Data-driven Solar Power Forecasting

Caroline Persson - s093051@student.dtu.dk
Peder Bacher - pbac@dtu.dk
Dynamical Systems Group, DTU Compute, Technical University of Denmark

What, Why, How?

What?
Multi-site forecasting of solar power production from 42 rooftop installations 1-6 hours ahead.

Why?
Better integration of weather dependent renewables into the electricity grid. More efficient utilization of renewable energy when it is available.

How?
A machine learning approach using Gradient Boosted Regression Trees (GBRT)

Forecasting Models

Persistence & Climatology

$$\hat{Y}_{t+k|t} = Y_t \tag{1}$$

$$\hat{Y}_{t+k|t} = \frac{1}{p}\sum_{n=0}^{p-1} Y_{t-n} \tag{2}$$

Recursive Auto-Regressive (AR) Model

A recursive AR model with forgetting factor γ.

$$\hat{\tau}_{t+k|t} = c + \alpha_t \tau_t + \beta_t \tau_{t-24+k}, \tag{3}$$

where τ is normalized power output.

Single-site GBRT

A GBRT for each individual PV installation using lagged variables and NWP data as input.

$$f_M(x) = \sum_{m=1}^{M} T(x; \Theta_m) \tag{4}$$

Multi-site GBRT

A GBRT using lagged variables, NWP data and location of PV installation as input resulting in a multi-site model. The model structure is equivalent to the single-site GBRT, but x constitutes data from all PV stations identified by their location (longitude and latitude).

Further Work

Point forecasts do not communicate the uncertainty of the forecast values. An obvious continuation of this project is to investigate the use of regression trees for probabilistic forecasting. The figure below shows an example of quantile regression using a GBRT model for a single PV installation for a forecast horizon of one hour.

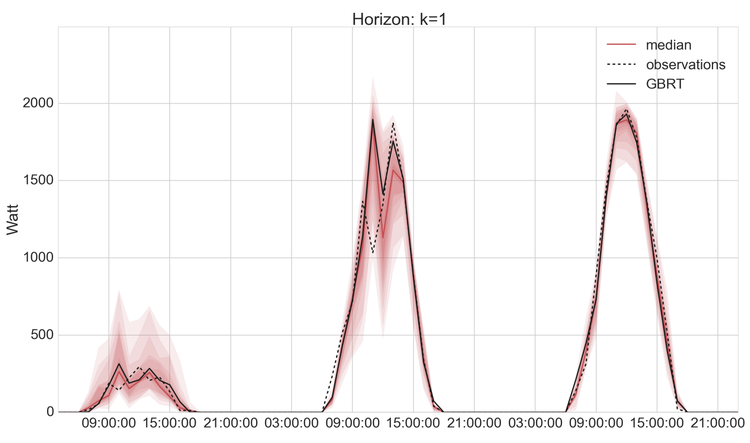

Figure 1: Probabilistic forecast of solar power using GBRT for quantile regression.

Furthermore a more exhaustive feature exploration and combinations of different models might result in an overall better performance.

