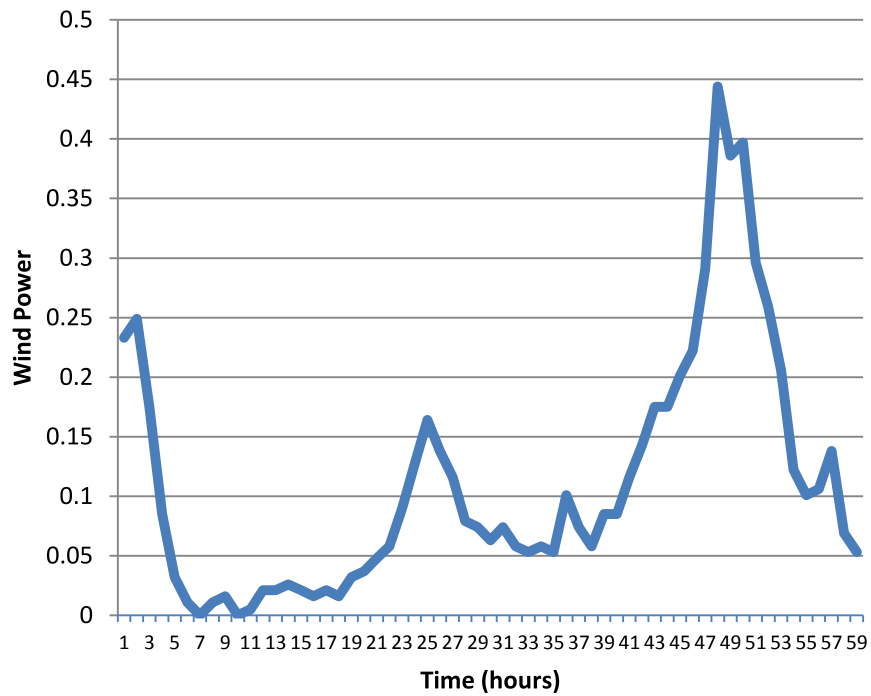
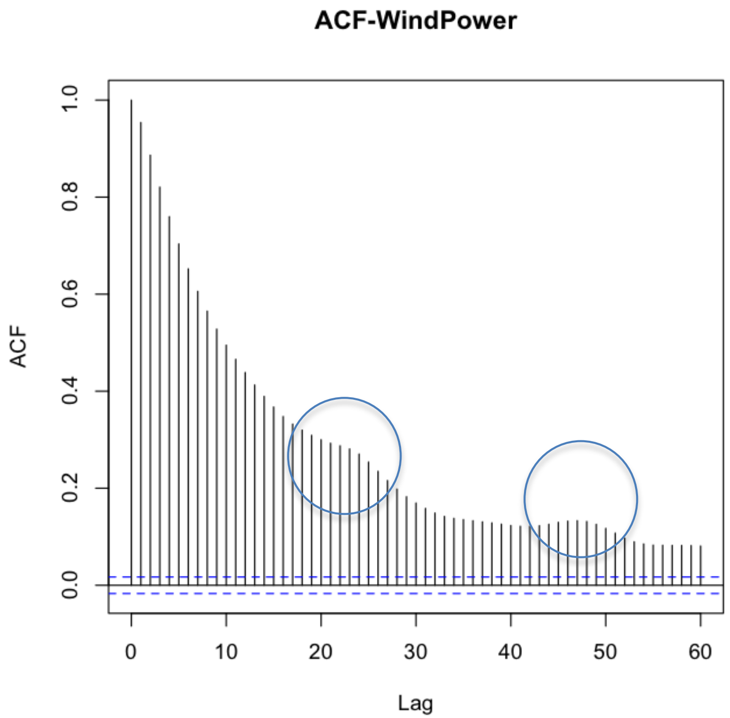
Wind Power Forecasting Problem

- Kaggle project sponsored by IEEE Power & Energy Society
- Given: Model identification period
- Need to forecast 48 hours ahead based on last 36 hours wind power
- Challenges:
  - Chaotic nature of atmosphere & incomplete understanding
  - Accuracy decreases when we forecast a period far ahead in the future
  - Sudden jumps in data series

