Performance Imputation Techniques for Assessing Costs of Technical Failures in PV Systems

S. Lindig, A. Louwen, M. Herz, J. Acencio-Vásquez, D. Moser, M. Topič

Pearl PV Workshop

23rd of September 2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under GA. No. 952957.

MOTIVATION

Statistical analysis of technical failures

Workflow

Failure appearance in PV plant

Creation of ticket in SCADA system

Classification of failure according to TRUST PV's Risk Matrix

Resolution of failure

Statistical analysis of failure (CPN[1])

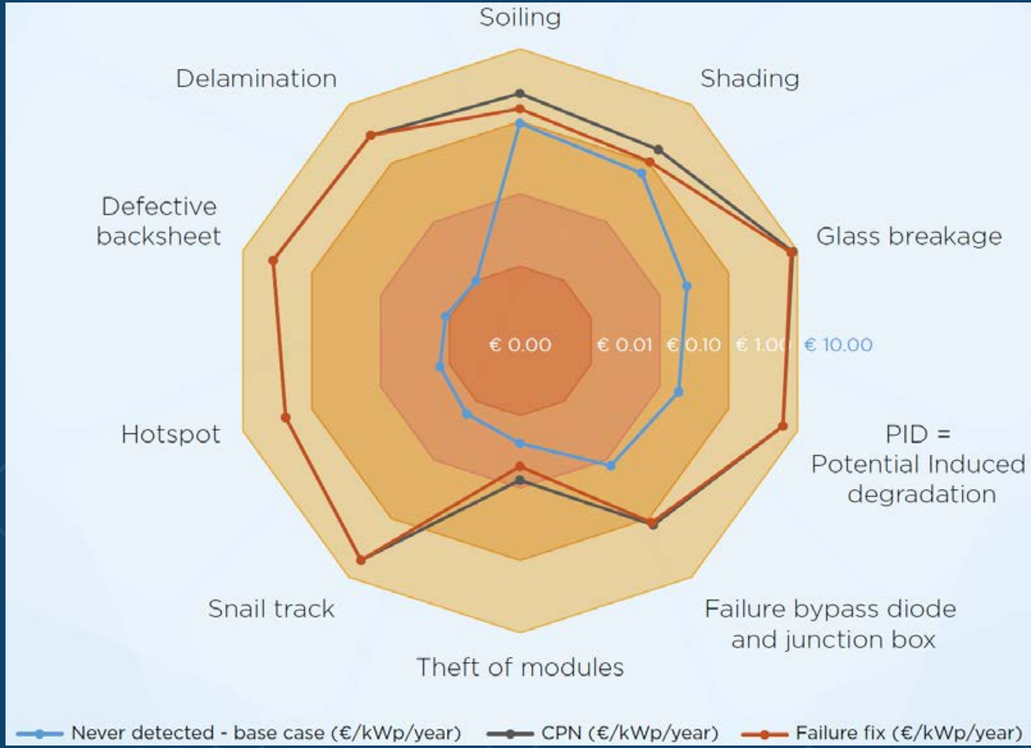
[1]H2020 Solar Bankability project http://www.solarbankability.org/home.html

MOTIVATION
Cost Priority Number[1]

Calculate the appearance likeliness & severity of technical failures in PV systems

$$\boxed{CPN = C_{down} + C_{fix}}$$

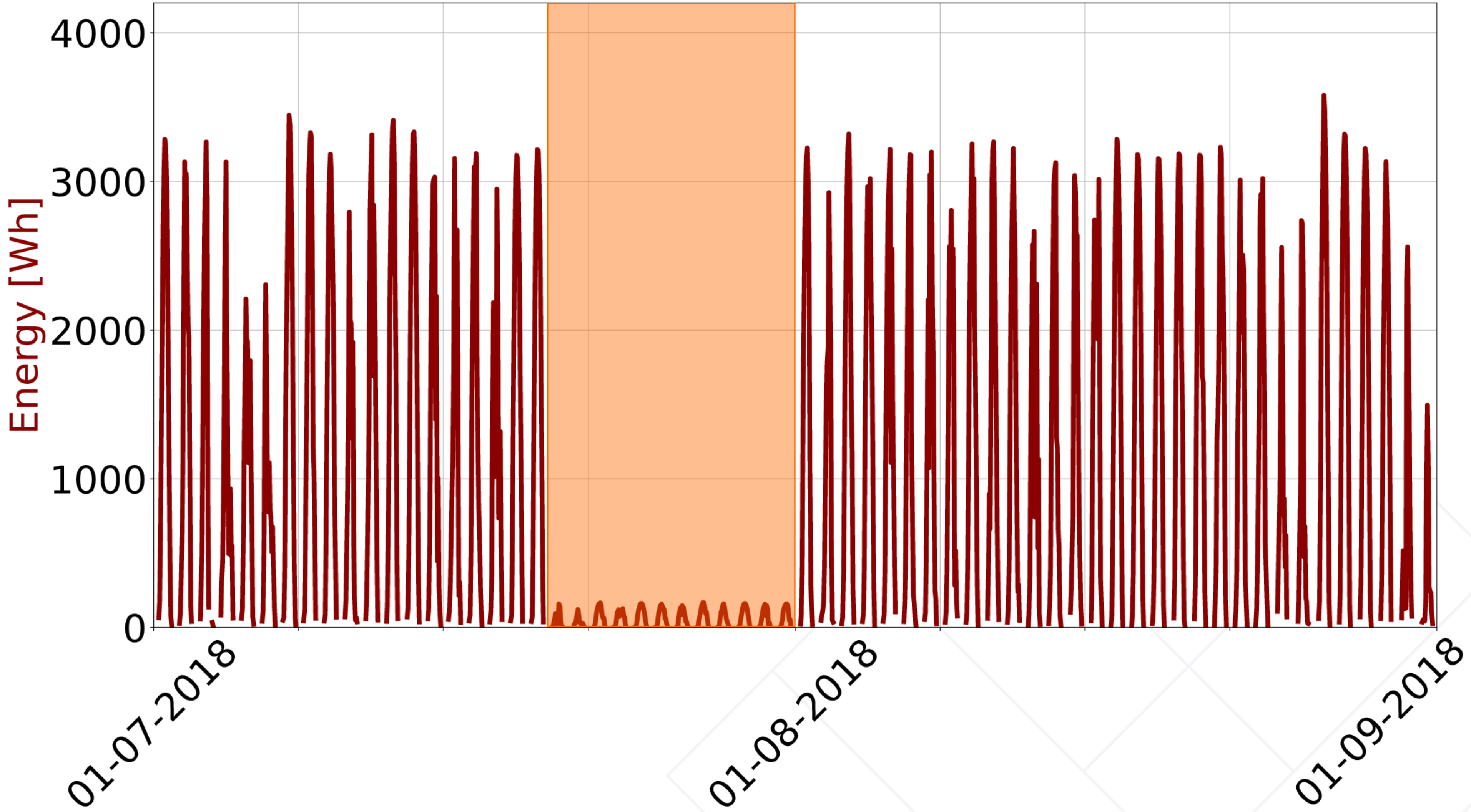
C_{down} – costs related to downtime of plant
C_{fix} – costs related to fix failures



