Data processing and quality verification for improved photovoltaic performance

PV Performance Assessment

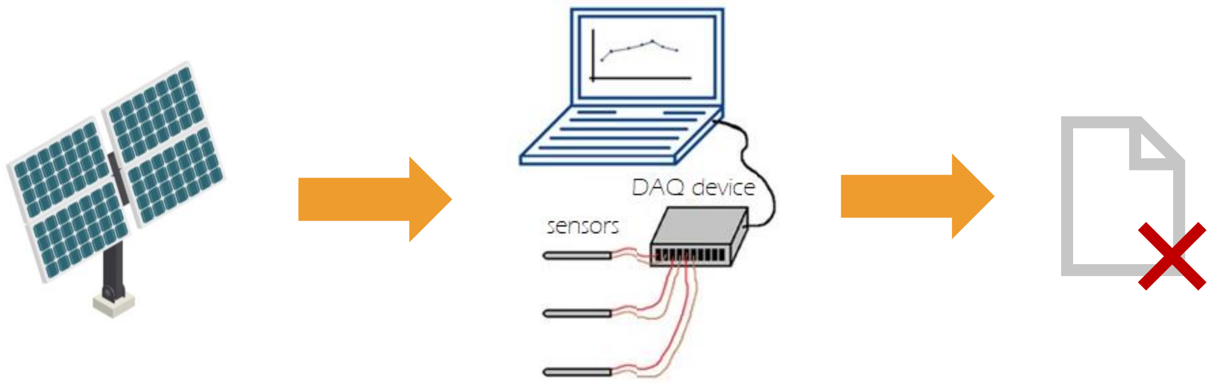
Andreas Livera
PV Technology Laboratory
FOSS Research Centre for Sustainable Energy
University of Cyprus

PEARL PV

FOSS
Research Centre for
Sustainable Energy

Why data quality matters

- Ensuring high-quality of data is crucial for the performance and reliability analytics of photovoltaic (PV) systems
- Field measurements commonly exhibit gaps, missing data and erroneous values caused by outages and component failures
- The processes applied to handle the acquired measurements can potentially introduce noticeable bias that obscures underlying PV performance analysis

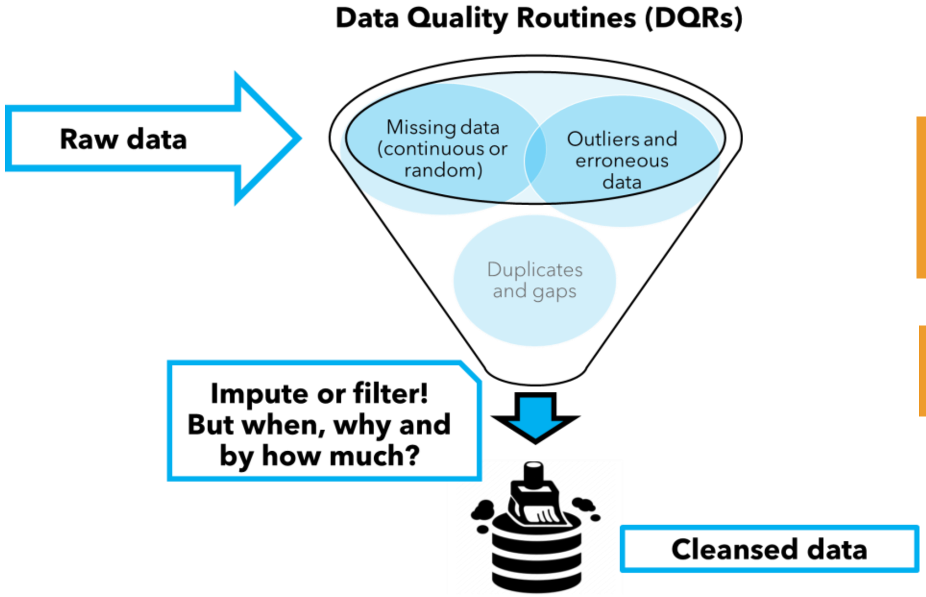


Background & Objective

- Existing standards (i.e. IEC 61724) and reports lack of specifications for handling invalid data (i.e. gaps, missing and erroneous values)

Specific Objective:

- To develop a complete and quantitative methodology of Data Quality Routines (DQRs) for data processing and quality checks



Provide a specific approach for handling invalid values and reconstructing PV datasets

Enable reproducible results in PV performance

Proposed DQRs

- Builds on IEC 61724 and other reports
- Comprises of 8 sequentially structured routines
- Main contribution on Step 6

