Random Forest Ramp Forecasts for Groups of Wind Farms

Pål Preede Revheim, Hans Georg Beyer
Dept. of Engineering Sciences
University of Agder, Grimstad, Norway

Outline

- Introduction
- Wind power ramps
- Data
- Classification trees
- Random forests
- Results
- Conclusions and outlook

Introduction

- Increasing volumes of wind power in electricity grids give an increasing need for reliable and accurate wind power forecasts

- One of the major challenges is sudden large changes in wind power production, normally referred to ramp events

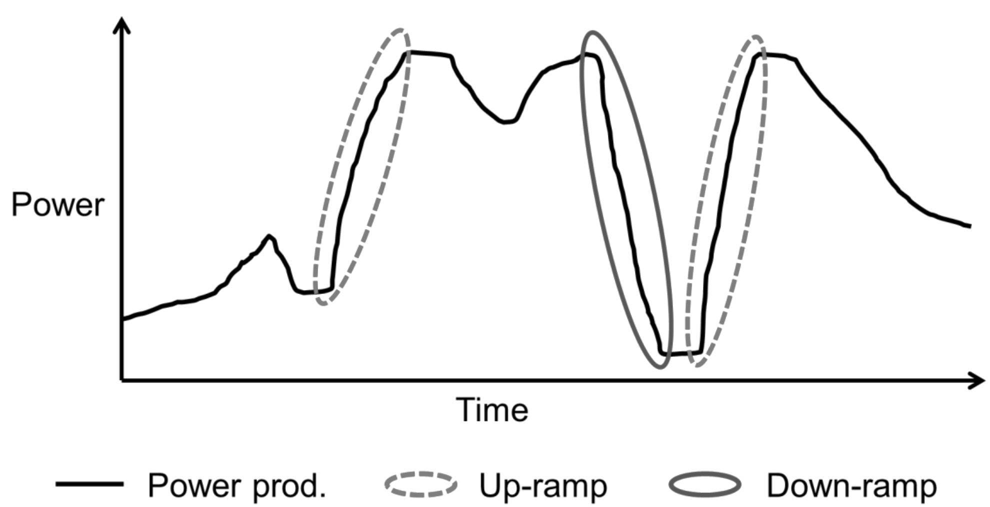
- The sooner and more accurate ramps can be predicted, the smoother and more efficiently they can be dealt with

- In this presentation we look at the possibilities of using the classifiction technique random forest to forecast ramps in the lumped output of a group of wind farms

Wind power ramps

- Wind power ramps are «large» and «sudden» changes in the wind power
  - No agreement on a formal definition, depends on the energy system
  - Here «large» is defined as a change that exceeds **50 %** of the rated output capacity, and «sudden» defined as **3 hours**

- Wind power ramps can be caused by:
  - Large changes in the wind speed
  - The wind speed exceeding the wind turbines upper cut-out speed (typically 25 m/s).
  - The wind speed dropping to below the lower cut-out limit



Wind power ramps

- Here we forecast wether the change in wind power production from time *t* to time *t+3 hours* will exceed the 50 % limit
  - Deterministic forecast with 3 classes (0 = no ramp, 1 = up-ramp, 2 = down-ramp)

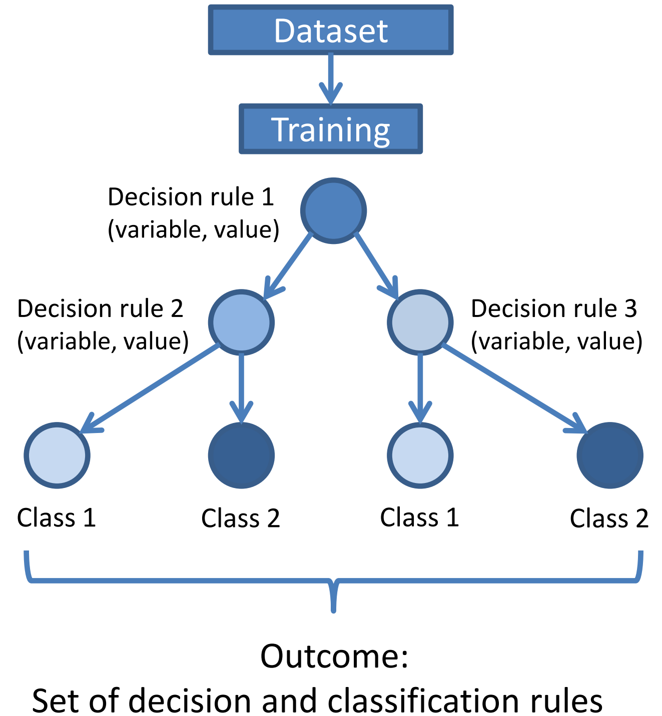
- Ramp forecasts are made using data inputs from:
  - NWP-based wind power forecasts for the group of interest and neighboring groups
  - Wind power measurements from the time the forecast is issued as well as the two preceding hours from the group of interest and neighboring groups