[1]H2020 Solar Bankability project http://www.solarbankability.org/home.html

1

Missing data



Parameter

Type of missing data
Amount of missing data
Ratio training set / test set
Predictor availability
Imputation models

Parameter

Type of missing data

pairwise

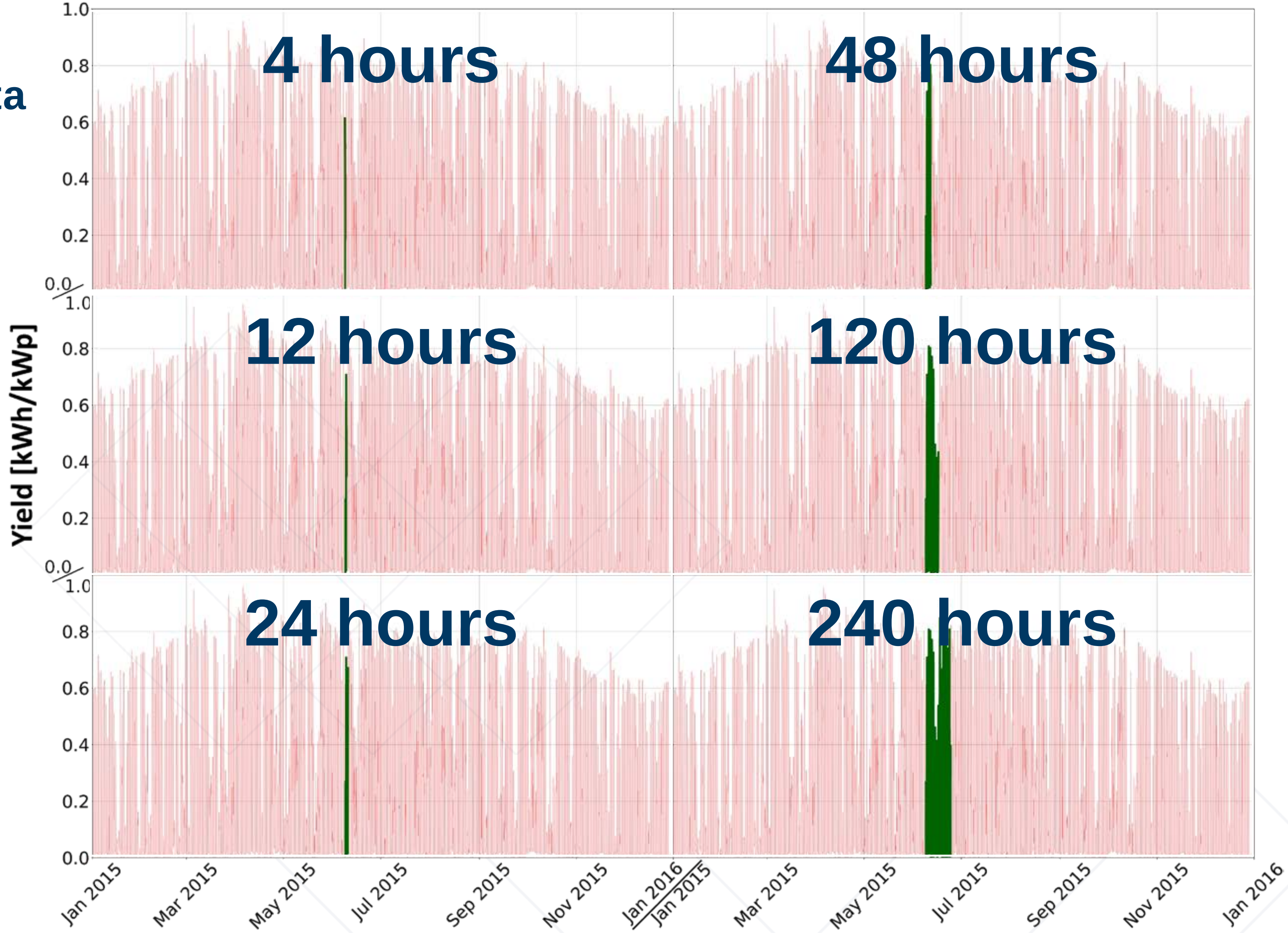
	time	Power	T_amb	G_POA	T_mod	wind_speed	G_POA_cams	G_POA_ERA5	T_amb_ERA5	pc_Si2
39	2018-07-04T09:00:00.000Z	895.39	20.9	274.82	28.1	0.7	486.66	313.36	17.89	956.88
40	2018-07-04T10:00:00.000Z	1634.4	22.89	489.39	37.75	0.87	636.12	512.85	20.47	1703.76
41	2018-07-04T11:00:00.000Z	2707.24	25.16	802.04	49.49	1.22	835.09	707.86	22.66	2721.69
42	2018-07-04T12:00:00.000Z	3130.9	28.18	933.61	58.42	1.23	838.03	866.99	24.42	3122.23
43	2018-07-04T13:00:00.000Z	1992.11	27.74	562.83	48.28	2.16	775.74	798.34	24.77	2105.26
44	2018-07-04T14:00:00.000Z	1112.32	26.31	283.08	35.89	2.07	474.78	717.59	22.19	1034.11
45	2018-07-04T15:00:00.000Z	496.63	25.28	148.99	28.49	1.14	169.76	472.58	21.79	542.97
46	2018-07-04T16:00:00.000Z	833.97	24.94	231.29	29.82	1.97	304.14	329.48	21.34	876.85
47	2018-07-04T17:00:00.000Z	933.38	25.1	263.18	30.95	1.1	160.01	239.42	21.67	959.09
48	2018-07-04T18:00:00.000Z	483.94	25.31	162.74	30.15	1.12	160.38	96	22.13	532.66
49	2018-07-05T07:00:00.000Z	119.25	18.58	42.36	17.98	1.5	60.94	55.94	12.11	143.13
50	2018-07-05T08:00:00.000Z	0	19.74	179.99	22.49	0.95	260.94	165.06	15.68	622.48
51	2018-07-05T09:00:00.000Z	0	22.23	356.7	32.96	1.08	456.49	109.3	16.97	1313.59
52	2018-07-05T10:00:00.000Z	0	23.91	483.78	38.04	0.95	646.77	318.99	19.46	1654.35
53	2018-07-05T11:00:00.000Z	0	25.91	631.53	46.75	1.18	778.12	475.87	20.79	2268.67
54	2018-07-05T12:00:00.000Z	0	27.07	544.2	46.19	1.08	801.76	548.3	21.59	1964.51
55	2018-07-05T13:00:00.000Z	0	27.71	553.7	45.87	1.27	928.87	744.71	22.51	1938.68
56	2018-07-05T14:00:00.000Z	0	28.25	317.26	39.05	1.13	456.44	761	23.12	1163.87
57	2018-07-05T15:00:00.000Z	0	28.66	445.82	45.05	1.1	714.91	664.84	23.66	1560.4
58	2018-07-05T16:00:00.000Z	0	29.53	538.83	44.41	1.73	571.2	519.58	23.25	1900.06
59	2018-07-05T17:00:00.000Z	0	28.88	318.98	38.51	1.55	381.87	319.91	22	1099.82
60	2018-07-05T18:00:00.000Z	0	26.93	190.84	31.69	3.5	106.99	89.39	19.81	722.91
61	2018-07-06T07:00:00.000Z	0	16.26	47.36	15.49	0.53	135.68	64.45	14.24	130.85