Methodology - Approach

1. Experimental setup

- Test PV modules and systems in Arizona and Cyprus

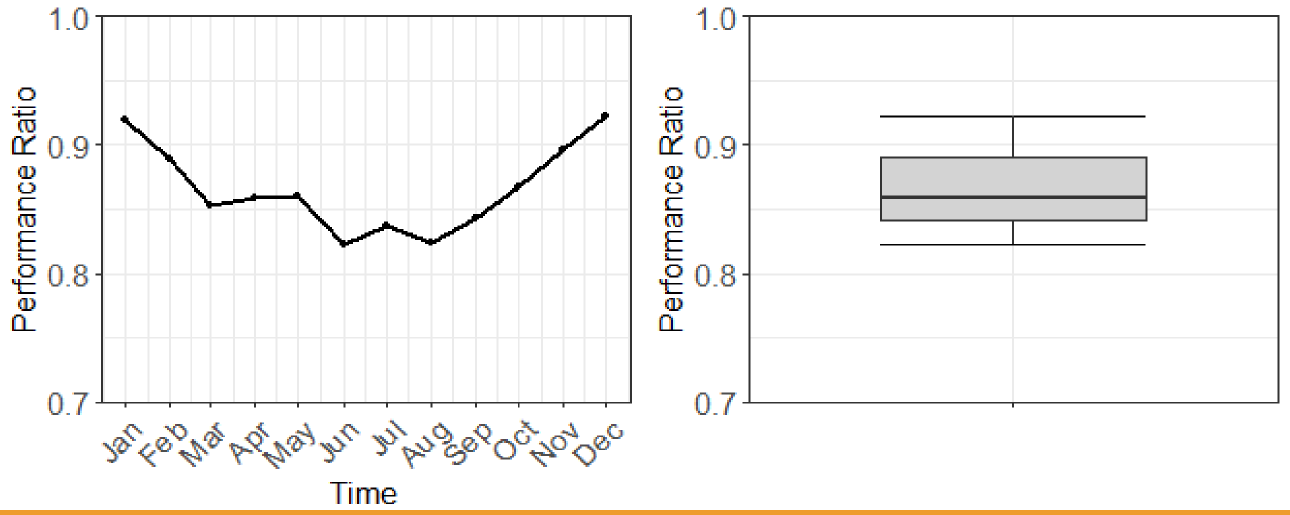
2. Data acquisition (DAQ) system

- Recording of electrical and meteo data
- Monthly DC Performance Ratio (PR) - Standard KPI in contractual agreements

3. Reference PV datasets

- 1-year and 5-year datasets (from a well-maintained PV module and system)

1-year reference PV dataset



Methodology - Approach

1. Experimental setup

- Test PV modules and systems in Arizona and Cyprus

2. Data acquisition (DAQ) system

- Recording of electrical and meteo data
- Monthly DC Performance Ratio (PR) - Standard KPI in contractual agreements

3. Reference PV datasets

- 1-year and 5-year datasets (from a well-maintained PV system)

4. Artificial invalid datasets

- Emulation of missing data into the reference dataset from 1% to 40% invalid data rate
- Process repeated 50 times for each invalid data rate (resulting in 2000 invalid datasets)

Methodology - Approach

5. Test cases
- Different test cases reflecting real PV monitoring scenarios of data loss

6. Invalid datasets reconstruction
- Invalid values treated by data deletion and inference techniques

7. Performance evaluation
- Comparison between reference and reconstructed datasets (PR and PLR metrics)
- Absolute percentage error (APE)

$$APE = \left|\frac{A_t - P_t}{A_t}\right|$$

where
A_t is the average monthly PR (or PLR) of the reference dataset
P_t is the average monthly PR (or PLR) of the 2000 invalid datasets

Methodology - 5. Test cases

- 6 different test cases were investigated reflecting real PV monitoring scenarios of data loss

Test Case	Missing measurements	Performance metric	Sequence of missing measurements
1	P_A	PR	Random
2	P_A	PR	Continuous
3	G_I	PR	Random
4	G_I	PR	Continuous
5	T_{mod}	PR_{TC}	Random
6	T_{mod}	PR_{TC}	Continuous

P_A is the array DC power
G_I is the in-plane irradiance
T_{mod} is the module temperature

Time	Variable X
t1	NA
t2	NA
t3	NA
t4	NA
…	…
tn	NA

Time	Variable X
t1	x1
t2	NA
t3	x3
t4	NA
…	…
tn	NA

Methodology - 6. Dataset reconstruction

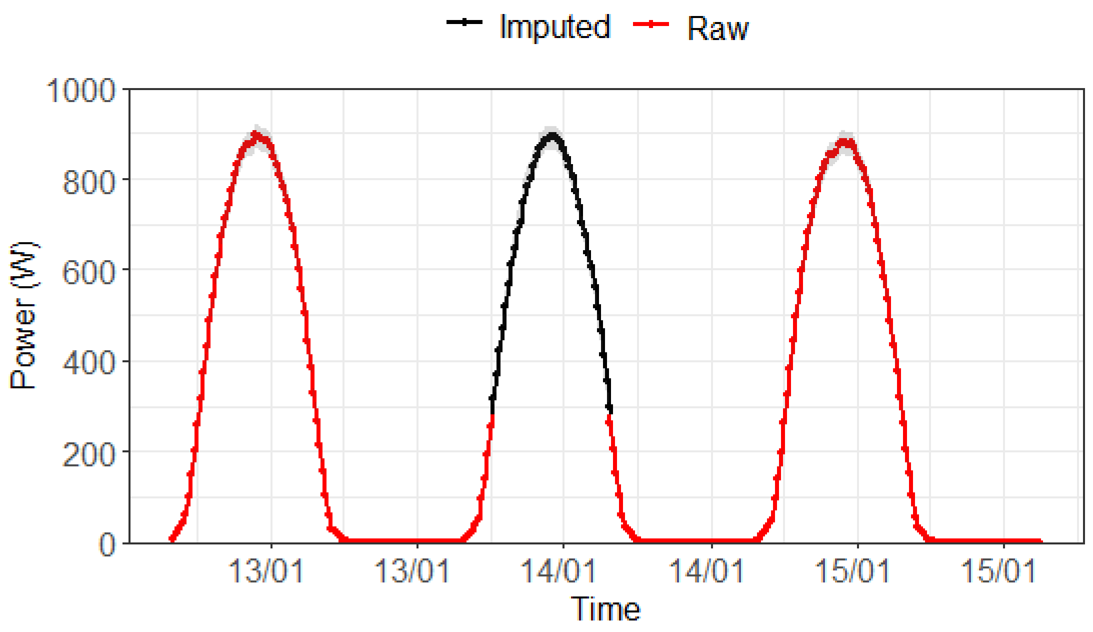
- Data deletion methods
  - “Row” deletion → all rows with at least one missing data point are deleted
  - “Value” deletion → only the missing values are removed

