Benchmark analysis of day-ahead solar power forecasting techniques using weather predictions

Lennard Visser, Copernicus Institute Utrecht University, the Netherlands

ISES Webinar August 27, 2020
Introduction

• Why day-ahead forecasting?
  • Most electricity traded in day-ahead market
  • Schedule dispatch of power generation

• Spot market trading:

- Lead time: 12 hours
- Horizon: 24 hours
Solar Forecasting Techniques

![Graph showing the relationship between Spatial Resolution (km) and Temporal horizon (hr). The x-axis represents Temporal horizon (hr) ranging from 0.001 to 1000, and the y-axis represents Spatial Resolution (km) ranging from 0.001 to 1000. The graph includes grid lines for easier visualization.](image)
Solar Forecasting Techniques

- Temporal horizon (hr)
- Spatial Resolution (km)

Techniques:
- Numerical Weather Prediction
- Statistical learning
- Satellite Imaging
- All Sky Imaging
Contribution

• Comparison of models that utilize NWP to forecast the PV power output
• Examining the value of aggregating PV systems for forecasting
PV-systems in Utrecht

UPP-network

• 200 PV-systems
• Utrecht (NL)
• 38 x 54 km²
• 2013 - 2017

Legend

● PV-system
◆ Weather station
Methods: Input Variables

ECMWF weather prediction

Additional exogeneous data

- Mean sea level pressure
- Surface temperature at 2m
- Dewpoint temperature at 2m
- Zonal wind vector at 10m
- Meridional wind vector at 10m
- Surface solar radiation downwards
- Cloud cover at low, mid and high altitude
- Total precipitation

- Clear sky irradiance
- Solar zenith angle
- Month
- Hour
Methods: Process Variables

ECMWF weather prediction archive

Variables

Additional exogeneous data

Standardize:

\[ v_{i,t} = \frac{x_{i,t} - \mu_i}{\sigma_i} \]
Methods: Forecasting Models

ECMWF weather prediction archive → Variables → Models

Additional exogeneous data

0) Smart Persistence (SP)
1) Multi-variate Linear Regression (MLR)
2) LASSO Regression (LASSO)
3) Linear Support Vector Machine (L-SVM)
4) Kernel Support Vector Machine (K-SVM)
5) Random Forests regression (RF)
6) Gradient Boosting regression (GB)
7) Feed-forward Neural Network (FNN)
Methods: Train Models

Training period 02/2014 - 02/2016

ECMWF weather prediction archive

Variables

Models

Additional exogeneous data

PV production measurements

PV measurements of 152 PV-systems for period 02/2014 - 02/2017
Methods: Train Models

PV production measurements

Training period
02/2014 - 02/2016

ECMWF weather prediction archive

Variables

Models

Pre-processing

PV measurements of 152 PV-systems for period 02/2014 - 02/2017

Normalize:

\[ y_{p,t} = \frac{y_{m,t}}{y_m} \]
Evaluate the performance of the forecast models with:

1) Mean Absolute Error (MAE):
   \[ MAE = \frac{1}{n} \sum_{t=1}^{n} |y_{p,t} - y_{m,t}| \]

2) Root Mean Square Error:
   \[ RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_{p,t} - y_{m,t})^2} \]

3) Skill Score:
   \[ Skill\ Score = 1 - \frac{RMSE_{for}}{RMSE_{ref}} \]

Methods: Forecast & Evaluation

ECMWF weather prediction archive

Variables \rightarrow Models \rightarrow Post-processing \rightarrow Output

Validation period 02/2016 - 02/2017

Remove outliers

Additional exogeneous data
Results: Time-series Forecast

- Example of forecast
Results: Forecasting single PV-system

- All statistical models perform better than SP
- The more sophisticated statistical models outperform the linear models
- Best performance RF and GB

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>Skill Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-SVM</td>
<td>8.04%</td>
<td>40.1%</td>
</tr>
<tr>
<td>RF</td>
<td>7.48%</td>
<td>41.2%</td>
</tr>
<tr>
<td>GB</td>
<td>7.63%</td>
<td>41.4%</td>
</tr>
<tr>
<td>FNN</td>
<td>7.71%</td>
<td>41.1%</td>
</tr>
</tbody>
</table>
Results: Forecasting multiple PV-systems (1)

- The performance of all models improve as the number of sites increase:
  - Statistical models (20-25%)
  - SP (10%)
- The rate of improvement decrease as the number of sites increase
- Deviation of forecast errors decrease as the number of sites increase
Forecasting multiple PV-systems

- All statistical models perform better than SP
- The more sophisticated models outperform the linear models
- RF best performance in terms of MAE
- K-SVM performs best in terms of the Skill Score

<table>
<thead>
<tr>
<th>Models</th>
<th>MAE (%)</th>
<th>Skill Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP</td>
<td>11.0</td>
<td>-</td>
</tr>
<tr>
<td>MLR</td>
<td>7.06</td>
<td>42.5</td>
</tr>
<tr>
<td>LASSO</td>
<td>7.06</td>
<td>42.0</td>
</tr>
<tr>
<td>L-SVM</td>
<td>7.20</td>
<td>42.5</td>
</tr>
<tr>
<td>K-SVM</td>
<td>6.29</td>
<td>46.5</td>
</tr>
<tr>
<td>RF</td>
<td>6.09</td>
<td>45.8</td>
</tr>
<tr>
<td>GB</td>
<td>6.19</td>
<td>45.9</td>
</tr>
<tr>
<td>FNN</td>
<td>6.30</td>
<td>46.1</td>
</tr>
</tbody>
</table>

MAE and Skill Score for 150 PV-systems
Conclusions

• Comparison of statistical PV power forecasting models

• Single PV-system
  ▪ Sophisticated models outperform the linear models
  ▪ RF and GB outperform the other models

• Aggregated PV-systems
  ▪ Benefits all forecasting models
  ▪ Reduces the difference in errors among the statistical models