- Random forests used to:
  - Identify situations in the empirical data when ramps have a high probability of happening
  - Select the variables that influence the ramp probability
  - Place new sets of forecasts and measurements into one of the three classes (0, 1 or 2)

- «True ramps» (for evaluation) are identified applying the 3 hr/ 50% definition to a time-series of measurements only

Classification trees 1

- The basis of a random forest is classification trees
  - Hierarchical structures that identifies and learns patterns in empirical data
  - Makes decision rules (forecast rules) through matching patterns with known class outcomes in a training dataset
- Patterns are expressed as a number of decision rules, where the decision rules aim at minimizing some property of the resultant classes
  - Gini impurity index most common
- Forecasts for new sets of observations are made by running the variables through the decision rules finding the best match with the identified patterns and known classes



Classification trees 2

Classification trees have the advantage that they are:

- Easy to interpret
- Require little data preparation
- Good at handling non-linearities, cirkular data etc.
- White-box model

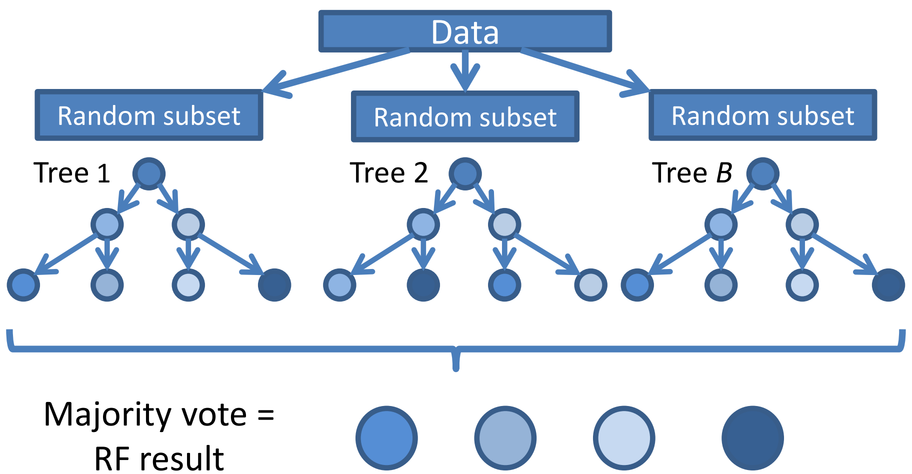
But they are also:

- High variance
- Easily overfitted

Random forests is a technique to deal with these problems

Random forests 1

- Random forests (Breiman, 2001) is an ensemble learning method for classification that operate by constructing a large collection of de-correlated classification trees by chosing a random subset of the input data

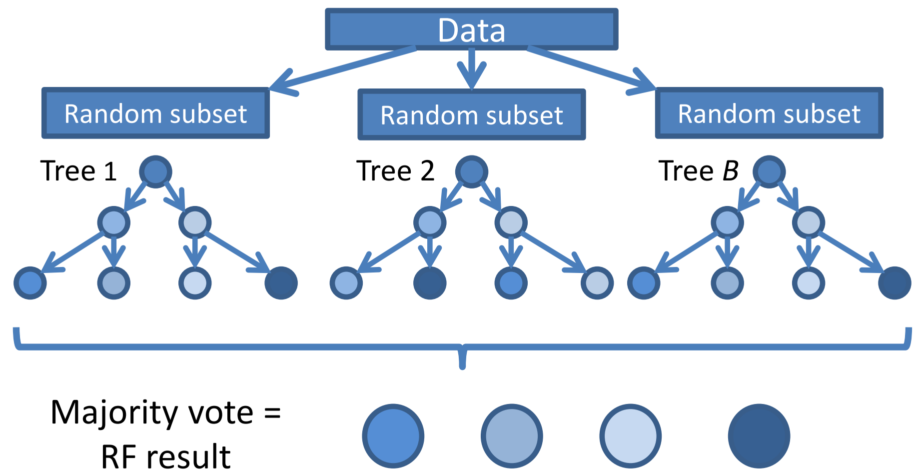


- Through the random forest procedure the variance of the classifier is reduced by $\frac{\sigma^2}{B}$, where B is the total number of trees, for as long as the trees are independent