Timestamp	Variable X	Variable Y
t1	x1	y1
~~t2~~	~~NA~~	~~y2~~
t3	x3	~~NA~~
…	…	…
tn	xn	yn

Row deletion (row t2)
Value deletion (NA in row t3)

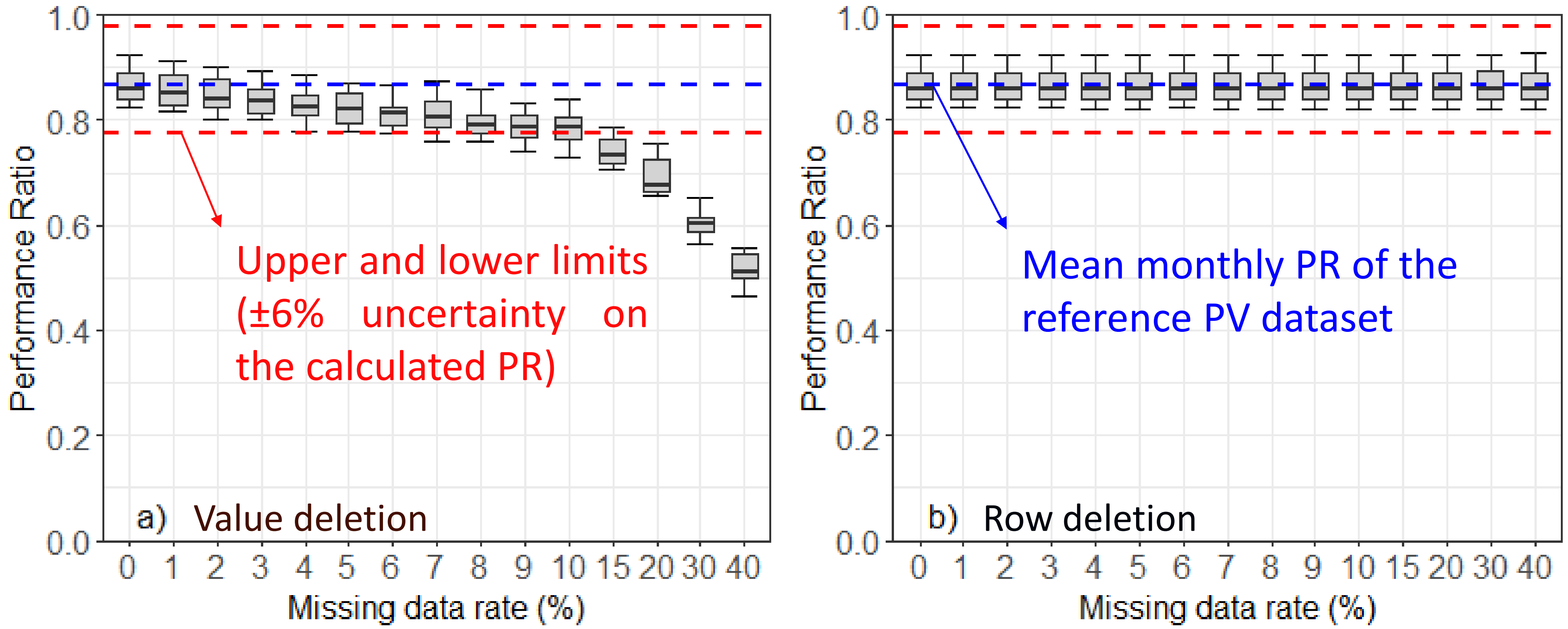
Methodology - 6. Dataset reconstruction

- Data inference methods (i.e. back-fill or impute missing data points)
  - Electrical models: Sandia Array Performance Model (SAPM), Sandia module temperature model (SMTM)
  - Multiple imputation: Predictive Mean Matching (PMM)
  - Univariate imputation: Random Forest (RF), bootstrap



