Application of a Random Forest Approach to Model Output Statistics for use in Day Ahead Wind Power Forecasts

Edward Natenberg, John Zack, John Manobianco, Glenn Van Knowe and Tim Melino
MESO, Inc. AWS Truepower, LLC

David John Gagne
University of Oklahoma

Motivation

- Balancing authorities, which include independent system operators and electric utilities, use day-ahead (48-hour) variable generation (e.g. wind and solar) forecasts to plan for and allocate the day-ahead generation resources.
- Currently, wind provides the majority of variable power production.
- The day-ahead wind forecast is an essential tool to estimate wind-generated power that will be available on the grid during the next operating day.
- New techniques that can reduce day-ahead wind power forecast error will provide additional value to balancing authorities and ultimately reduce the cost of integrating wind into electric systems.
- Day-ahead wind power forecasts are primarily produced from an ensemble of numerical weather prediction(NWP) forecasts which exhibit systematic non-linear error patterns.
- Systematic NWP forecast error can vary by:
  - Location
  - Season
  - Time of Day
  - Weather Regime
- Various forms of statistical models, known as model output statistics (MOS), are commonly used to minimize the impact of systematic model error on forecasts.
- Traditional MOS techniques have several limitations:
  - Artificial neural networks require long training periods and tend to overfit
  - Support vector machines are slow to train over large datasets and depend on the choice of kernel parameters
  - Linear regression assumes normal distribution and is overly sensitive to outliers
- These limitations make the ensemble decision tree technique known as "random forest" a promising and interesting method that may be applied to NWP forecasts in order to more effectively reduce model forecast error.

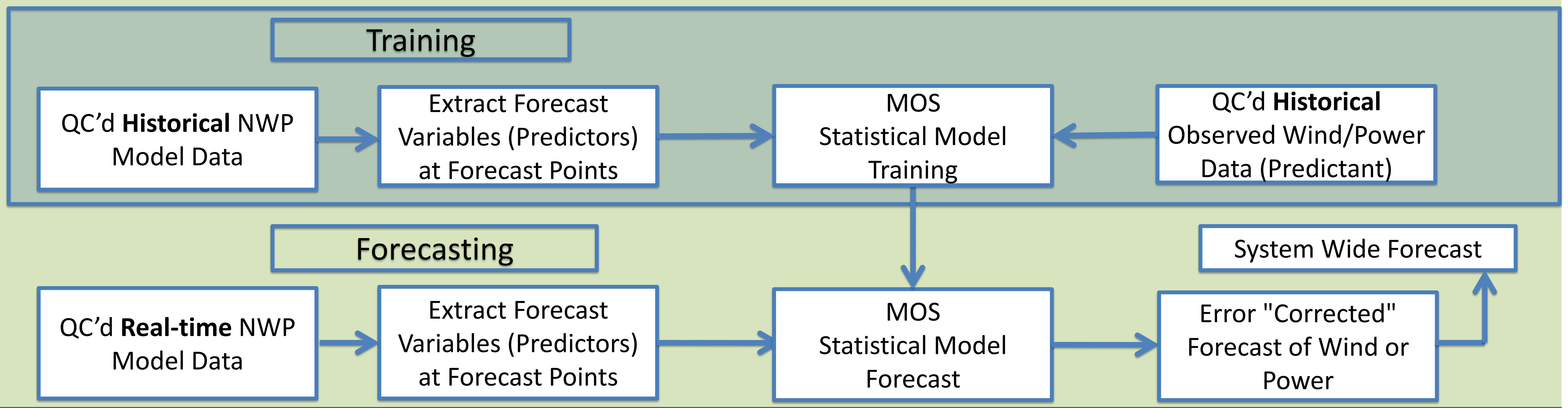
Method

- Random forest (RF)[1] is a non-parametric machine learning method that improves on the decision tree approach[2] in several ways:
  - Random forest trains from an ensemble of trees on resampled versions of the original training set using a bootstrap technique to explore the range of dataset variability.
  - At each node of the tree, only a random subset of the data attributes are evaluated.
- Unlike other techniques random forest trains quickly, is less prone to overfitting and can account for non-linear relationships.

