Big Data Models for Forecasting of Solar Power

C. Persson

DTU Compute, Technical University of Denmark

PURPOSE

This project presents a new approach for spatio-temporal solar power forecasting using machine learning algorithms suited for handling large amounts of data.

THEORY AND METHODS

Using highly non-linear regression trees it's possible to incorporate both location of a particular PV-installation and desired forecast horizon into one model, the model presented thus represents a multi-site, multi-horizon model. In this context, a tree model is trained on historical power data and associated meteorological forecasts to learn the patterns and interdependencies of future values of power production (at a given time and location) and the mentioned explanatory variables.
A gradient boosting algorithm is used for iteratively training of simple regressors (small decision trees), which end up constituting an ensemble of regression trees referred to as a gradient boosting tree.

RESULTS

Figure 1 shows an comparison of the performance of the gradient boosting tree (gbt) and three benchmark models; a persistence (pers), auto-regressive (linear) and adaptive auto-regressive (rec) model. Common for each model is that for a horizon of 1 hour mainly endogenous inputs are used (historical observations), where for longer forecast horizons exogenous inputs such as meteorological variables are used in the model estimation.

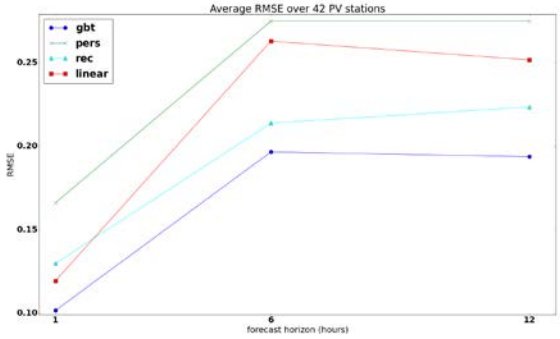

Figure 1: Model comparison of the average root mean square error of 42 PV-installations for three different forecast horizons (1, 6 and 12 hours)

CONCLUSION

The gradient boosting tree outperforms the benchmark models on different forecast horizons. Moreover, the model has the benefit of incorporating temporal and spatial information, such that one model represent a greater geographic area, and thus overcome the need of having one model for each time horizon and each PV-installation.