Results - Data deletion methods

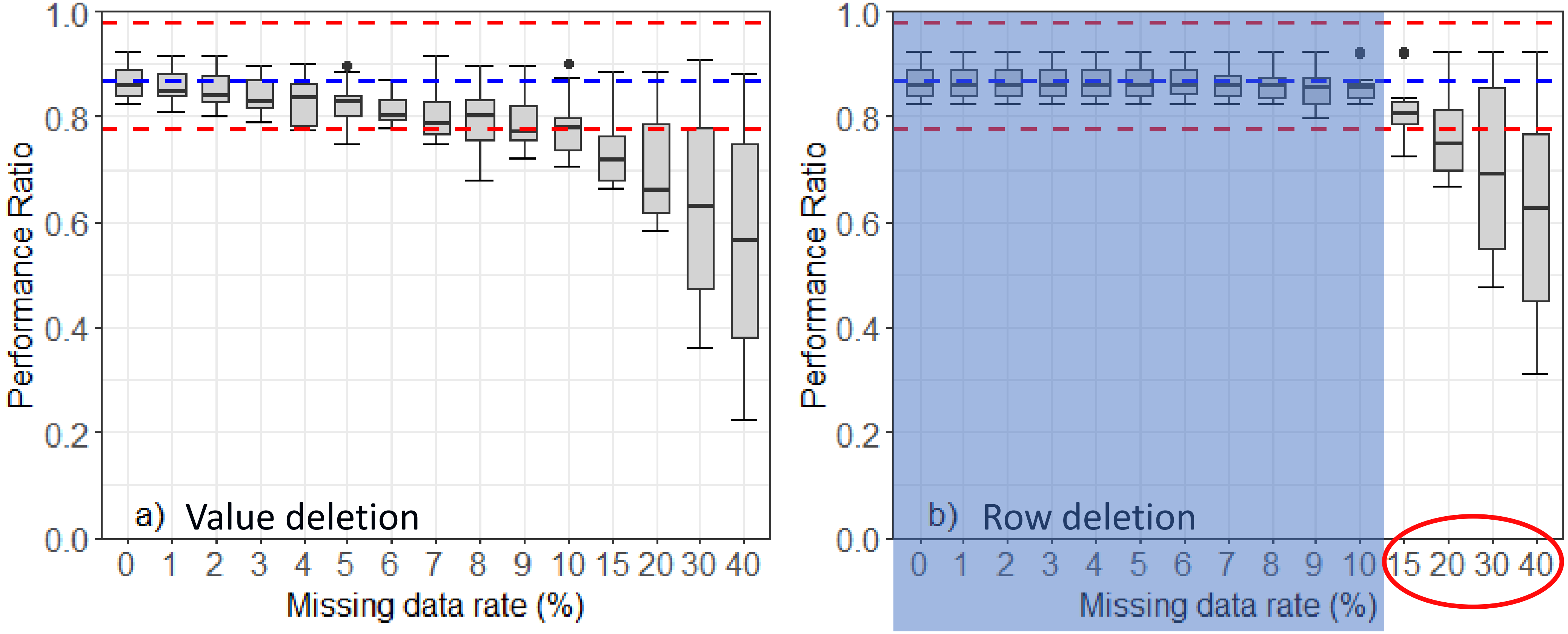
- 1-year PV datasets with random missing power measurements



The effect of random missing power measurements was mitigated by **row deletion**

Results - Data deletion methods

- 1-year PV datasets with continuous missing power measurements

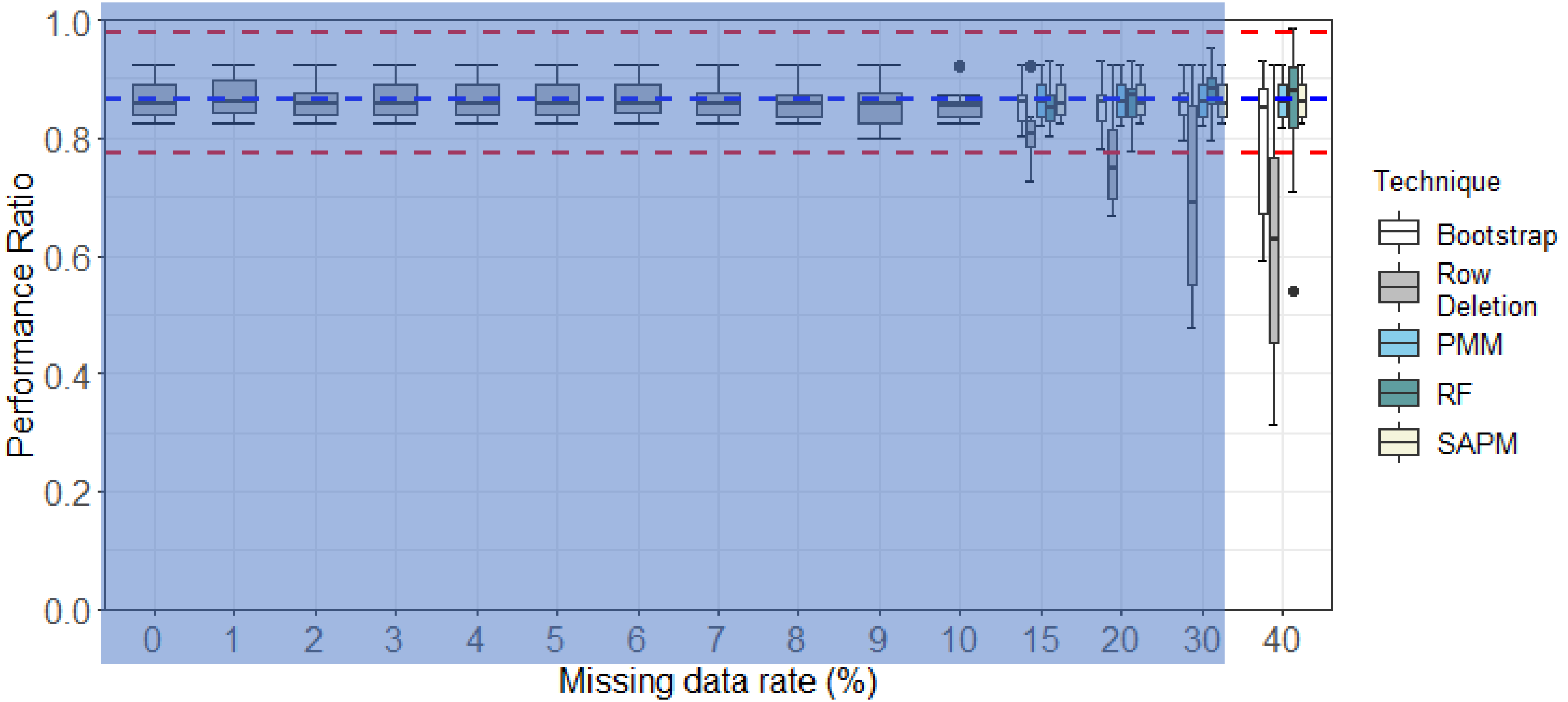


Row deletion was effective for handling continuous missing power measurements up to 10%