Results

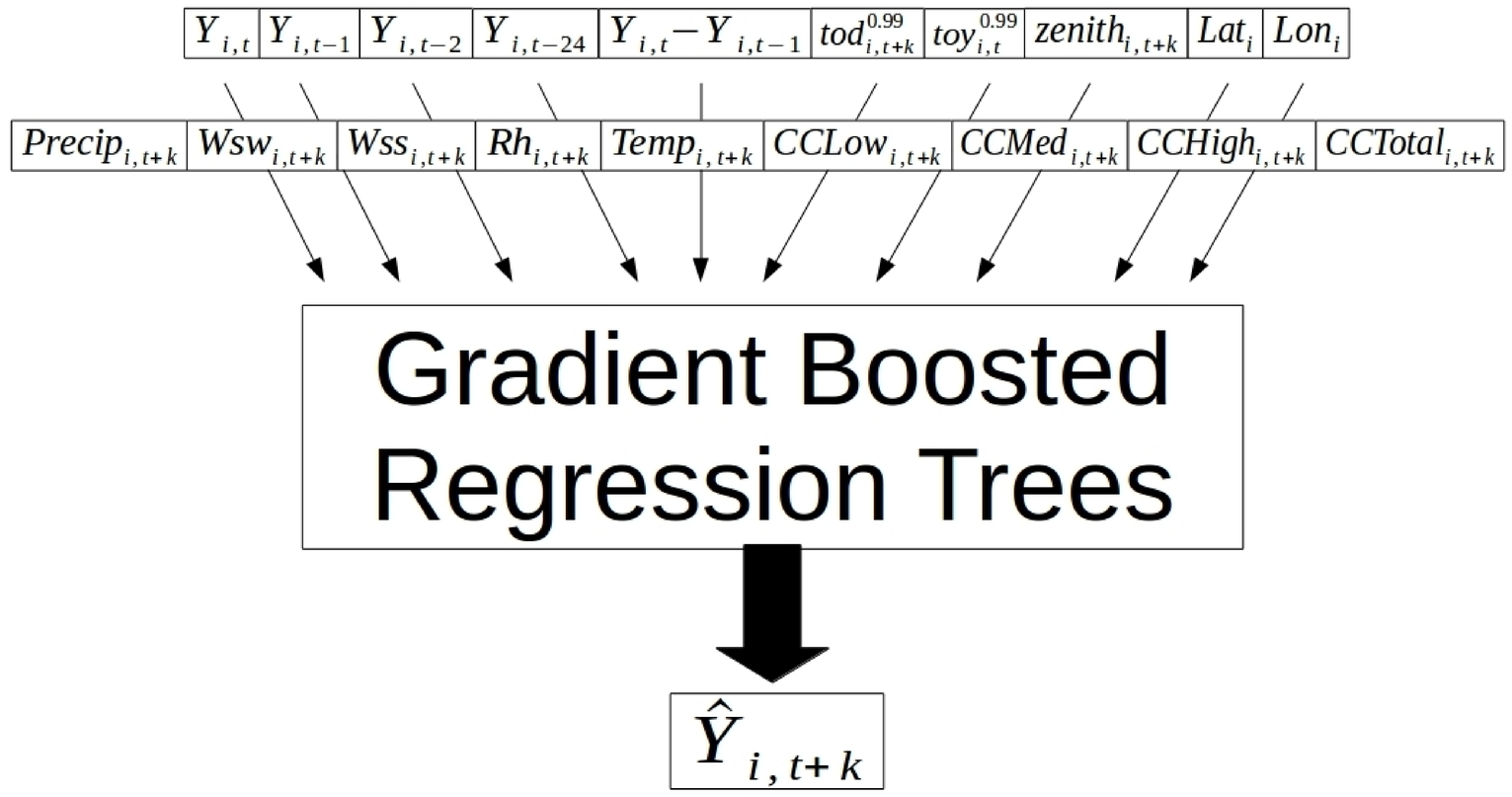

Figure 2: Outline of the model approach. 19 input variables are fitted to predict the power output $\hat{Y}_{i,t+k}$ using gradient boosted regression trees.

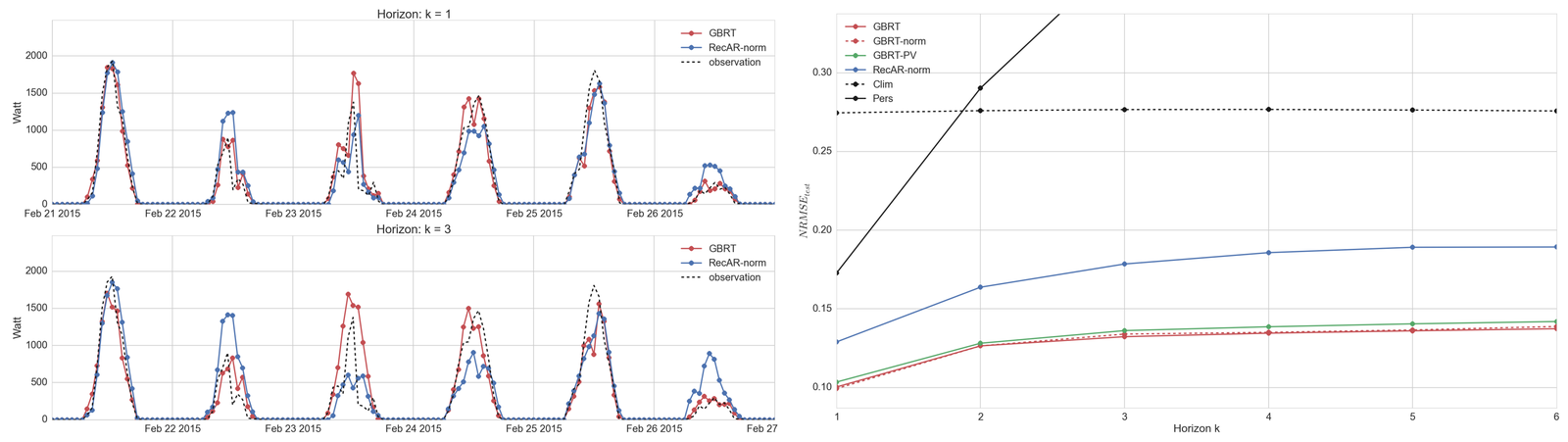

Figure 3: [left] One and three hours ahead predictions comparing multi-site GBRT and RecAR; [right] $NRMSE_{test}$ weighted over all 42 PV stations for forecast horizon of $1-6$ hours ahead.