listwise

	time	Power	T_amb	G_POA	T_mod	wind_speed	G_POA_cams	G_POA_ERA5	T_amb_ERA5	pc_Si2
39	2018-07-04T09:00:00.000Z	895.39	20.9	274.82	28.1	0.7	486.66	313.36	17.89	956.88
40	2018-07-04T10:00:00.000Z	1634.4	22.89	489.39	37.75	0.87	636.12	512.85	20.47	1703.76
41	2018-07-04T11:00:00.000Z	2707.24	25.16	802.04	49.49	1.22	835.09	707.86	22.66	2721.69
42	2018-07-04T12:00:00.000Z	3130.9	28.18	933.61	58.42	1.23	838.03	866.99	24.42	3122.23
43	2018-07-04T13:00:00.000Z	1992.11	27.74	562.83	48.28	2.16	775.74	798.34	24.77	2105.26
44	2018-07-04T14:00:00.000Z	1112.32	26.31	283.08	35.89	2.07	474.78	717.59	22.19	1034.11
45	2018-07-04T15:00:00.000Z	496.63	25.28	148.99	28.49	1.14	169.76	472.58	21.79	542.97
46	2018-07-04T16:00:00.000Z	833.97	24.94	231.29	29.82	1.97	304.14	329.48	21.34	876.85
47	2018-07-04T17:00:00.000Z	933.38	25.1	263.18	30.95	1.1	160.01	239.42	21.67	959.09
48	2018-07-04T18:00:00.000Z	483.94	25.31	162.74	30.15	1.12	160.38	96	22.13	532.66
49	2018-07-05T07:00:00.000Z	119.25	18.58	42.36	17.98	1.5	60.94	55.94	12.11	143.13
50	2018-07-05T08:00:00.000Z	0	0	0	0	0	0	0	0	0
51	2018-07-05T09:00:00.000Z	0	0	0	0	0	0	0	0	0
52	2018-07-05T10:00:00.000Z	0	0	0	0	0	0	0	0	0
53	2018-07-05T11:00:00.000Z	0	0	0	0	0	0	0	0	0
54	2018-07-05T12:00:00.000Z	0	0	0	0	0	0	0	0	0
55	2018-07-05T13:00:00.000Z	0	0	0	0	0	0	0	0	0
56	2018-07-05T14:00:00.000Z	0	0	0	0	0	0	0	0	0
57	2018-07-05T15:00:00.000Z	0	0	0	0	0	0	0	0	0
58	2018-07-05T16:00:00.000Z	0	0	0	0	0	0	0	0	0
59	2018-07-05T17:00:00.000Z	0	0	0	0	0	0	0	0	0
60	2018-07-05T18:00:00.000Z	0	0	0	0	0	0	0	0	0
61	2018-07-06T07:00:00.000Z	0	0	0	0	0	0	0	0	0