FOSS
Research Centre for Sustainable Energy

Results - Data inference techniques

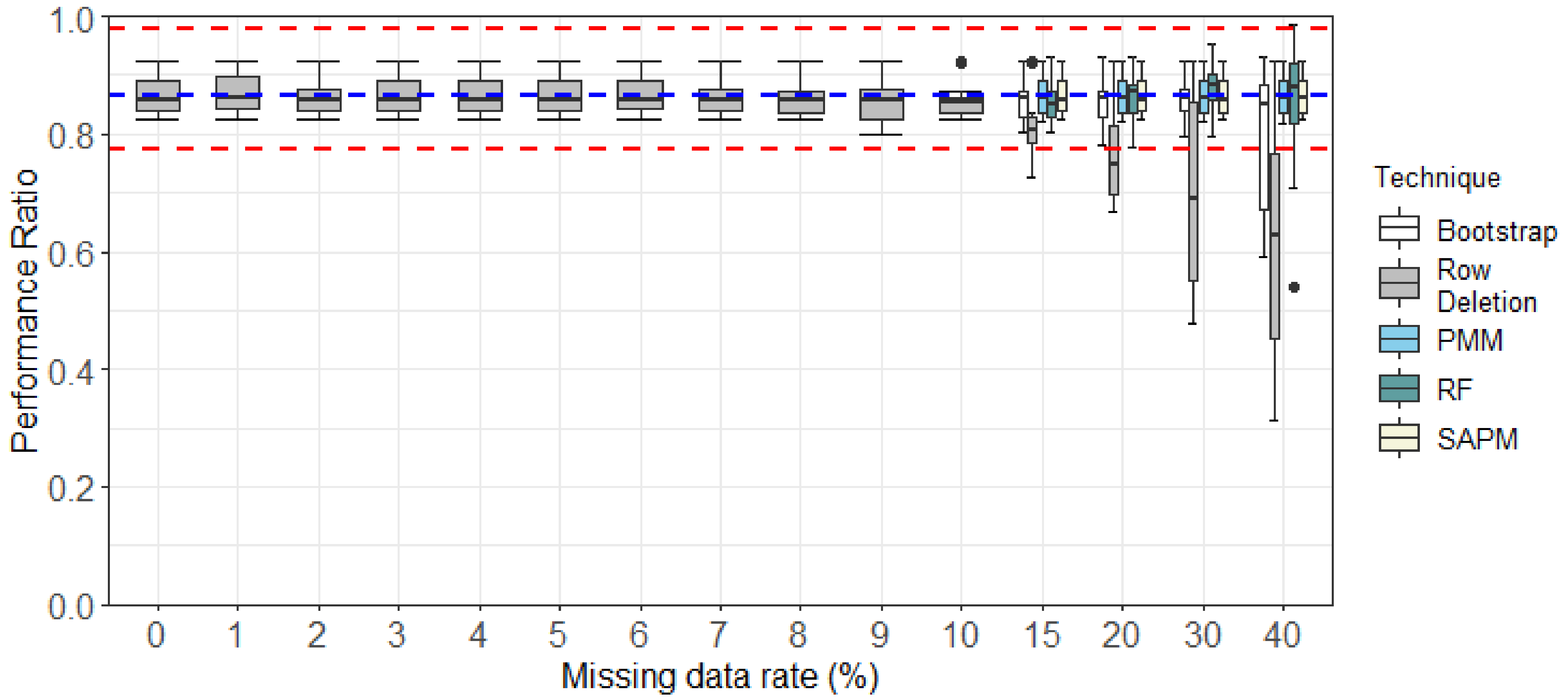
- Invalid missing power datasets were reconstructed by data inference techniques