Experimental Design

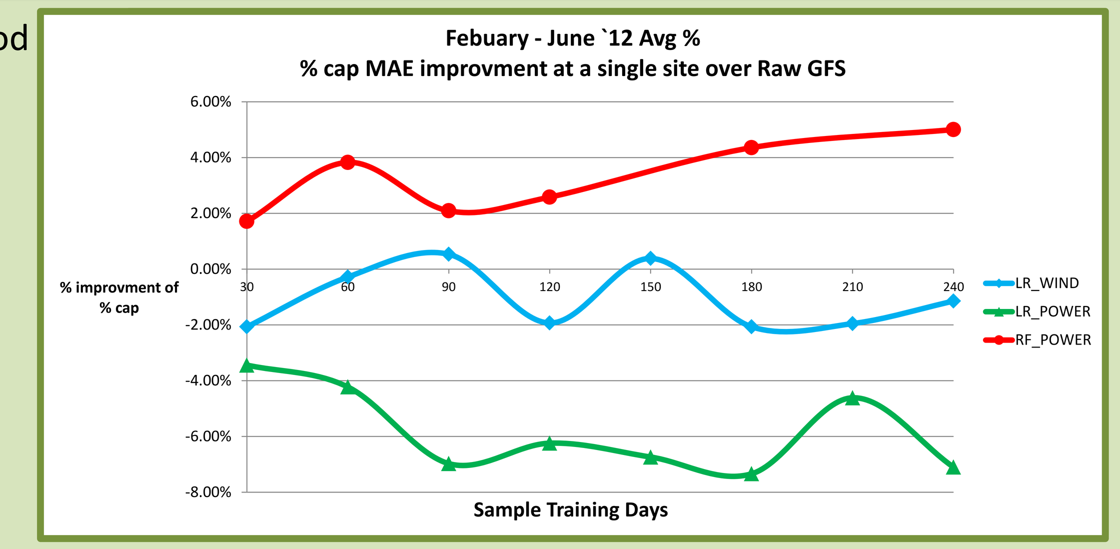
- A RF approach was used to directly predict wind power production (% of plant capacity) and compared to a screening multiple linear regression (LR) technique and model forecasts without bias correction (raw).
- Forecasts were produced for a number of wind farms in Texas and aggregated to generate a system wide forecast (AGG) using two years of observations and model data.
- Predictor variables (input) from the North American Mesoscale Model (NAM) and Global Forecast System (GFS) NWP models at selected heights for each forecast location included:
  - Wind Speed and Direction
  - Temperature
  - Geopotential height
- The RF model was retrained monthly using various training periods sizes
- Each method can be trained to predict either wind or power

Forecast Technique



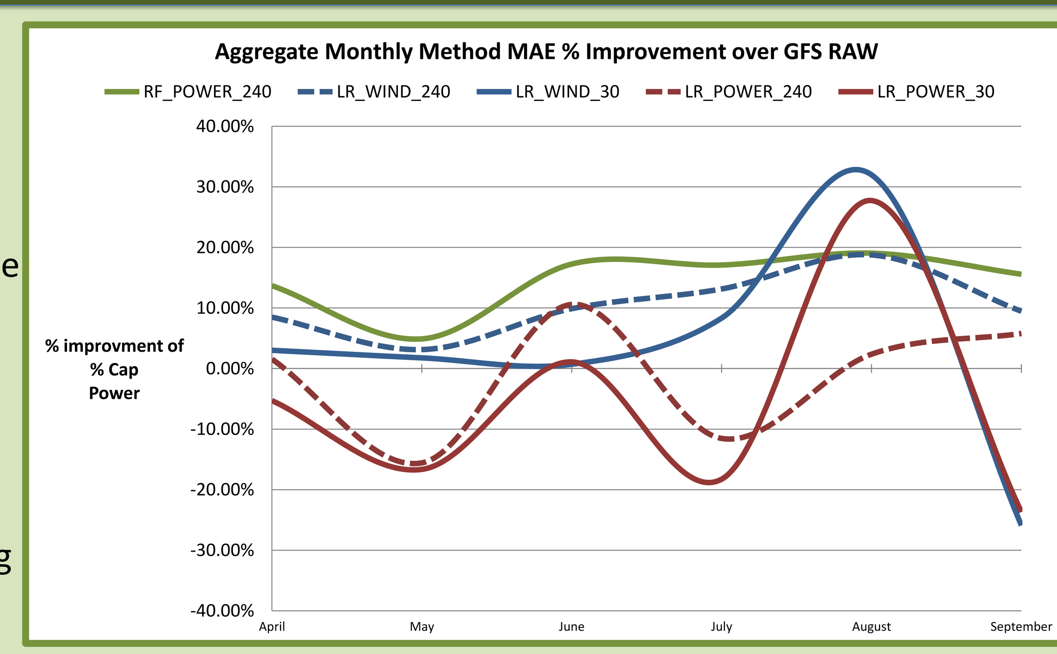
Single Site Sensitivity to Training Sample Size by Method Type