Random forests 2

- The building of a random forest goes through four steps:

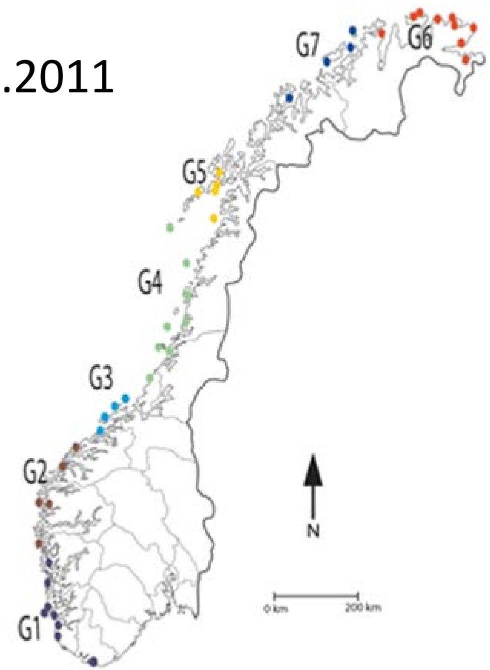
  1. The size of the random subsets for each tree is determined
  2. The importance of each variable is calculated
  3. The dataset is reduced by excluding the less important variables
  4. A final model is built based on the reduced dataset.



- To classify new observations each set of predictor variables are classified using all trees in the final model
- The most common outcome (majority vote) is used as the random forests classification

Data

- wind data from time period 01.01.2009 to 17.12.2011
- 43 weather stations run by Met.no
  - 10 m height
  - Hourly updates
- Hirlam 4*4 km forecasts
  - Hourly forecasts
  - Updated every 24 hours
- Height transformation
- Transformation to power output
- data gathered in 7 groups

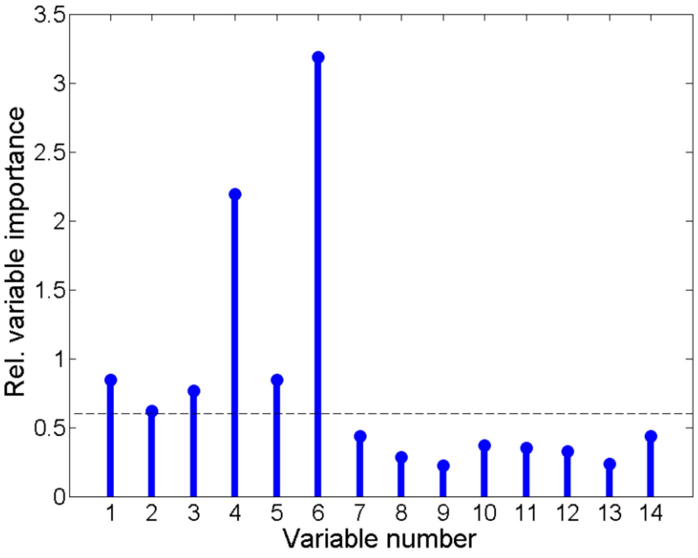


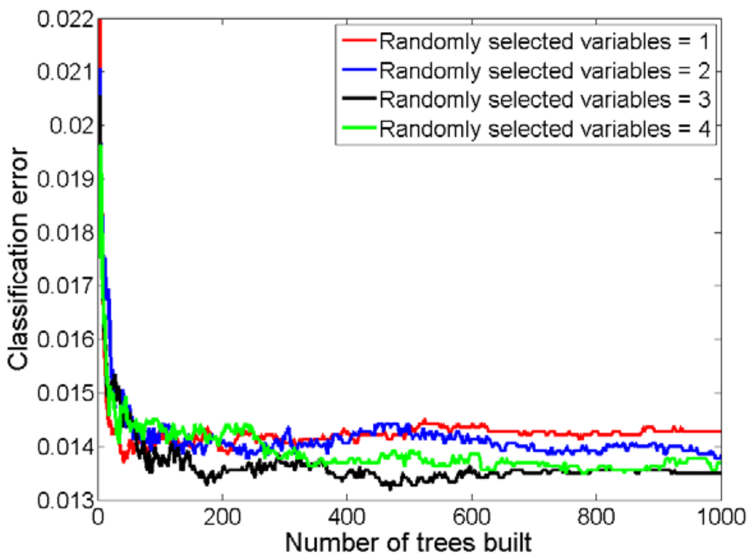
Location of the 43 data points and the 7 groups created through a clustering procedure