Data Series And Autocorrelation

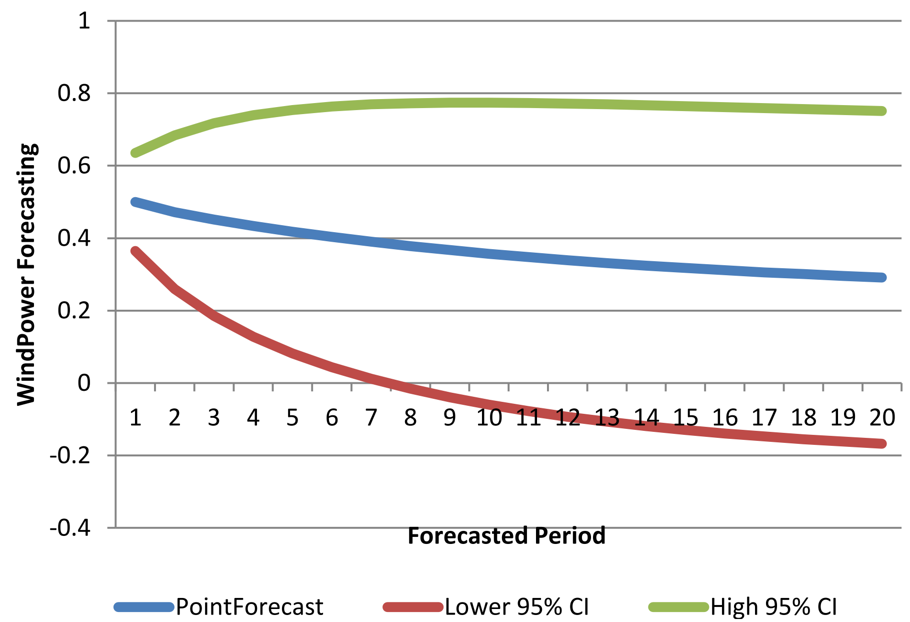




- Unpredictability of weather
- Strong correlation at lag 1, lag 24, and lag 48

ARIMA

- Can't use plain off the shelf ARIMA model to forecast 48 periods ahead
  - Wide confidence interval
  - Negative values!



- Need to take advantage of inherent structure in the system

$$y_t = \frac{\phi_0 + \phi_1 y_{t-1} + \ldots \phi_p y_{t-p}}{\theta_0 a_t + \theta_1 a_{t-1} + \ldots \theta_q a_{t-q}}$$

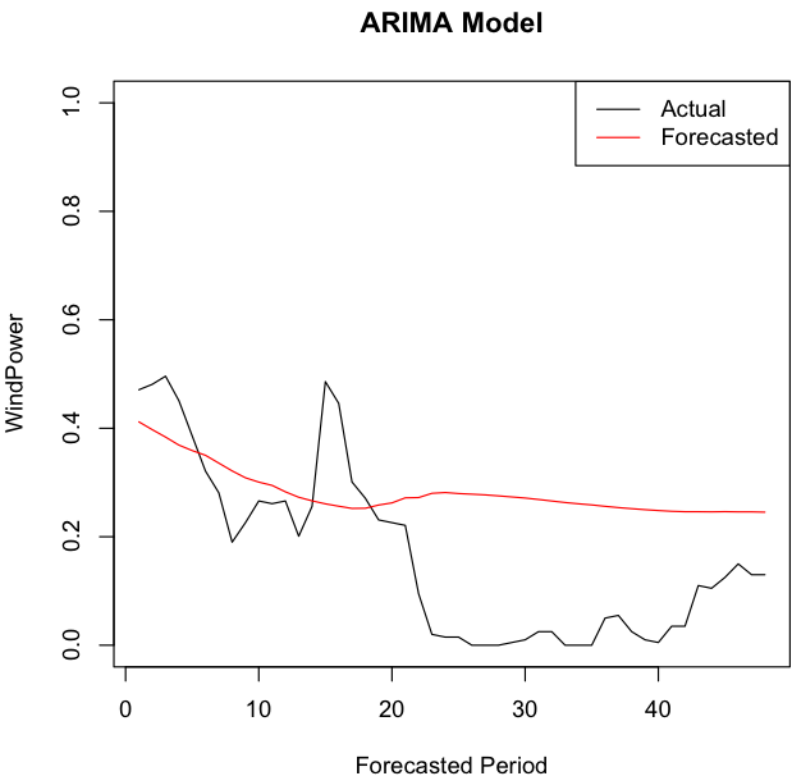
ARIMA Model

- Use
  - High correlation at lag 1 and lag 24
  - Last known good value

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-24}$$

$$y_{t+l} = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t+l-24} \qquad 1 \leq l \leq 23$$

$$y_{t+l} = \phi_0 + \phi_1 \hat{y}_{t-24} + \phi_2 y_{t+l-48} \qquad 24 \leq l \leq 47$$