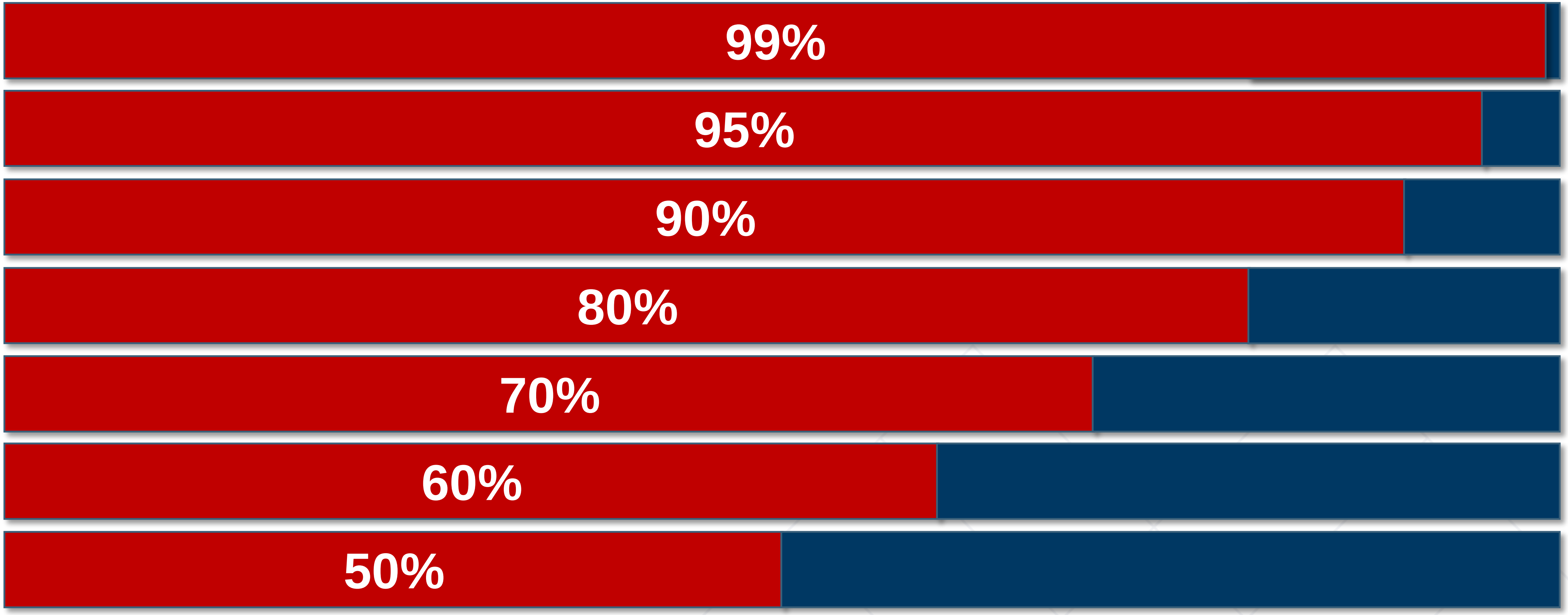
Parameter
Amount of missing data

- Depends on
  - Failure
  - O&M strategy (detection method, spare part management, etc.)



