Universiteit Utrecht

Faculty of Geosciences
Copernicus Institute of Sustainable Development
Energy & Resources

Applied Computational Sciences (ACOS) symposium 2019

Day-ahead solar power forecasting with machine learning

L. Visser, T. AlSkaif and W. van Sark

Energy & Resources, Copernicus Institute for Sustainable Development, Utrecht University (UU), Utrecht, The Netherlands

Introduction

The increasing penetration of distributed PV-systems form a threat to reliable grid operation. PV-systems impede load balancing due to variable power production.

The development of highly accurate forecasting techniques is essential to support a high PV penetration rate in the electricity grid.

Methods

This research examines the performance of different models that predict day-ahead power production of PV-systems. The forecasts are based on historic power production and weather forecasts. The models considered are:

- Smart persistence (SP)
- Lasso Regression (LASSO)
- Random Forest (RF)
- Gradient Boosting (GB)
- Linear Support Vector Machine (L-SVM)
- Kernel Support Vector Machine (K-SVM)
- Multi-variate linear regression (MLR)
- Feed forward neural network (FNN)

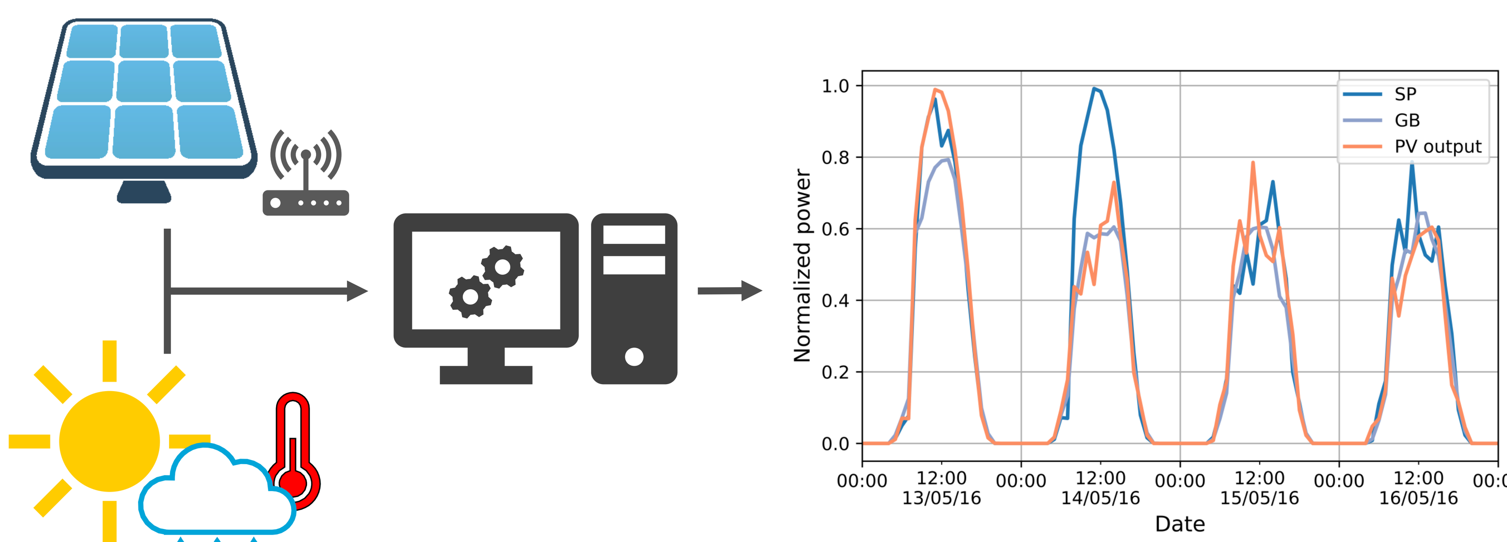

Figure 1: Flow diagram of methods.