RMSE = 0.2609

- Regularization

$$\hat{\phi} = \arg\min_{\phi} ||y_t - \phi^T y_{t-l}||^2$$

$$0 \leq \phi_1 \leq C$$

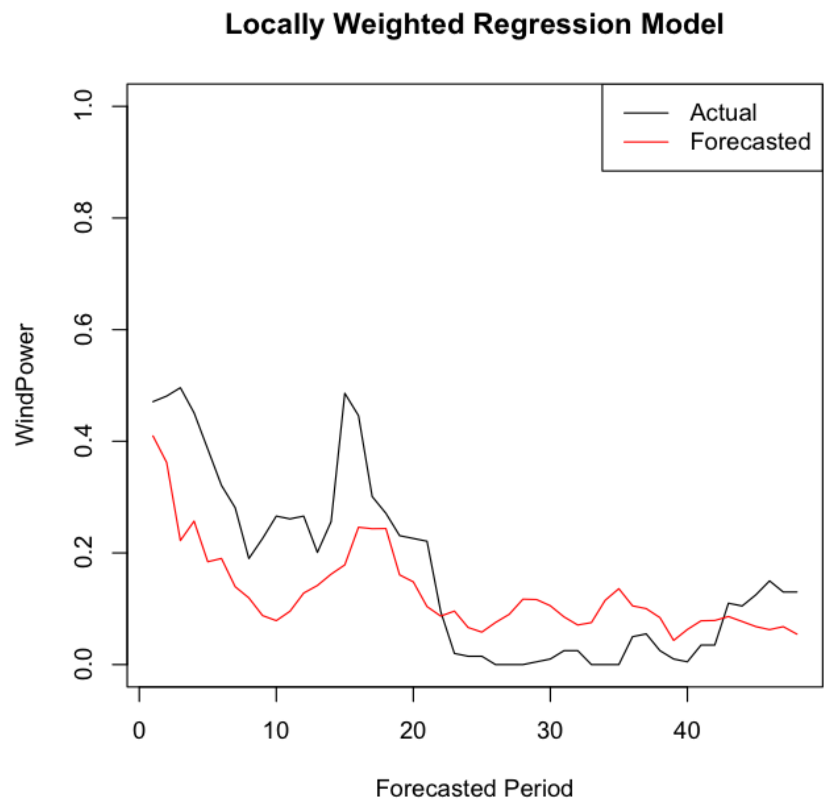
$$0 \leq \phi_2 \leq C$$

Locally Weighted Regression

- Divide the model identification data in chunk of 36+48 periods
- Use the similarity of first 36 period to forecast the next 48 period

$$w^{(i)} = \exp\left(-\frac{(x^{(i)} - x)^2}{2\tau^2}\right)$$

- Cross-Validation
  – Use cross-validation to find the value of bandwidth parameter