Parameter

Ratio training set / test set



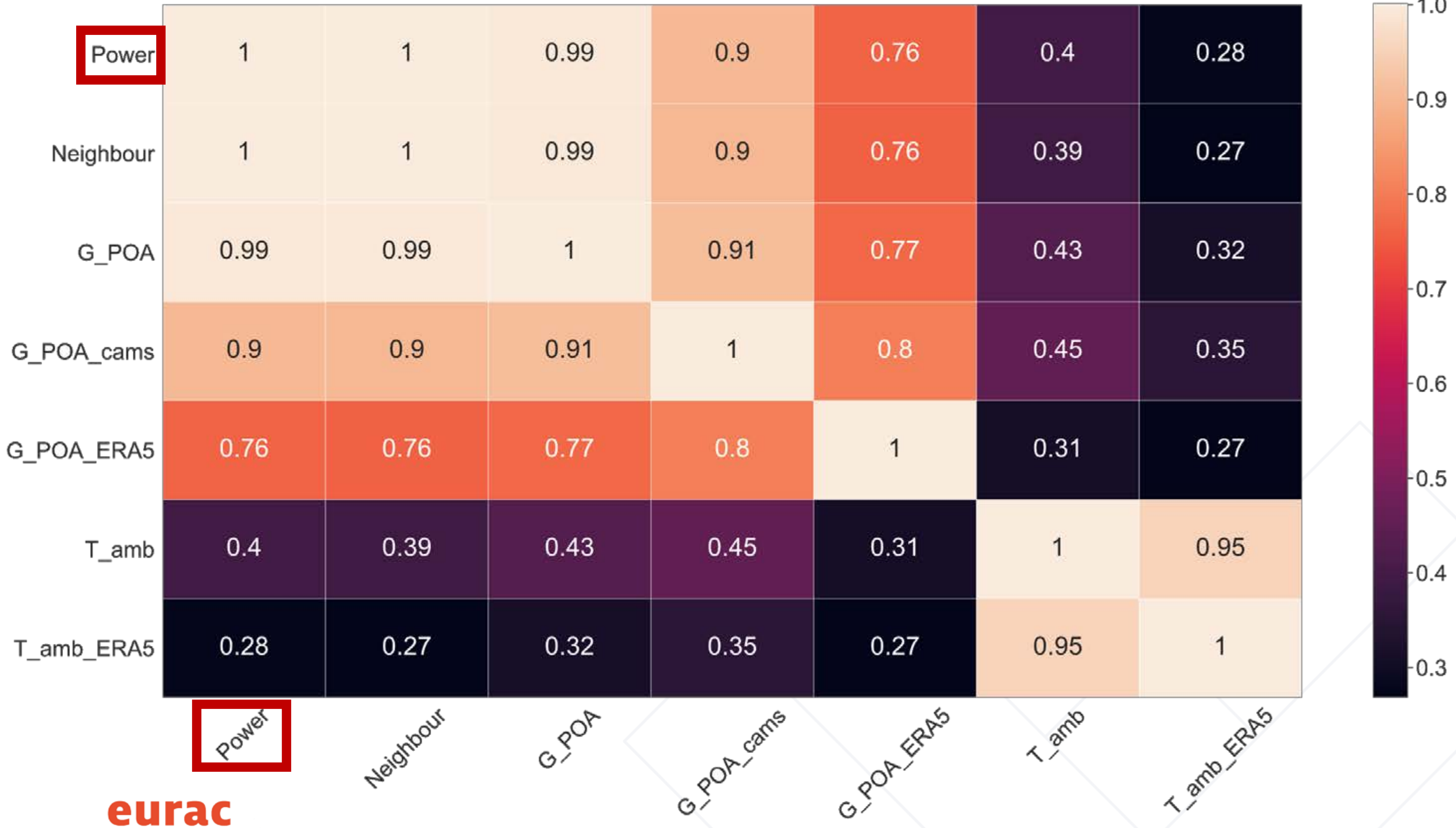
Parameter
Predictor availability

Measured parameter	Climate reanalysis data	No data
Climate data	ERA5[2] + tilt & azimuth	Univariate imputation methods
G_{POA} T_{amb}	G_{POA} T_{amb}	
Neighbouring string/plant	Cams[3] + tilt & azimuth	
Power	G_{POA}	

[2] Copernicus Climate Change Service (C3S) ERA5, “Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus climate change service climate data store (cds),» https://cds.climate.copernicus.eu/cdsapp/home, 2017.
[3] CAMS: Surface solar radiation data, [Online]. Available: https://atmosphere.copernicus.eu. [Accessed 12 01 2021].

Parameter

Predictor availability



Parameter
Imputation models

Regression / ML models	Empirical models	Univariate models
Multivariate linear regression[4]	PVWatts[5]	Hourly average
K-nearest neighbors[4]	PVGIS[6]	Random value[8]
Decision tree[4]	3 parameter model[7]	Kalman smooth[8]
Random forest[4]		Interpolation[9,10]
Extra tree[4]		
Gradient boosting regressor[4]		
Histogram based gradient boosting regressor[4]		

[4] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, p. 2825-2830, 2011.
[5] NREL, PVWatts Version 5 Manual, National Renewable Energy Laboratory, 2014.
[6] D. L. King, W. Boyson and J. Kratochvill, "Photovoltaic Array Performance Model," Sandia National Laboratories, Alberquerque, 2004.
[7] K. Ding, Z. Ye and T. Reindl, "Comparison of Parameterisation Models for the Estimation of the Maximum Power Output of PV Modules," Energy Procedia, vol. 25, p. 101–107, 2012.
[8] S. Moritz, T. Bartz-Beielstein, “imputeTS: Time Series Missing Value Imputation in R,” The R Journal, vol. 9 (1), p. 207-218, 2017.
[9] R. J. Hyndman et al., “forecast: Forecasting functions for time series and linear models,” R package version 8.11, 2020.
[10] R. J. Hyndman, Y. Khandakar, “Automatic time series forecasting: the forecast package for R,” Journal of Statistical Software, vol. 26 (3), p. 1-22, 2008.

PV plant

Poly-crystalline Si system

Nominal power 4.2kWp

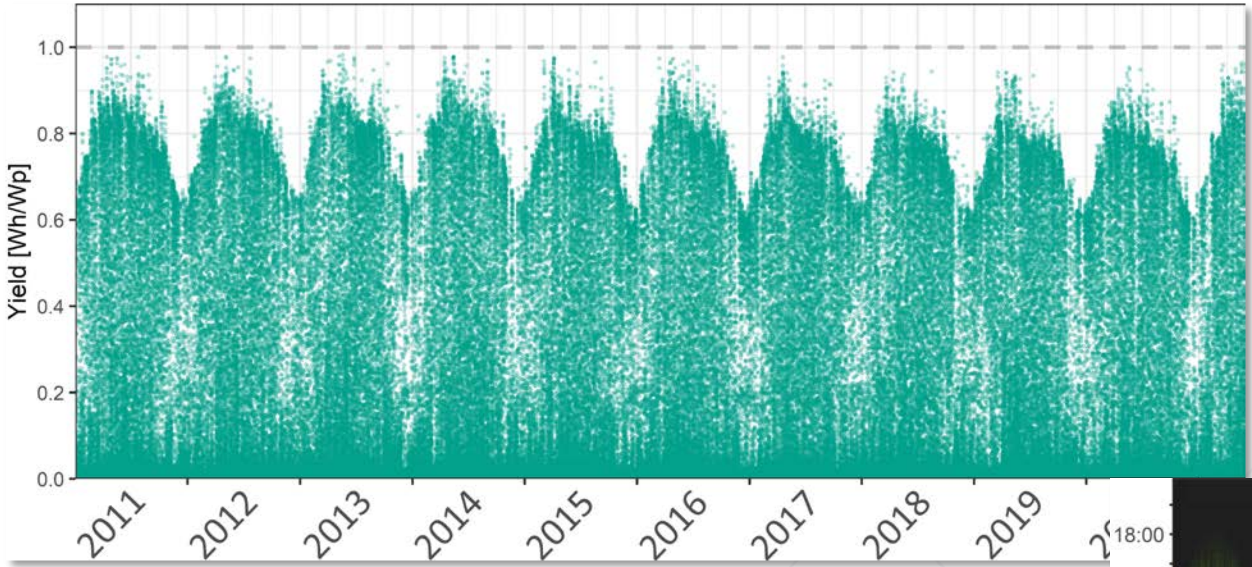
In operation since 2010

Weather station:

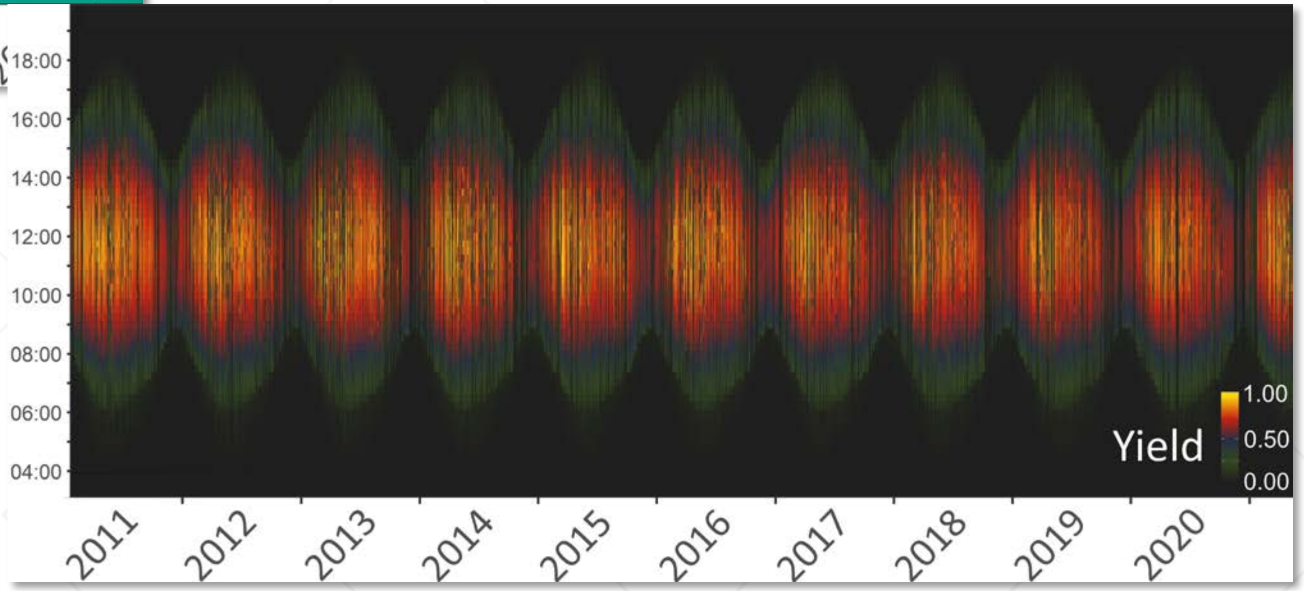
Plane-of-array irradiance

Ambient temperature

PV plant



Yield [Wh/Wp]
1.0
0.8
0.6
0.4
0.2
0.0
2011
2012
2013
2014
2015
2016
2017
2018
2019

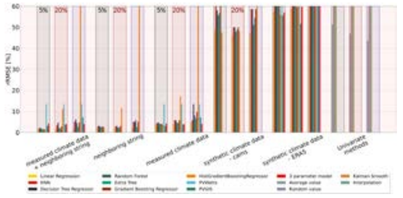


18:00
16:00
14:00
12:00
10:00
08:00
06:00
04:00
Yield
1.00
0.50
0.00
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020

Statistical metrics

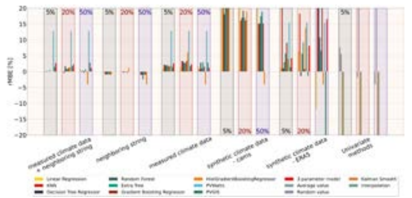
$$rRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{N}(\tilde{y}_i - y_i)^2}{N}}}{\frac{1}{N}\sum_{i=1}^{N} y_i} * 100\%$$

Root mean square error –
Overall prediction accuracy



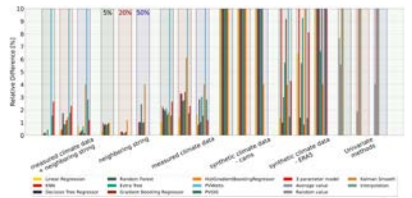
$$rMBE = \frac{\sum_{i=1}^{N}(\tilde{y}_i - y_i)}{\sum_{i=1}^{n} y_i} * 100\%$$

Mean bias error –
over- or underestimating



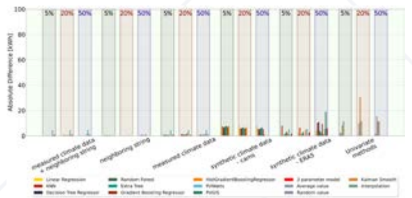
$$rD = \frac{\left|\sum_{i=1}^{N} y_i - \sum_{i=1}^{n} \tilde{y}_i\right|}{\sum_{i=1}^{n} y_i}$$