Univariate imputation can be used for up to 30% missing data rate

Results - Data inference techniques

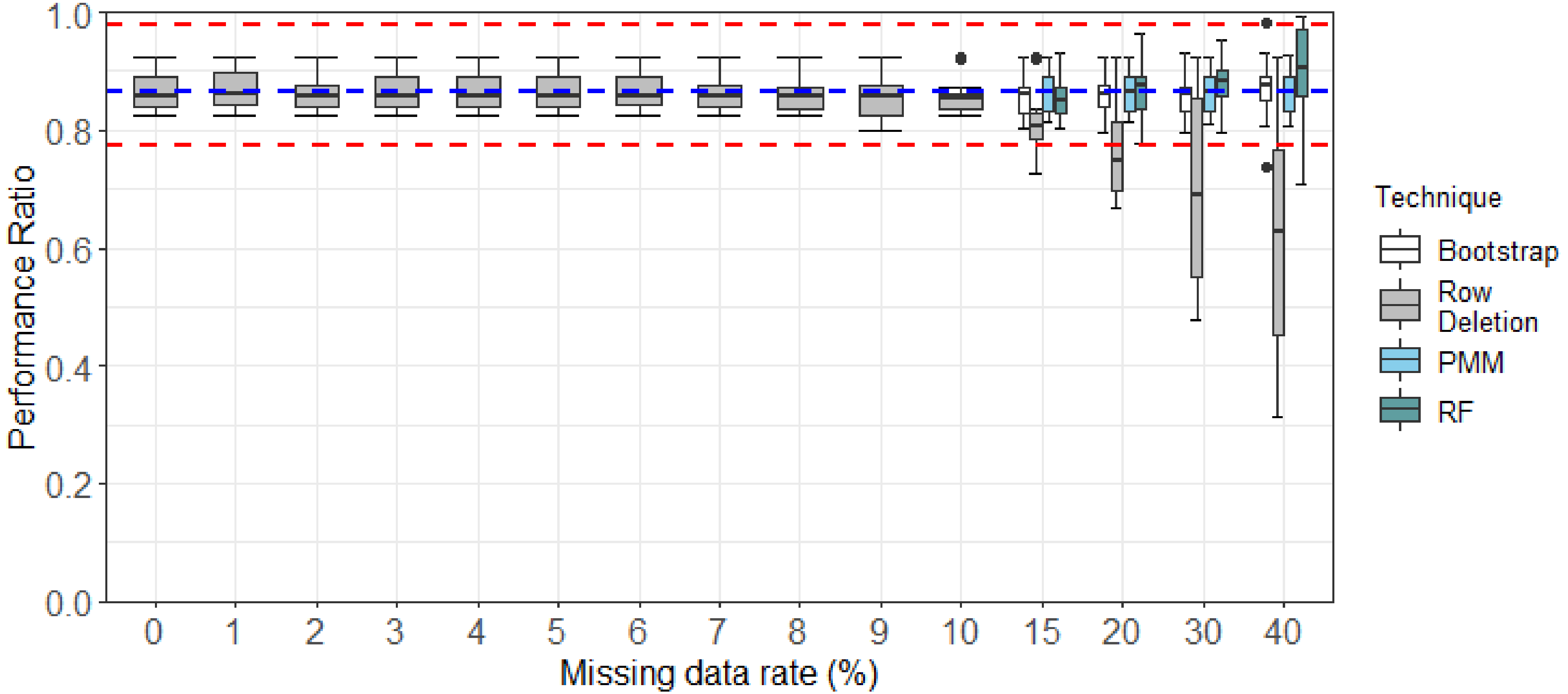
- Invalid missing power datasets were reconstructed by data inference techniques



Data inference with SAPM yielded the lowest error (max APE of 0.81%)

Results - Data inference techniques

- Invalid missing irradiance datasets were reconstructed by data inference techniques

