Performance evaluation of PV power predictive models for real-time monitoring

Andreas Livera\textsuperscript{1}, Marios Theristis\textsuperscript{1}, George Makrides\textsuperscript{1}, Juergen Sutterlueti\textsuperscript{2}, Steve Ransome\textsuperscript{3} and George E. Georghiou\textsuperscript{1}

\textsuperscript{1}PV Technology Laboratory, University of Cyprus, Nicosia, Cyprus
\textsuperscript{2}Gantner Instruments GmbH, Schruns, Austria
\textsuperscript{3}Steve Ransome Consulting Ltd, Kingston upon Thames, UK
Acknowledgement

**Specific Objective:** Development of an innovative condition monitoring platform for proactive and reactive O&M with enhanced data analytic functionalities

Advanced baseline condition monitoring solution to ensure operational quality and optimise energy production

**Partners:** GI and UCY  
**Project:** Innovative Performance Monitoring System for Improved Reliability and Optimized Levelized Cost of Electricity IPERMON [Solar-ERA.net project]  
**Budget:** €400,000  
**Duration:** 36 Months (April 2016 – Sept 2019)  
**Weblink:** [http://www.pvtechnology.ucy.ac.cy/projects/ipermont/](http://www.pvtechnology.ucy.ac.cy/projects/ipermont/)
Introduction

• Accurate output power prediction is crucial for PV performance assessment
• Predictive models are required for data-analytic features of advanced PV monitoring systems

Data-analytic features

• System health state
• Failure diagnosis
Objective

Development of an optimized location- and technology-independent predictive modeling methodology at **minimum requirements**

**Input**
- Features
- Dataset split method
- Dataset split partition
- Filtering stages
- Weather conditions

**Output**
Methodology – Approach

1. Get Data
2. Train Model
3. Clean, Prepare & Manipulate Data
4. Test Data
5. Improve
Methodology – Experimental setup

• Recording of meteorological and PV operational measurements (IEC 61724)
• Measurement resolution 1-sec and recording intervals 1-, 15-, 30- and 60-min
Methodology – Data quality routines (DQRs)

- Identification of repetitive data and duplicates
- Identification of missing or erroneous data, outliers and outages
- Correction of erroneous/missing data through data imputation techniques
Methodology – Data quality routines (DQRs)

1. **Identification of duplicates**
   - Check timestamp measurements against known timestamp series
   - Check for row measurement duplicates

2. **Identification of missing data**
   - Search for NAN values from the dataset

3. **Identification of erroneous data**
   - Set threshold ranges for:
     - 0 < Irradiance < 1300 (W/m²)
     - 0 < DC Power < STC power x 1.3
     - 0 < DC Voltage < STC Voltage x 1.1
     - 0 < DC Current < STC Current x 1.25

4. **Data filtering**
   - Night time effects can be removed (e.g. Irradiance < 50 W/m²)

5. **Data correction**
   - Data imputation techniques for handling erroneous or missing data
Methodology – Predictive model selection

**Empirical**

**MECHANISTIC PERFORMANCE MODEL ‘MPM’**

\[ PR = \left( \frac{P_{\text{MEAS}}}{P_{\text{NOM}}/G_1} \right) = C_1 + C_2 * T_{\text{mod}} + C_3 * \log_a(G_i) + C_4 * G_i + C_5 * W_S \]

- **P TOLERANCE**: %
- **GAMMA**: %/K
- **LLEC**: %/STC
- **RS**: %/STC
- **WIND**: %/(ms⁻¹)

**Machine Learning**

Feed-Forward Neural Network (FFNN)
Methodology – Train model and test data

Dataset (1 year of hourly historical actual data)

- **Measured Inputs**
  - $G_i$
  - $T_{mod}$
  - RH
  - WS
  - $W_{alpha}$
  - AzS
  - AlS
  - Pmp

- **Calculated Inputs**

- **Output**

**Dataset split method**
- Continuous
- Random

**Dataset split partition**
- 70:30% train and test set
- 30:30% train and test set
- 10:30% train and test set
Results – Input features (Machine Learning)

- Machine learning model with measured and calculated features

<table>
<thead>
<tr>
<th>Inputs</th>
<th>2 Inputs</th>
<th>4 Inputs</th>
<th>7 Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>nRMSE</td>
<td>1.13%</td>
<td>1.12%</td>
<td>0.91%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.93%</td>
<td></td>
</tr>
</tbody>
</table>