Relative difference –
aggregated distance to target

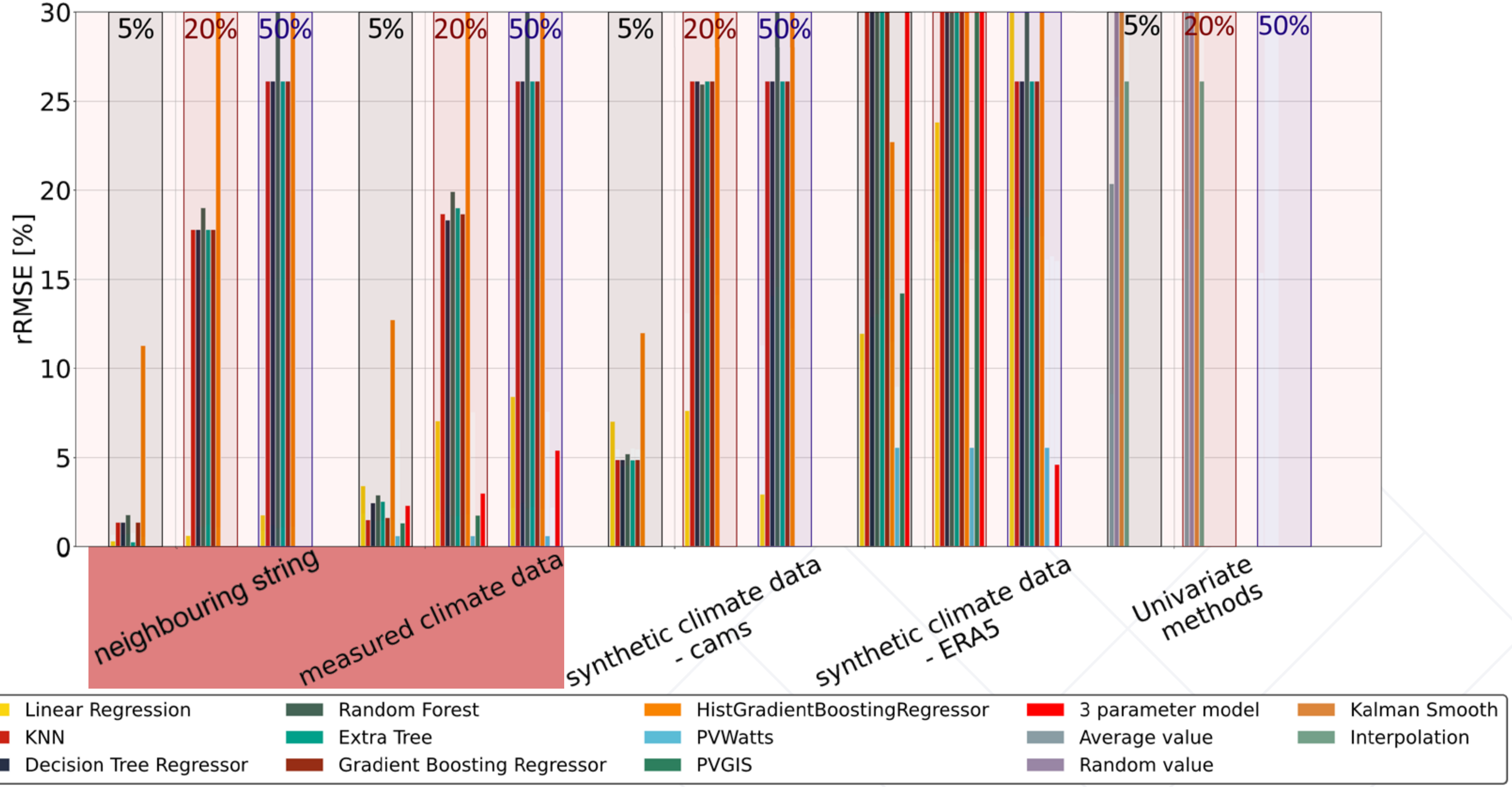


$$aD = \left|\sum_{i=1}^{N} y_i - \sum_{i=1}^{N} \tilde{y}_i\right|$$