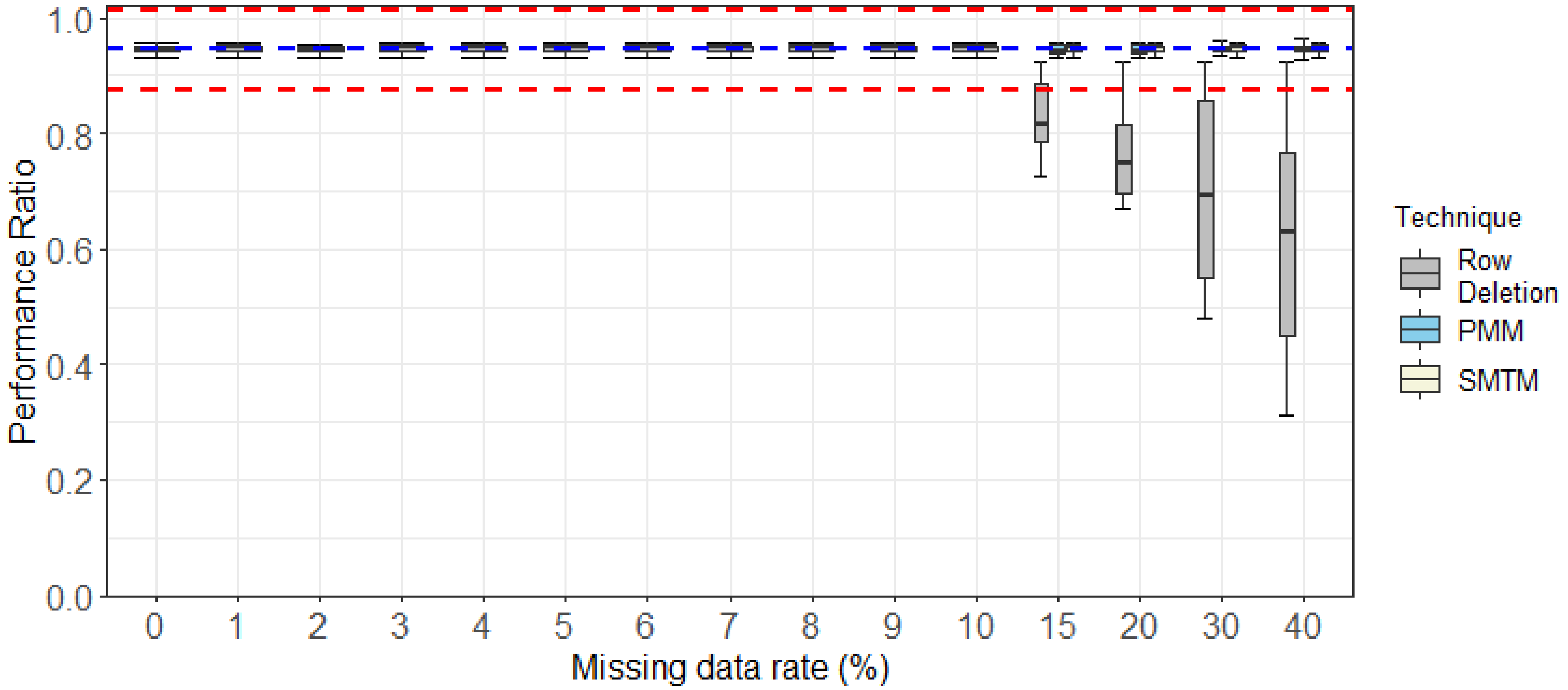


Data inference with PMM yielded the lowest error (max APE of 1.97%)

FOSS
Research Centre for Sustainable Energy

Results - Data inference techniques

- Invalid missing temperature datasets were reconstructed by data inference techniques

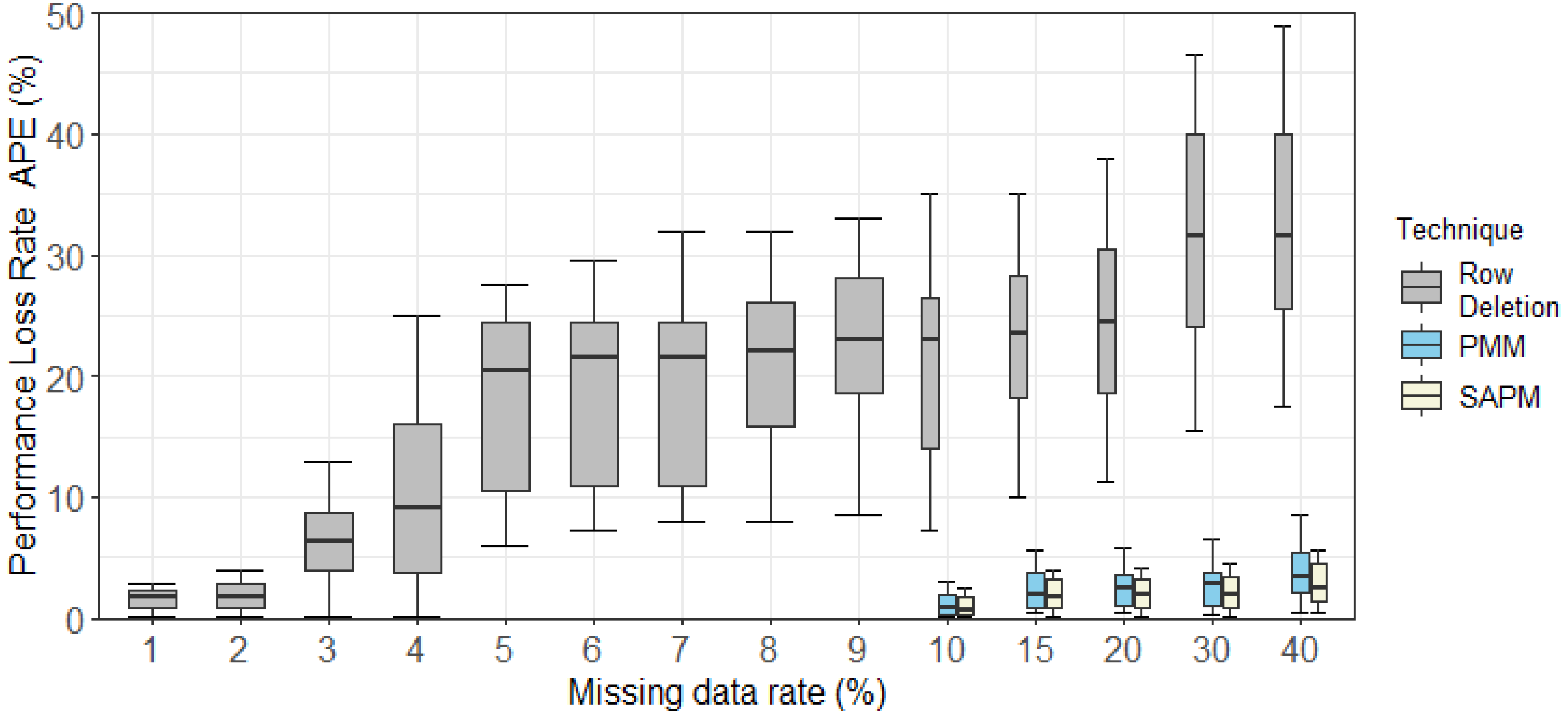


Data inference with SMTM yielded the lowest error (max APE of 0.28%)