- Here ramp forecasts are made for Group 3, with NWP- forecasts and measurements from the three past hours from Groups 2, 3 and 4 as input data

Results – Model building

- The best model is found to apply six variables (exceeding the dashed line)
- These are the wind power measurements for the time of the forecast and past two hours, and wind power forecasts for the three next hours





- 600 trees is needed for the classification error to reach a stable level
- Randomly selecting 3 of the 6 variables in the final model for building each tree gives the lowest classification error

Results

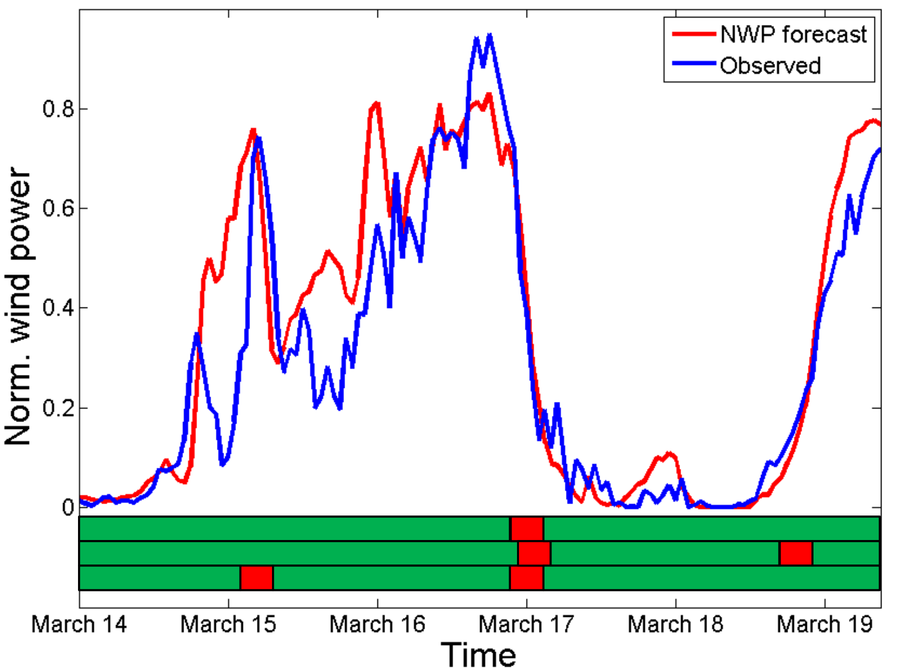
- Figure shows forecasted and observed wind power output for five days in March 2009

- The red/green lines show the ramp forecasts (red indicates ramps)
  - Random forests model (upper)
  - Raw NWP ramp forecast (middle)
  - Observed ramps (lower)

- Both ramp forecasts misses the first up-ramp

- Both ramp forecasts catches the down-ramp, but the NWP ramp forecast with one hour time delay

- The NWP ramp forecasts a false up-ramp late March 18.



Results

- The fraction of forecastsed ramps that actually occur are increased from **15 %** for the NWP-forecast to **53 %** for the random forests forecast

- The raw NWP ramp forecast predicts **3 times** as many ramps as is observed. With the random forest ramp forecast this is reduced to a under-prediction of **30 %**.

- The improvement over a random forecasting strategy (Hanssen and Kuipers skill score) is increased from **15 %** for the raw NWP forecast to **54 %** for the random forests forecast

- No improvements in the ramp forecast accuracy are found from the inclusion of wind power forecasts and measurements from neighboring groups

Conclusions and outlook

- The random forests model give an increase in ramp forecast accuracy, but suffers from a tendency of under-forecasting

- When classifying rare events, the random forests procedure is vulnerable to hedging – over-classification of the majority category
  - Here the «true ramps» constitute approx. 1.5 % of the total observations.
  - Techniques to reduce this problem through up/down-sampling of the training data are found in the litterature

- Available evaluation metrics are not ideal. Would be benificial with a metric that gives uneven weight to differen kinds of forecast errors
  - Economic gain from using the model would be an option
  - Not a trivial task to assign precise values (time-dependency)
  - Could not be used to optimize the model

Thank you for the attention!