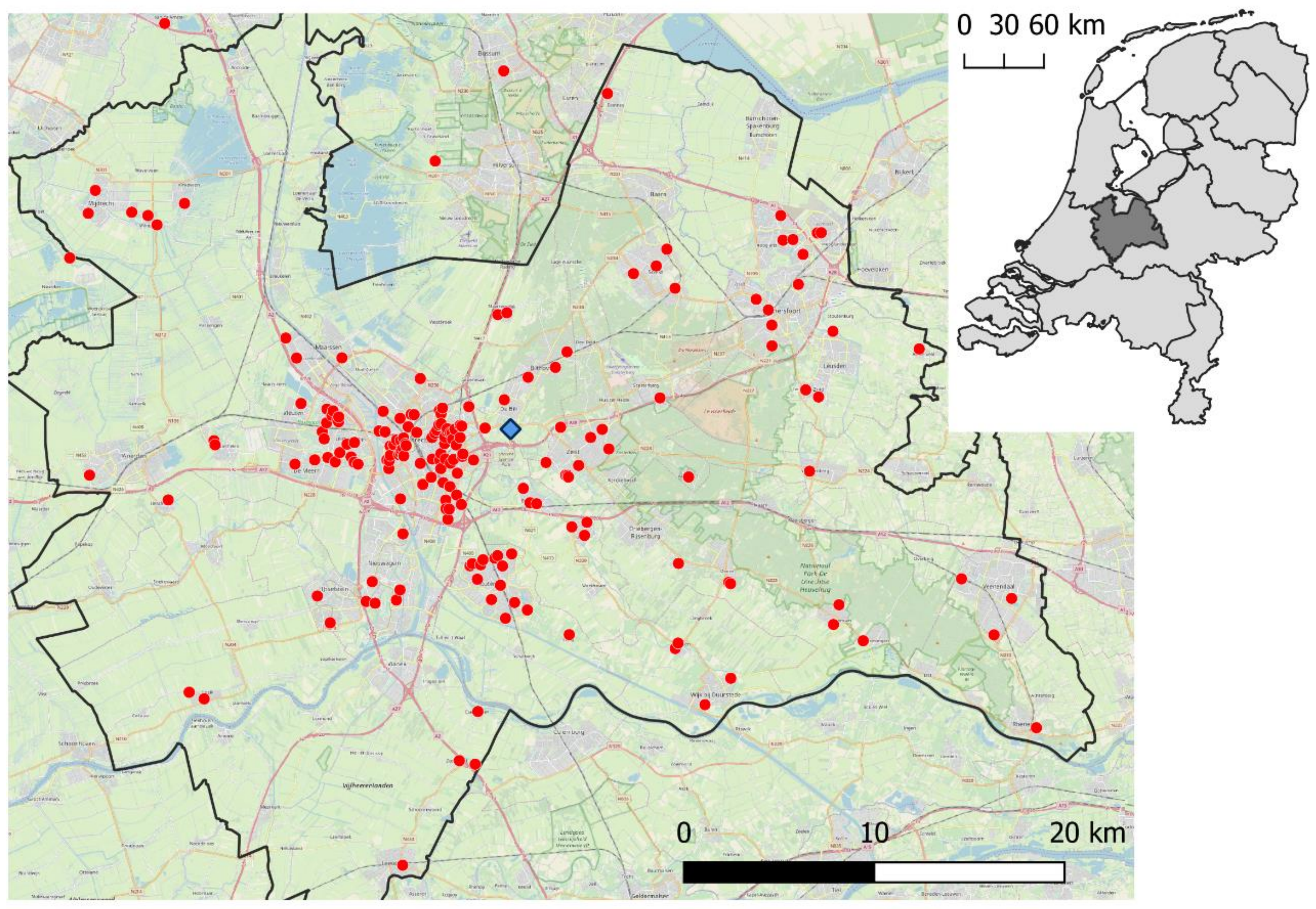

Figure 2: Distribution of rooftop PV-systems in Utrecht.

Data

The PV production data is collected from 152 rooftop PV-system in Utrecht, the Netherlands (Figure 2). Weather forecasts are collected from the ECMWF. Variables include the cloud cover, solar irradiance, temperature, pressure, windspeed and direction, All data is collected on an hourly basis for the period February 2014 until February 2017.

Results

Single PV-system forecasting:

The boxplots in figure 3 show that RF and GB outperform the other ML models in terms of the MAE, RMSE and Skill Score.

The spread in the boxplots indicate that for each model the forecast accuracy obtained can deviate significantly per site.

All forecasting models except for SP and RF have a negative bias. This implies that the PV power forecast generated by the models is structurally too high.