FOSS
Research Centre for Sustainable Energy

Results - Data integrity effect on PLR analysis

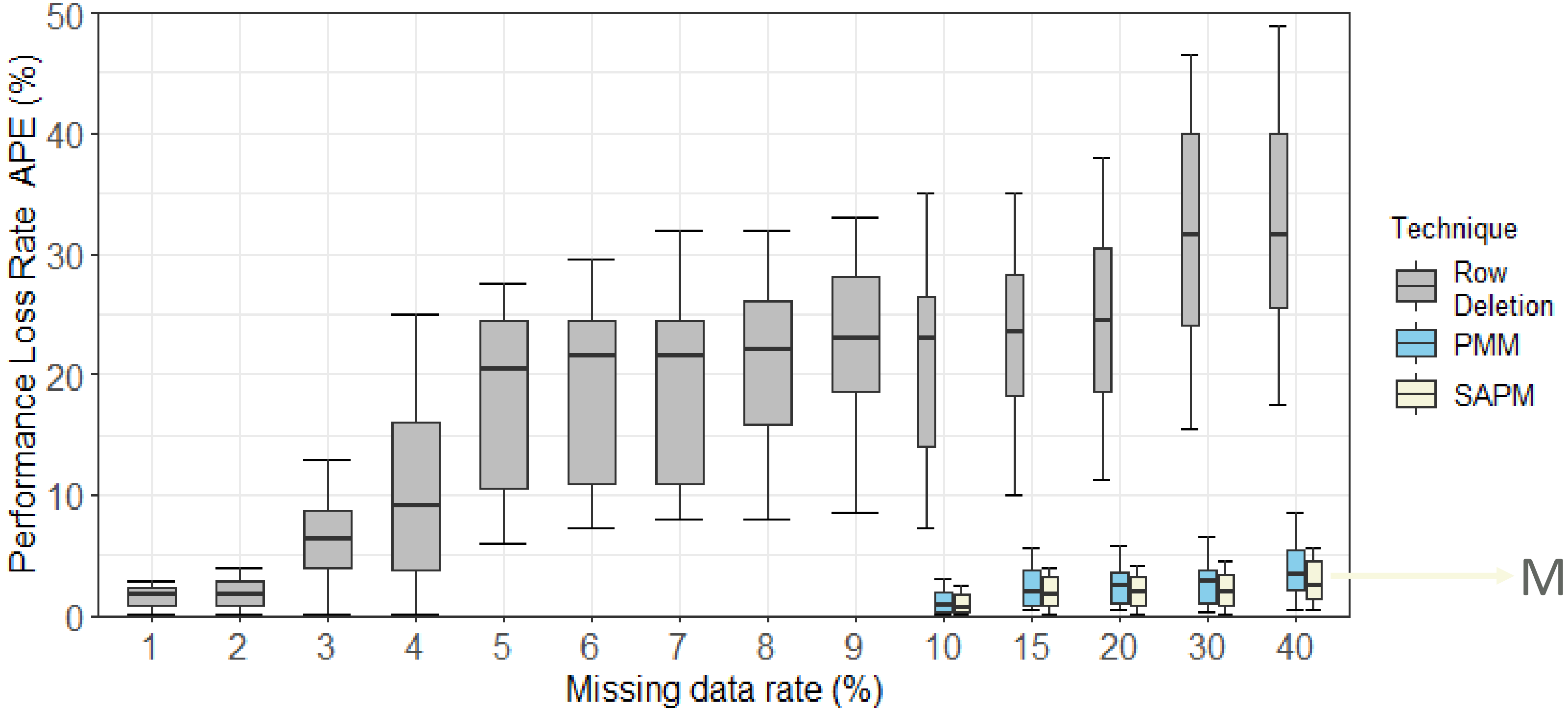
- Linear PLR was estimated by Ordinary Least Squares (OLS)
- 5-year PV dataset with continuous missing power measurements



Row deletion introduces a bias in the PLR calculation with the increasing missing data rate

Results - Data integrity effect on degradation analysis

- Linear PLR was estimated by Ordinary Least Squares (OLS)
- 5-year PV dataset with continuous missing power measurements



Max APE of 5.69%

SAPM yielded more accurate results when compared to the PLR estimates of row deletion

Conclusions and Future Work

- DQRs were developed to ensure data validity and bridge the quantitative gap that exists in current practices
- The results demonstrated that datasets with:
  - missing data rates < 10%, can be reconstructed by row deletion
  - missing data rates > 10%, can be reconstructed by data inference techniques
- For PLR estimates, data inference techniques are recommended even at missing data rates as low as 3% (in the case of OLS)
- Electrical models yielded the lowest error among the investigated techniques
- Future work will focus on determining the impact of invalid data and reconstruction routines on additional performance indices

Acknowledgements

SMARTER-PV project (BILATERAL/ISRAEL/0518/0017)

RESTART 2016-2020 RESEARCH
ΠΡΟΓΡΑΜΜΑΤΑ ΕΡΕΥΝΑΣ, ΤΕΧΝΟΛΟΓΙΚΗΣ ΑΝΑΠΤΥΞΗΣ ΚΑΙ ΚΑΙΝΟΤΟΜΙΑΣ

RESEARCH & INNOVATION FOUNDATION

SOLAR ENERGY TECHNOLOGIES OFFICE
U.S. Department Of Energy

Solar Energy Technologies Office Award Number 34364

Thank you for your attention

Andreas Livera
Special Scientist

University of Cyprus
FOSS Research Centre
1 University Avenue

livera.andreas@ucy.ac.cy

Related paper

PROGRESS IN
PHOTOVOLTAICS

RESEARCH ARTICLE | Open Access

Data processing and quality verification for improved photovoltaic performance and reliability analytics

Andreas Livera, Marios Theristis, Elena Koumpli, Spyros Theocharides, George Makrides, Juergen Sutterlueti, Joshua S. Stein, George E. Georghiou

First published: 07 October 2020 | https://doi.org/10.1002/pip.3349

The Team

solarcentury

Gantner instruments

University of Cyprus PV Technology

Sandia National Laboratories