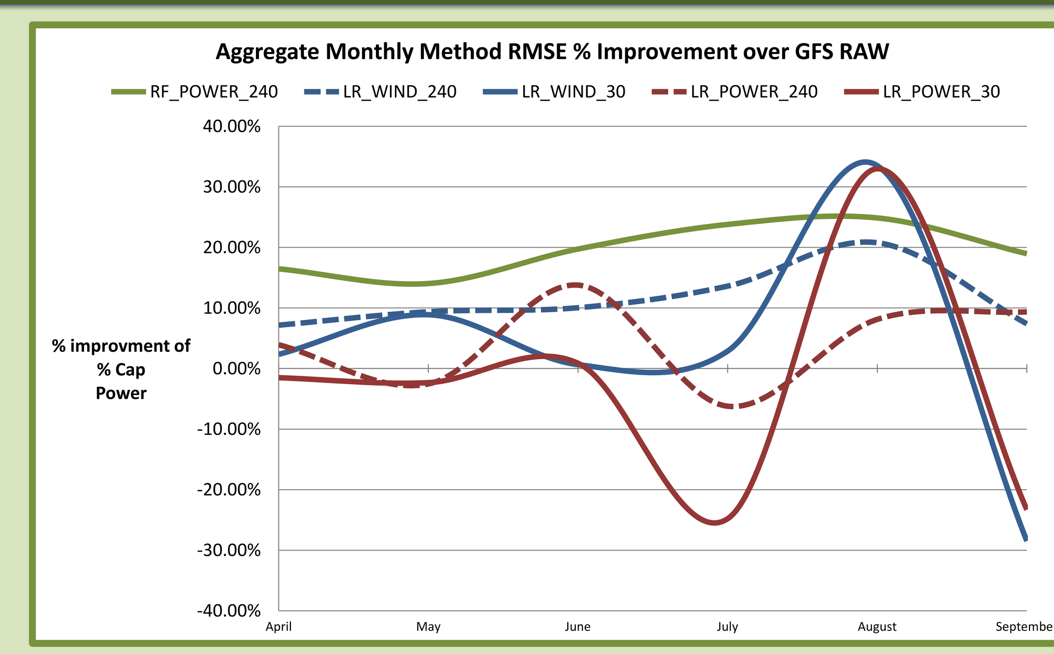
- 48-hour forecasts of hourly power production were produced daily for one site over a 5 month period
- Each method was retrained monthly using 30 -240 training days
- 3 Methods were trained using the GFS model variables (predictors):
  - **LR_Wind:** Linear regression trained to wind speed then converted to power
  - **LR_Power:** Linear regression trained directly to power
  - **RF_Power:** Random forest trained directly to power
- Methods are compared to the raw GFS wind speed forecast put through a power curve
- RF_Power performed the best, with increased improvement for larger training sample sizes
- LR_Power performed the worst most likely due to non-linear correlation from model state to power



Aggregate Forecast Performance by Month

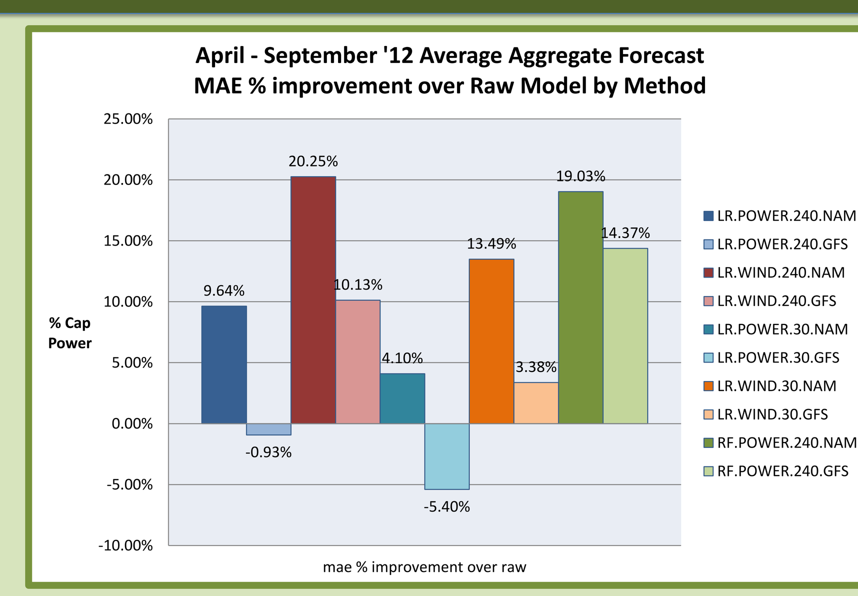
- Trained for both **30** and **240** days once per month
- Aggregate of 50 wind sites
- RF 6 month average RMSE improvement was 19.64% over the raw GFS method as compared to 11.39% and 4.40% for wind and power LR methods (240 training days).
- 30-day training linear regression methods did best in August during a very persistent weather regime.