Best performance FFNN

Random 70:30% nRMSE 1.18%
Continuous 70:30% nRMSE 1.33%
UCY OTF
Results – Output features (Machine Learning)

- Machine learning model with measured and calculated features

7 Inputs – $PR$ output  
7 Inputs – $P_{mp}$ output

**Random** 70:30% | **Continuous** 70:30%  
**nRMSE 1.30%** | **nRMSE 1.33%**  
**nRMSE 0.91%** | **nRMSE 0.93%**

Best performance FFNN

Random – Recommended dataset split method
Results – Input features (Mechanistic)

• Mechanistic model with measured and meaningful, orthogonal, robust and normalized features

\[
PR = \left( \frac{P_{\text{MEAS}}}{P_{\text{Nom}}}/G_i \right) = C_1 + C_2 T_{\text{mod}} + C_3 \log_{10}(G_i) + C_4 G_i + C_5 WS
\]

Inputs:
• Module temperature \((T_{\text{mod}})\)
• Global irradiance \((G_i)\)
• Wind speed \((WS)\)

Requirements for optimal devised model:
• Irradiance Filter \((G_i > 100 \, W/m^2)\)
• Time Filter \((08:00 \leq \text{Time} \leq 17:00)\)
## Results – Input features (Mechanistic)

### Random 70:30% - GI OTF

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 (%)</td>
<td>114.09</td>
</tr>
<tr>
<td>C2 (%/K)</td>
<td>-0.39</td>
</tr>
<tr>
<td>C3 (%)</td>
<td>25.05</td>
</tr>
<tr>
<td>C4 (%)</td>
<td>-17.87</td>
</tr>
<tr>
<td>C5 (%/ms(^{-1}))</td>
<td>0.08</td>
</tr>
</tbody>
</table>

The formula for PR is:

\[
PR = \left( \frac{P_{\text{MEAS}}}{P_{\text{NOM}}/G_i} \right) = C_1 + C_2 \cdot T_{\text{mod}} + C_3 \cdot \log(G_i) + C_4 \cdot G_i + C_5 \cdot \text{WS}
\]

**TOLERANCE**
- C1: Global Tolerance (%)
- C2: Global Tolerance (%/K)
- C3: Local Tolerance (@LIC)
- C4: Local Tolerance (@STC)
- C5: Local Tolerance (%/ms\(^{-1}\))

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>STC</td>
<td>Standard Test Conditions (1000 W/m(^2), 25°C)</td>
</tr>
<tr>
<td>NOCT</td>
<td>Nominal Operating Cell Temperature</td>
</tr>
<tr>
<td>GI OTF</td>
<td>Global Irradiance Operating Temperature Factor</td>
</tr>
<tr>
<td>PMAX</td>
<td>Maximum Power (W)</td>
</tr>
<tr>
<td>LLEC</td>
<td>Light Level Efficiency Curve</td>
</tr>
<tr>
<td>GAMMA</td>
<td>Gamma</td>
</tr>
<tr>
<td>Gi (kW/m(^2))</td>
<td>Irradiance (W/m(^2))</td>
</tr>
</tbody>
</table>

### Diagram

- **PRdC** vs. **Gi (kW/m\(^2\))**
- **Pmax**/**Tmod**
- **Gi**
- **Gamma**
- **LLEC**
- **STC**
- **NOCT**
- **Series Resistance**
- **thermal rise**
- **Actual/Nameplate Pmax**

Illustrating good PV Performance:
- Uniform vertical separation means gamma = constant
- Smooth behaviour at lowest and highest light levels
Results – Influence of filtering (Mechanistic)

Random 70:30% - GI OTF

MPM – Improved performance at high irradiance levels
Results – Influence of filtering (Mechanistic)

- Filtering at $G_I > 100 \, W/m^2$ (GI OTF)

  - $G_I > 100 \, W/m^2$
    - nRMSE 1.03%
    - MPM – Higher accuracy by applying irradiance filters (2.15% without any filter)

  - $G_I > 400 \, W/m^2$
    - nRMSE 0.88%

  - $G_I > 600 \, W/m^2$
    - nRMSE 0.87%
Results – Influence of filtering (Mechanistic)