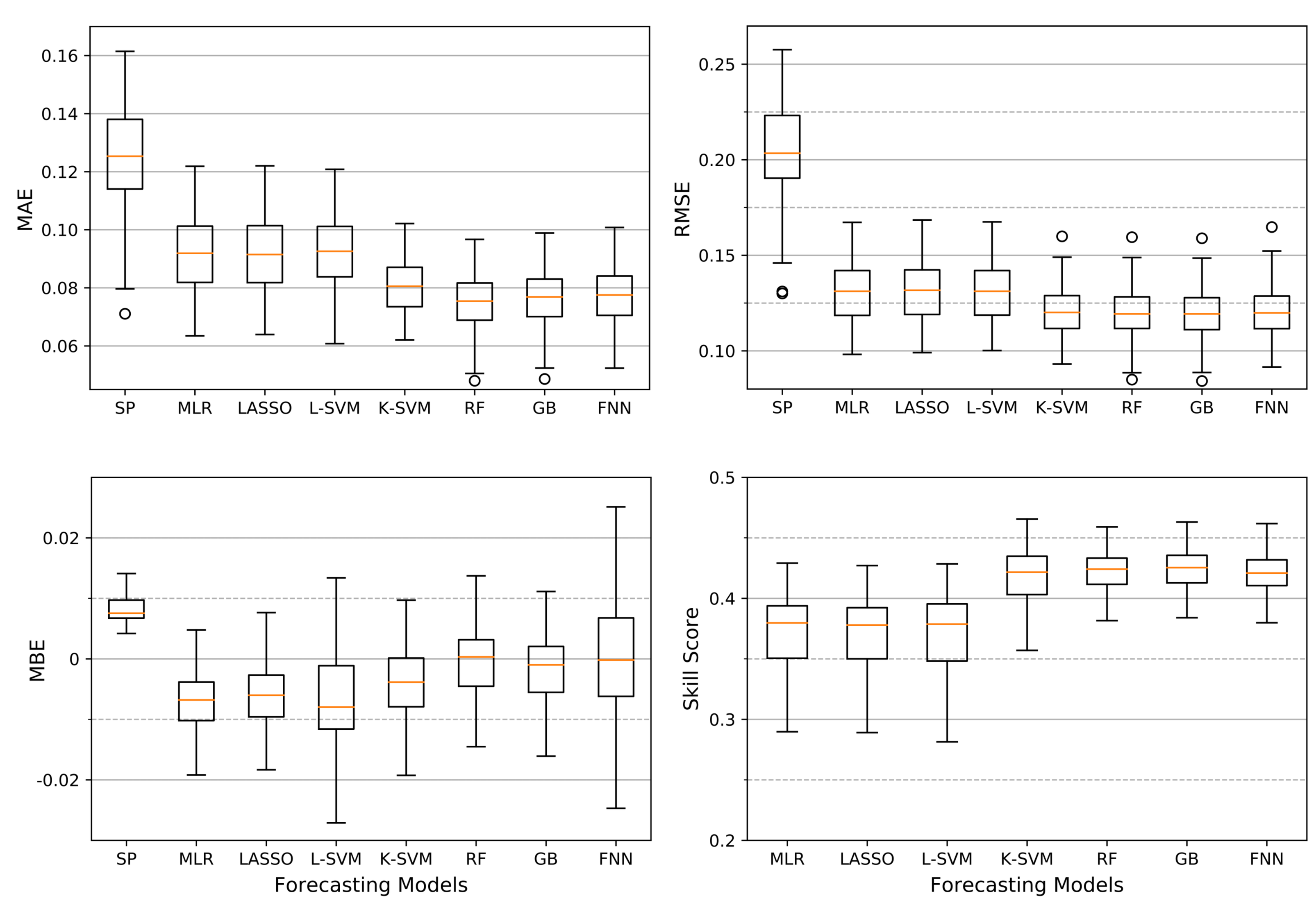

Figure 3: Boxplots containing the performance of the day-ahead forecasting models for 152 PV-systems in terms of the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Bias Error (MBE) and Skill Score (based on the RMSE).

Aggregated PV-system forecasting

Figure 4 and table 1 demonstrate that the forecast accuracy of all models improve as the number of sites considered increase.

This improvement is most significant when the number of systems considered is relatively low.

As the amount of aggregated systems increases K-SVM proves to become competitive to RF. RF and K-SVM outperform all other models when 150 PV-systems are considered.