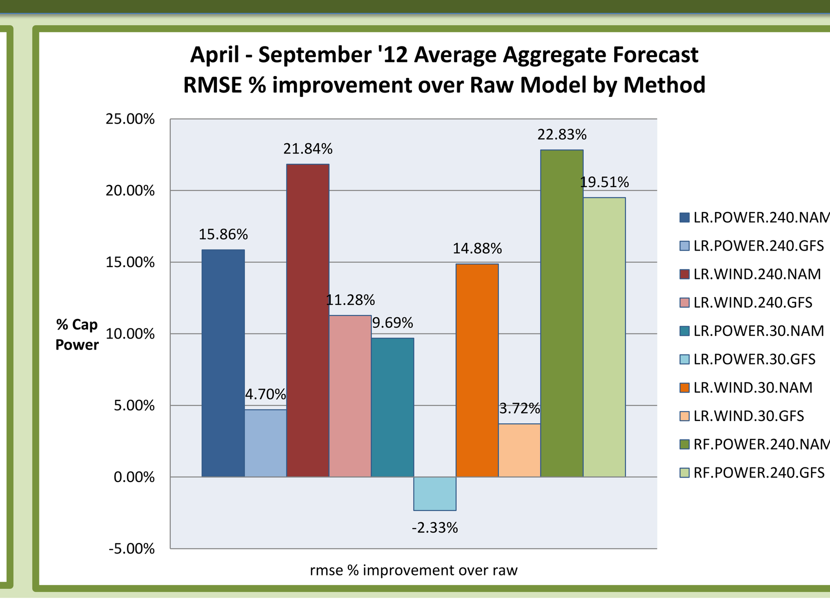




Sensitivity to Input Model

- Compared linear regression to random forest for both GFS and NAM models with 30 and 240 training days.
- Results show % decrease in MAE and RMSE over either NAM or GFS (raw) model wind forecast put through a power curve.
- Darker colors denote NAM methods compared to raw NAM and lighter colors denote GFS methods compared to raw GFS.
- **RF decreases RMSE more than MAE and improves the NAM forecast more than GFS.**
- RF with 240 training days is slightly better than RF for MAE using NAM input, mostly due to the improved performance in July-August months (shown above).