- Filtering at $G_I > 100 \text{ W/m}^2$ (GI OTF)

72% of days exhibiting daily nRMSE accuracies below 1% independent of the type of day (clearness index)
Results – Influence of filtering (Machine Learning)

- Filtering at different irradiance levels (UCY OTF)

Without filter:
- nRMSE 0.91%

ML – Improved performance at increased data for training:
- $G_I > 100 \, W/m^2$
- nRMSE 1.31%

ML – Accuracy not improved by applying irradiance filter:
Results – Influence of filtering (Machine Learning)

- Filtering at \( G_I > 100 \, W/m^2 \) (UCY OTF)

  \[ G_I > 100 \, W/m^2 \]
  nRMSE 1.31%

  ML – Accuracy not improved by applying irradiance filter

  \[ G_I > 400 \, W/m^2 \]
  nRMSE 1.36%

  \[ G_I > 600 \, W/m^2 \]
  nRMSE 1.29%
Results – Influence of filtering (Machine Learning)

• Filtering at $G_I > 100 \text{ W/m}^2$ (UCY OTF)

62% of days exhibiting daily nRMSE accuracies below 1.3% independent of the type of day (clearness index)
Results – Dataset split partitions

- Training at different dataset split partitions (10, 30 and 70% of yearly data)

Random training - Accurate predictions for both models even at small amount of training data partitions

Continuous training – Seasonal errors

- FFNN
- MPM
## Summary

<table>
<thead>
<tr>
<th>Mechanistic</th>
<th>Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Simple implementation (low complexity)</td>
<td>• Higher complexity for implementation</td>
</tr>
<tr>
<td>• Robustness at high irradiance conditions</td>
<td>• Robust at all irradiance conditions only after training at different data combinations</td>
</tr>
<tr>
<td>• Irradiance filter improves prediction accuracy</td>
<td>• No data filtering requirements</td>
</tr>
<tr>
<td>• Robust model at low duration data set partitions</td>
<td>• Higher training data partitions yield more accurate predictions</td>
</tr>
<tr>
<td>• Useful, physically meaningful coefficients</td>
<td>• No direct usable coefficients</td>
</tr>
</tbody>
</table>
Conclusions

• The MPM and the FFNN predictive models were compared in terms of input/output features (model complexity), filtering criteria, dataset split method and partition

• Optimal models: 7 inputs parameter FFNN compared with 5 inputs parameter MPM

• Application of irradiance filter yielded higher predictive accuracy only for the MPM

• Random dataset split method is recommended for both models

• FFNN - Lowest nRMSE of 0.91% for a random 70:30% train/test set approach (UCY OTF)

• MPM - Lowest nRMSE of 1.12% for a random 10:30% train/test set approach (GI OTF)
Next steps...

- Influence of irradiance profiles classification and establishment of minimum requirements for daily weather classification
- Further improvement of the MPM (spectral and AOI corrections) for more accurate predictions
- Benchmarking on several PV systems installed at different locations

<table>
<thead>
<tr>
<th>Class</th>
<th>$k_d$</th>
<th>$POP_d$</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$k_d \geq 0.6$</td>
<td>$POP_d \geq 0.9$</td>
<td>High Quantity High Quality</td>
</tr>
<tr>
<td>2</td>
<td>$0.3 \leq k_d &lt; 0.6$</td>
<td>$POP_d \geq 0.9$</td>
<td>Medium Quantity and High Quality</td>
</tr>
<tr>
<td>3</td>
<td>$k_d &lt; 0.3$</td>
<td>$POP_d \geq 0.9$</td>
<td>Low Quantity High Quality</td>
</tr>
<tr>
<td>4</td>
<td>$k_d \geq 0.6$</td>
<td>$0.7 \leq POP_d &lt; 0.9$</td>
<td>High Quantity Medium Quality</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>8</td>
<td>$0.3 \leq k_d &lt; 0.6$</td>
<td>$0.5 \leq POP_d &lt; 0.7$</td>
<td>Medium Quantity Low Quality</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>$POP_d &lt; 0.5$</td>
<td>Very Low Quality</td>
</tr>
</tbody>
</table>
Thank you for your attention

Marios Theristis, PhD
PV Technology Laboratory
University of Cyprus
Email: theristis.marios@ucy.ac.cy

Website: www.pvtechnology.ucy.ac.cy
www.gi-cloud.io