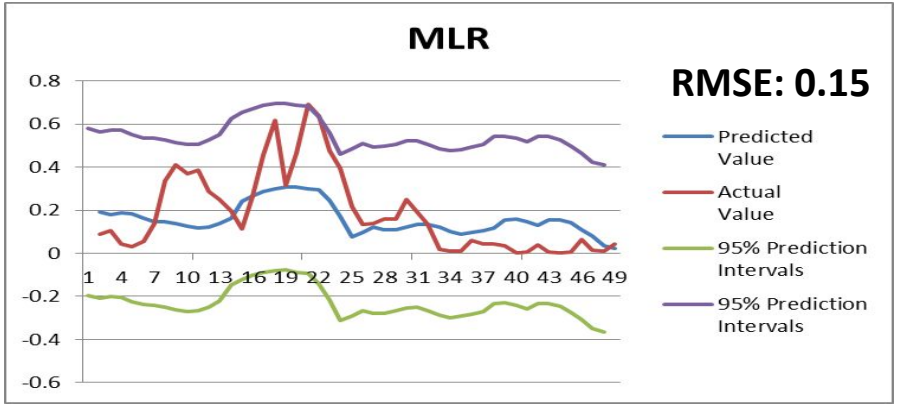
Models used

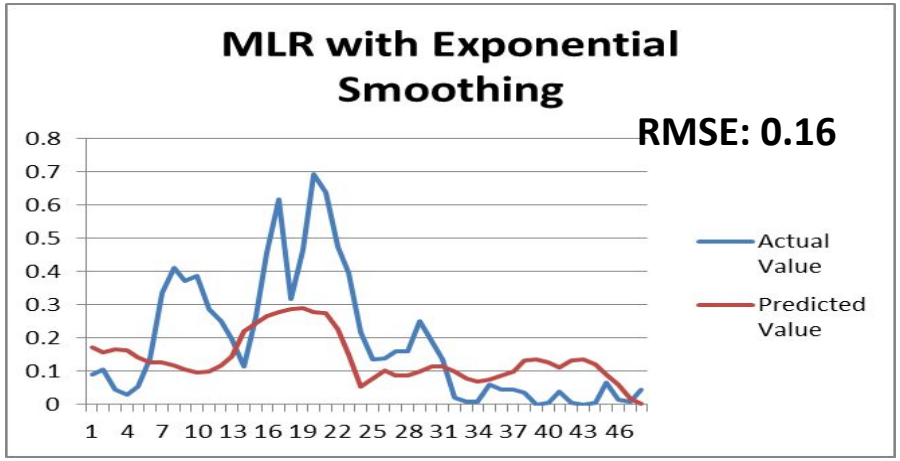
- Based on the above approach, we tried the following models:
  - Naïve (to establish a benchmark)
  - MLR to de-trend, followed by Simple Exponential Smoothing
  - MLR to de-trend, followed by another MLR with Lag 24 and Lag 48 as explanatory variables (see ACF plot below)

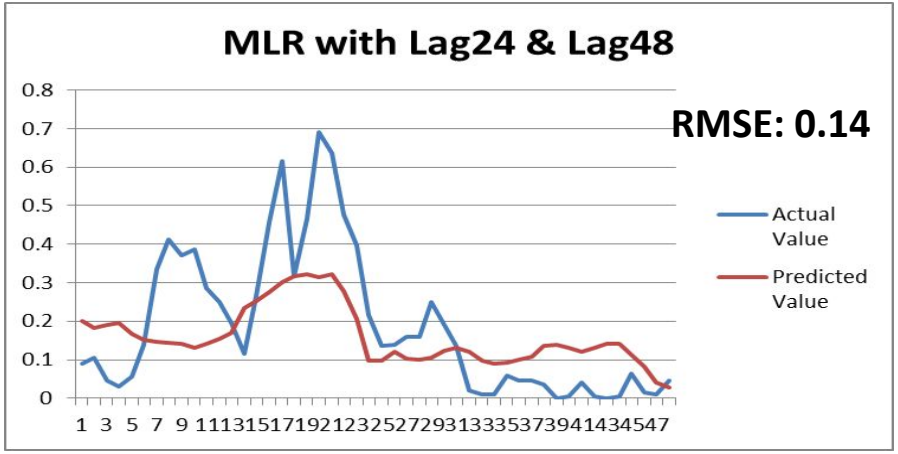





Neural Network
RMSE: 0.18
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 111315171921232527293133353739414345 47
Predicted Value
Actual Value


MLR
RMSE: 0.15
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49
Predicted Value
Actual Value
95% Prediction Intervals
95% Prediction Intervals


MLR with Exponential Smoothing
RMSE: 0.16
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Actual Value
Predicted Value


MLR with Lag24 & Lag48
RMSE: 0.14
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 111315171921232527293133353739414345 47
Actual Value
Predicted Value

Conclusion & Recommendations

- Cost of over-forecasting greater than under-forecasting. Must pay attention to the prediction intervals.

- Simple MLR performs better than a Neural Network in explaining the variation. Extracting more information using Exponential smoothing and MLR of residuals with lagged predictors gets us only a marginal improvement in RMSE.

- Predictions from different models follow the same trend, therefore taking ensembles will not result in improved performance. Best method to use is simple MLR followed by