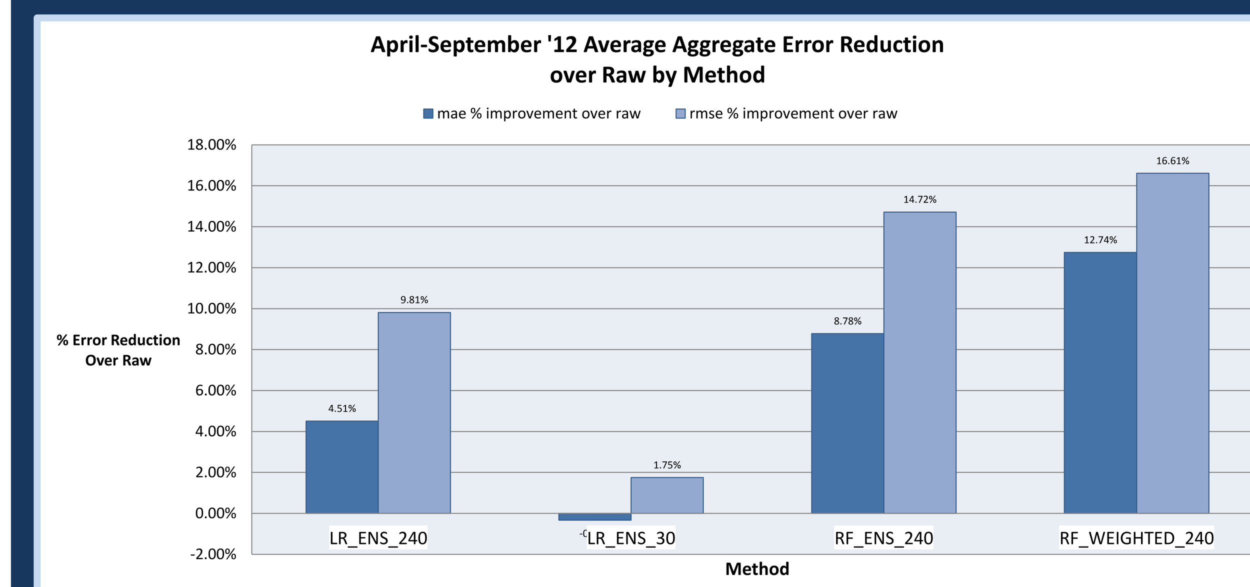


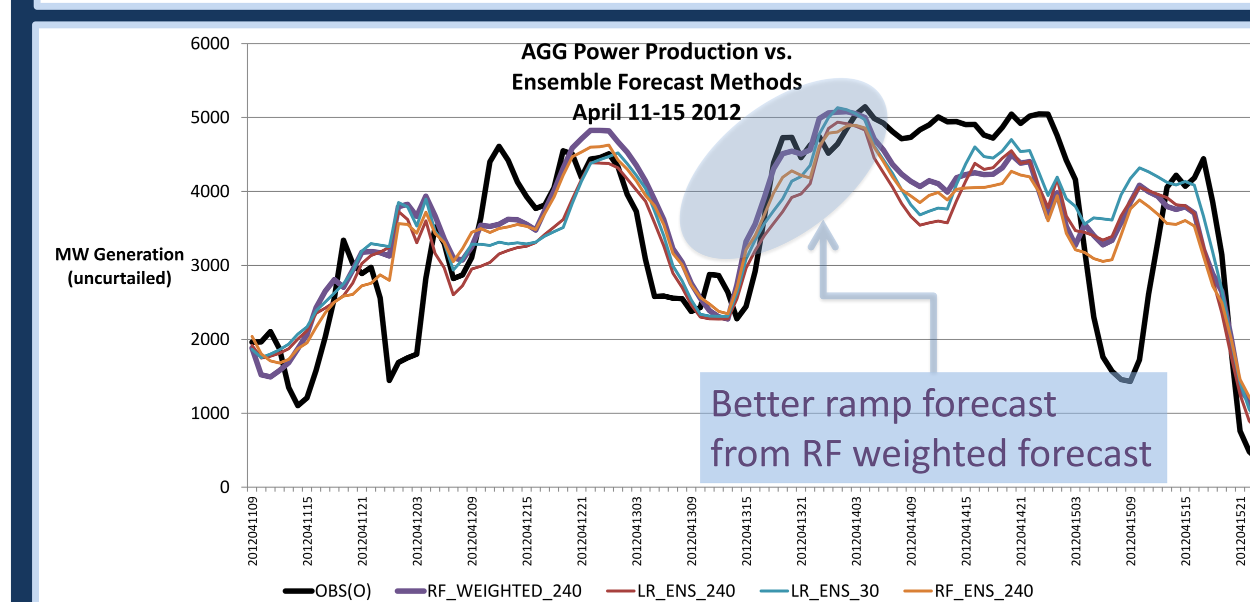


1. Breiman, L., 2001 Random Forests. *Machine Learning*,**45**, 5-32.
2. Breiman, L.,1984: *Classification and Regression Trees*, Chapman & Hall, 358 pp.

Ensemble Results

- To determine the impact of using predictors from more than one model, several ensemble aggregate forecasts were generated for a six month period from April – September 2012:
  - **Raw**: unweighted average of both the GFS and NAM wind speed forecast after a power curve is applied.
  - **LR_ENS_240**: unweighted average of 4 methods; LR_Wind and LR_Power for GFS and NAM data using 240 training days.
  - **LR_ENS_30**: unweighted average of 4 methods; LR_Wind and LR_Power for GFS and NAM data using 30 training days.
  - **RF_ENS_240**: unweighted average of RF_Power for GFS and NAM data using 240 training days.
  - **RF_WEIGHTED_240:** RF weighting predictors from both GFS and NAM data using 240 training days.
- Uneven weighting of model state variables in RF reduces the power production forecast error by 3.96% for MAE and 1.89% for RMSE over an average of RF-adjusted NAM and GFS for a 6 month period.





Conclusions and Future Work

- RF is an efficient machine learning technique that can reduce the error of NWP-based day-ahead wind power generation forecasts.
  - RF trains quickly, is less prone to overfitting than other methods, and can account for non-linear interactions among predictors.
- RF took advantage of larger training samples more effectively than a screening linear regression technique for a 5-month evaluation period.
- RF-based forecasts using predictors from 2 NWP models improved upon a simple average of RF-adjusted forecasts from each model suggesting that RF has skill in unevenly weighting the individual forecasts.
- Area for further development:
  - Screening of candidate predictors to identify which are most useful in order to limit predictor to a smaller set of the most effective predictors without overfitting.
  - Regime-based training (pre-partitioning of sample into subsamples or addition of regime based predictors)