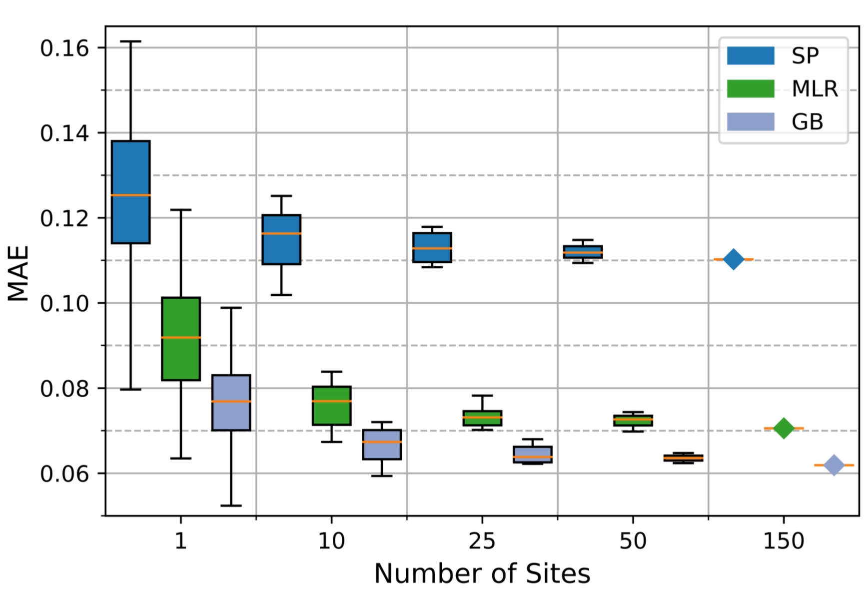

Figure 4: Boxplot of the MAE obtained by SP, MLR and GB for different levels of aggregated systems considered in the forecast models.

	Single site		10-sites		25-sites		50-sites		150-sites	
Models	MAE (std)	RMSE (std)	MAE (std)	RMSE (std)	MAE (std)	RMSE (std)	MAE (std)	RMSE (std)	MAE	RMSE
SP	12.5 (1.83)	20.4 (2.64)	11.5 (0.69)	18.2 (0.97)	11.3 (0.43)	17.9 (0.49)	11.2 (0.27)	17.7 (0.30)	11.0	17.4
MLR	9.16 (1.37)	13.1 (1.70)	7.59 (0.52)	10.8 (0.59)	7.34 (0.30)	10.5 (0.33)	7.23 (0.23)	10.3 (0.21)	7.06	10.0
LASSO	9.15 (1.37)	13.1 (1.71)	7.58 (0.53)	10.9 (0.60)	7.34 (0.30)	10.5 (0.34)	7.23 (0.22)	10.4 (0.20)	7.06	10.1
L-SVM	9.23 (1.20)	13.1 (1.56)	7.72 (0.49)	10.8 (0.58)	7.49 (0.28)	10.5 (0.32)	7.38 (0.20)	10.3 (0.18)	7.20	10.0
K-SVM	8.02 (0.86)	12.1 (1.23)	6.80 (0.41)	10.1 (0.51)	6.58 (0.22)	9.80 (0.27)	6.45 (0.13)	9.63 (0.11)	6.29	9.31
RF	7.48 (1.01)	11.9 (1.41)	6.60 (0.39)	10.3 (0.53)	6.40 (0.25)	10.0 (0.33)	6.28 (0.14)	9.76 (0.15)	6.09	9.43
GB	7.63 (1.02)	11.9 (1.40)	6.67 (0.39)	10.2 (0.52)	6.45 (0.24)	9.87 (0.33)	6.36 (0.12)	9.70 (0.13)	6.19	9.41
FNN	7.71 (1.01)	12.0 (1.41)	6.79 (0.43)	10.3 (0.53)	6.52 (0.23)	10.0 (0.27)	6.34 (0.11)	9.70 (0.09)	6.30	9.38

Table 1: The MAE, RMSE and corresponding standard deviations in percentages for different numbers of aggregated PV-systems considered in day-ahead forecasting by each model.

Conclusion

The results show that all ML models considered perform better than the reference model, SP. Moreover, the more sophisticated models (K-SVM, RF, GB and FNN) achieve better results compared to the linear models. RF is found to outperform all other models on a single PV-system level, while RF and K-SVM perform best when PV-systems are aggregated.

References

[1] Elsinga, B., & van Sark, W. (2015). Spatial power fluctuation correlations in urban rooftop photovoltaic systems. Progress in Photovoltaics: Research and Applications, 23(10), 1390-1397.

[2] L. Visser, T. Alskaif and W. Van Sark, "Benchmark analysis of day-ahead solar power forecasting techniques using weather predictions," IEEE 46th Photovoltaic Specialists Conference (PVSC), pp. 1-6, 2019.

Lennard Visser
PhD Candidate, Energy & Resources, Copernicus Institute of Sustainable Development, Utrecht University
Email: l.r.visser@uu.nl