Input Data

For the multi-site GBRT model 19 explanantory variables are included:

- Endogenous power data (7)
- Weather forecasts (9)
- System data (3)

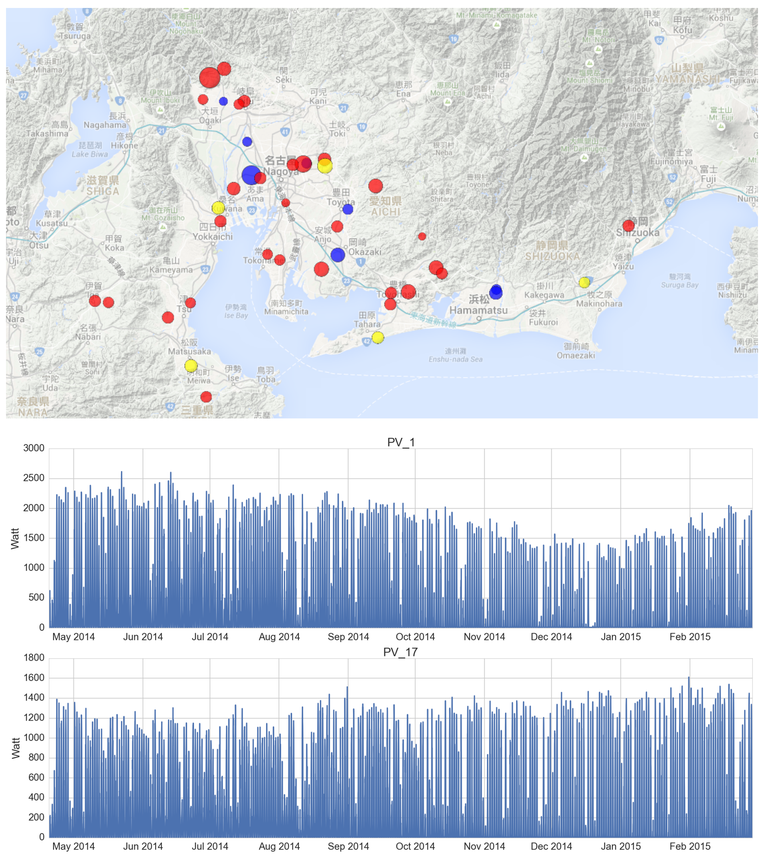

Figure 4: [up] Location of the 42 PV installations; [down] Yearly pattern of two installations.

Feature Analysis

Top 5 most important features of the multi-site GBRT. Notice how recent observations are important on short lead times ($k = 1$) and weather forecasts are more important for longer lead times ($k = 6$).

Ranking	$k = 1$	$k = 6$
1	Y_t	$\text{tod}^{0.99}$
2	$\text{tod}^{0.99}$	CCMid
3	$Y_t - Y_{t-1}$	zenith
4	zenith	Y_{t-24+k}
5	CCLow	CCLow

Table 1: Feature importance for multi-site GBRT for a forecast horizon of one and six hours ahead.

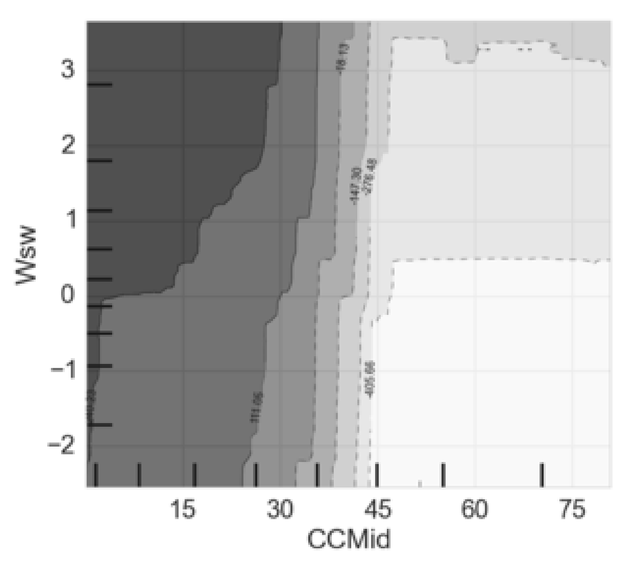

Figure 5: Partial dependence plot of medium cloud cover and western wind.