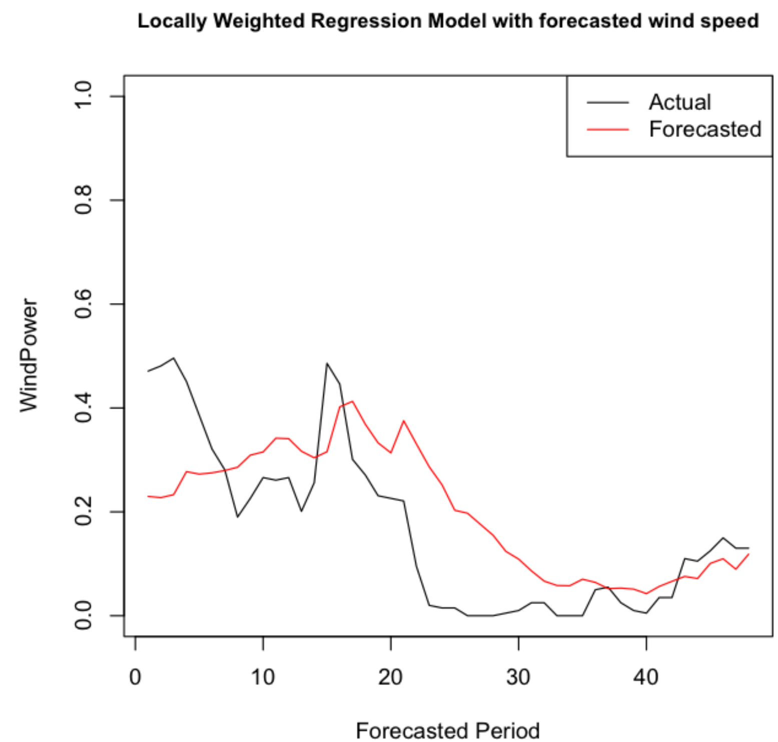
RMSE = 0.3343

Locally Weighted Regression With Forecasted Wind Speed

- Instead of relying only on univariate wind power, use
  - Forecasted wind speed
  - Forecasted direction
  - Forecasted zonal component
  - Forecasted meridional component

- Using cross-validation test set:
  - Only forecasted wind speed was had a significant impact

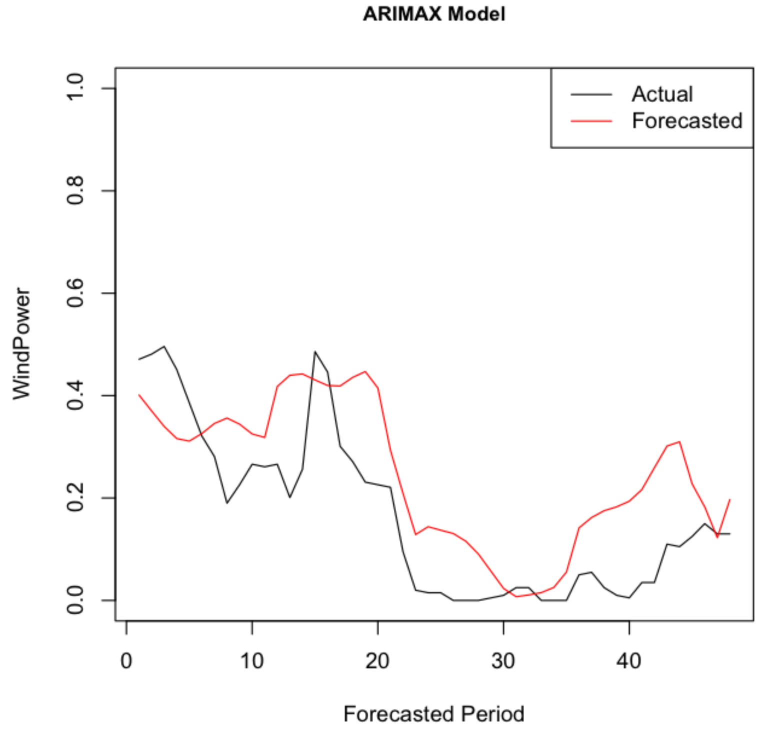
$$w^{(i)} = \exp\left(-\frac{(x^{(i)} - x)^2}{2\tau^2}\right)$$



RMSE = 0.2050

ARIMAX Model with Forecasted Wind Speed

- High correlation between wind power and wind speed
- Wind direction, zonal, meridional components were not found to have significant correlation
- Include as part of linear regression model



RMSE = 0.19

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-24} + \phi_3 w_t$$

$$y_{t+l} = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t+l-24} + \phi_3 w_{t+l} \qquad 1 \leq l \leq 23$$

$$y_{t+l} = \phi_0 + \phi_1 \hat{y}_{t-24} + \phi_2 y_{t+l-48} + \phi_3 w_{t+l} \qquad 24 \leq l \leq 47$$

Summary

- Best indicators for forecasting wind power in a given hour:
  - Predicted wind power generated over the past hour
  - Wind power generated over the same hour one day ago
  - Most recent wind speed forecast for the current hour
  - Forecasts for wind direction, zonal and meridional wind components were not significant
- Performance of ARIMAX model is comparable to Kaggle's best entries
- ARIMAX model is more interpretable

Results

Model	RMSE
Best Kaggle Entry	0.15
ARIMAX Model with Wind Speed	0.19
Locally Weighted Regression with Wind Speed	0.21
ARIMA Model	0.26
Locally Weighted Regression	0.33
Benchmark	0.35