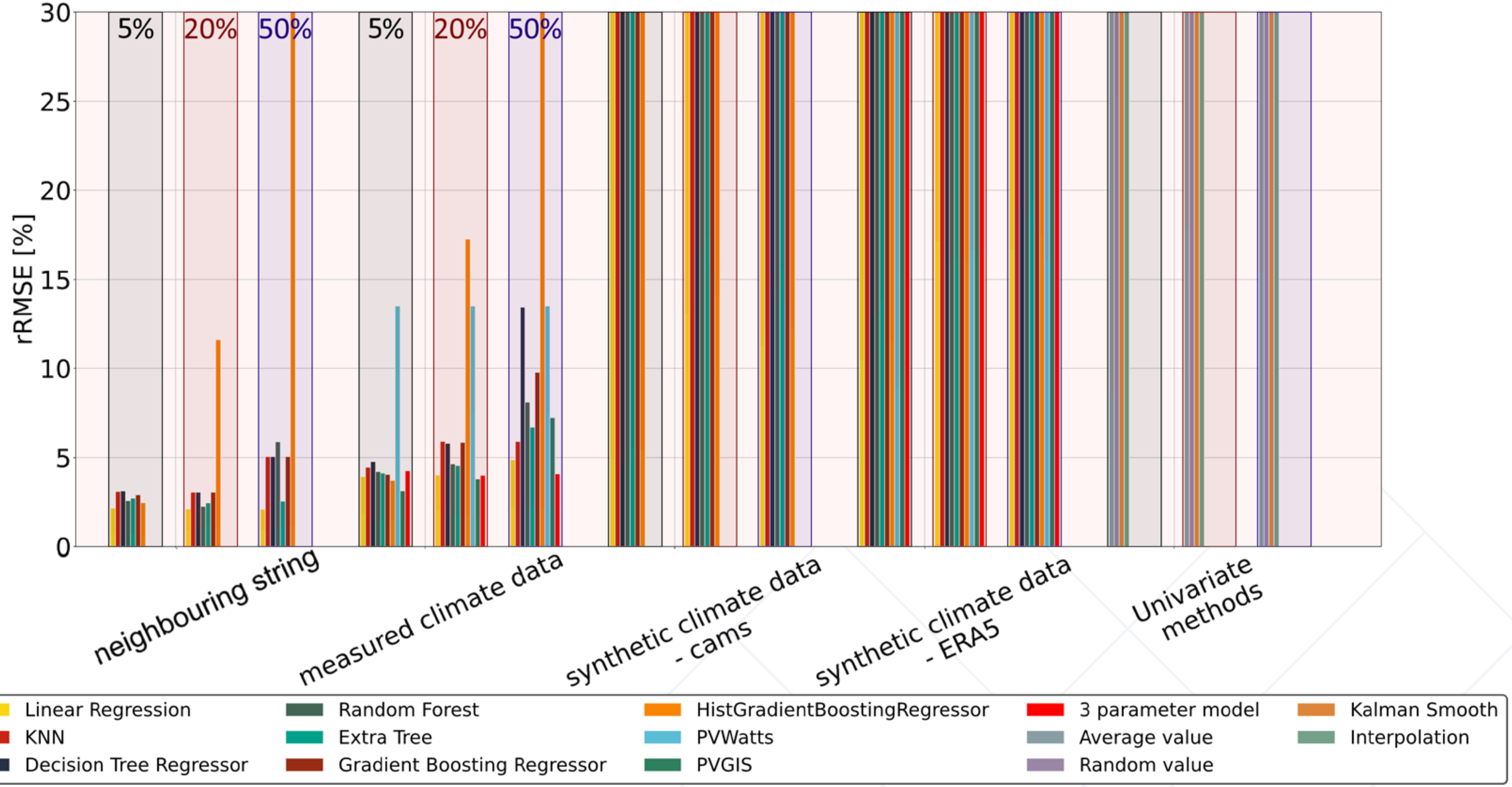
Absolute difference –
kWh from target



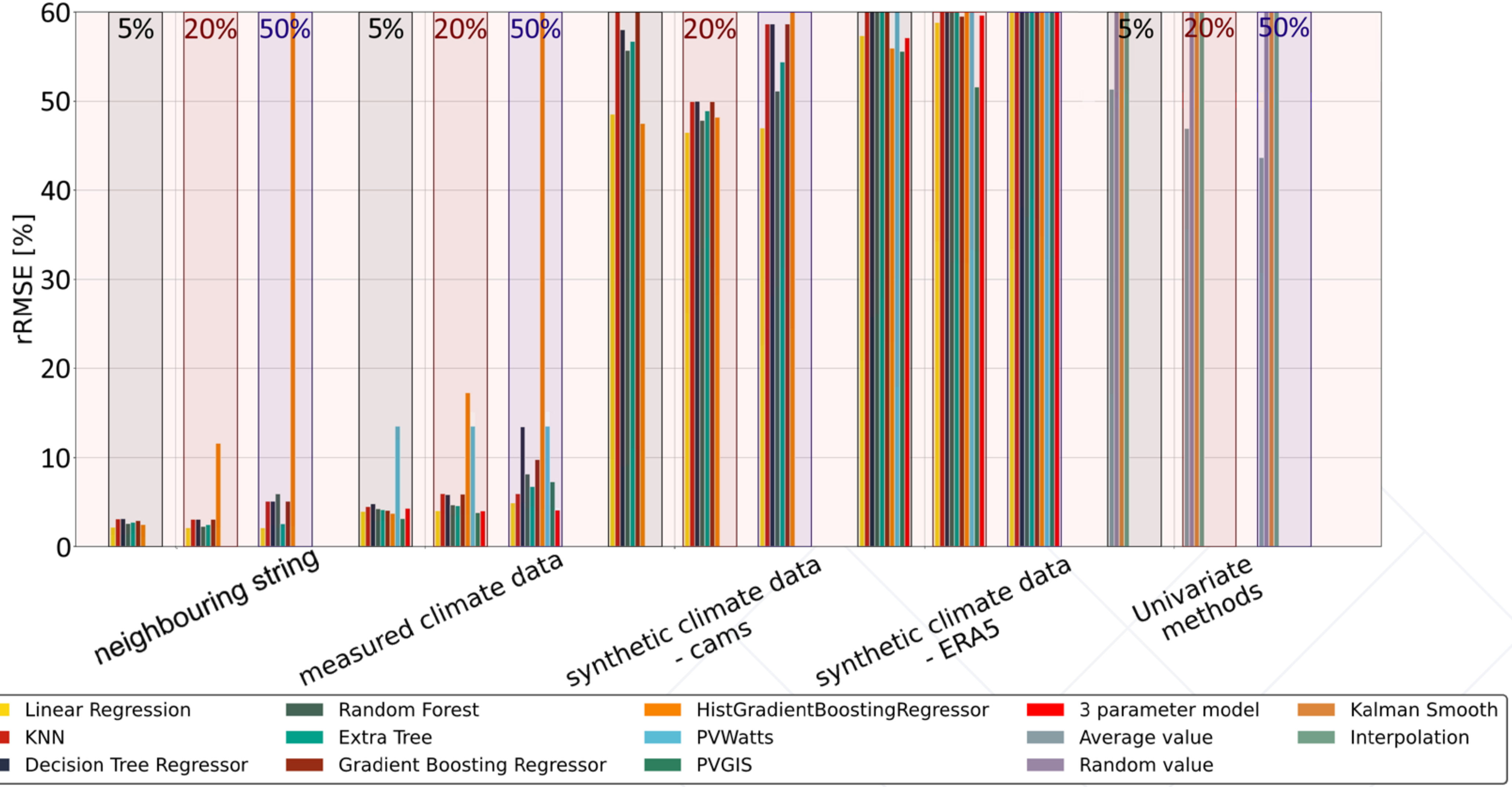
Results – rRMSE – 4 hours



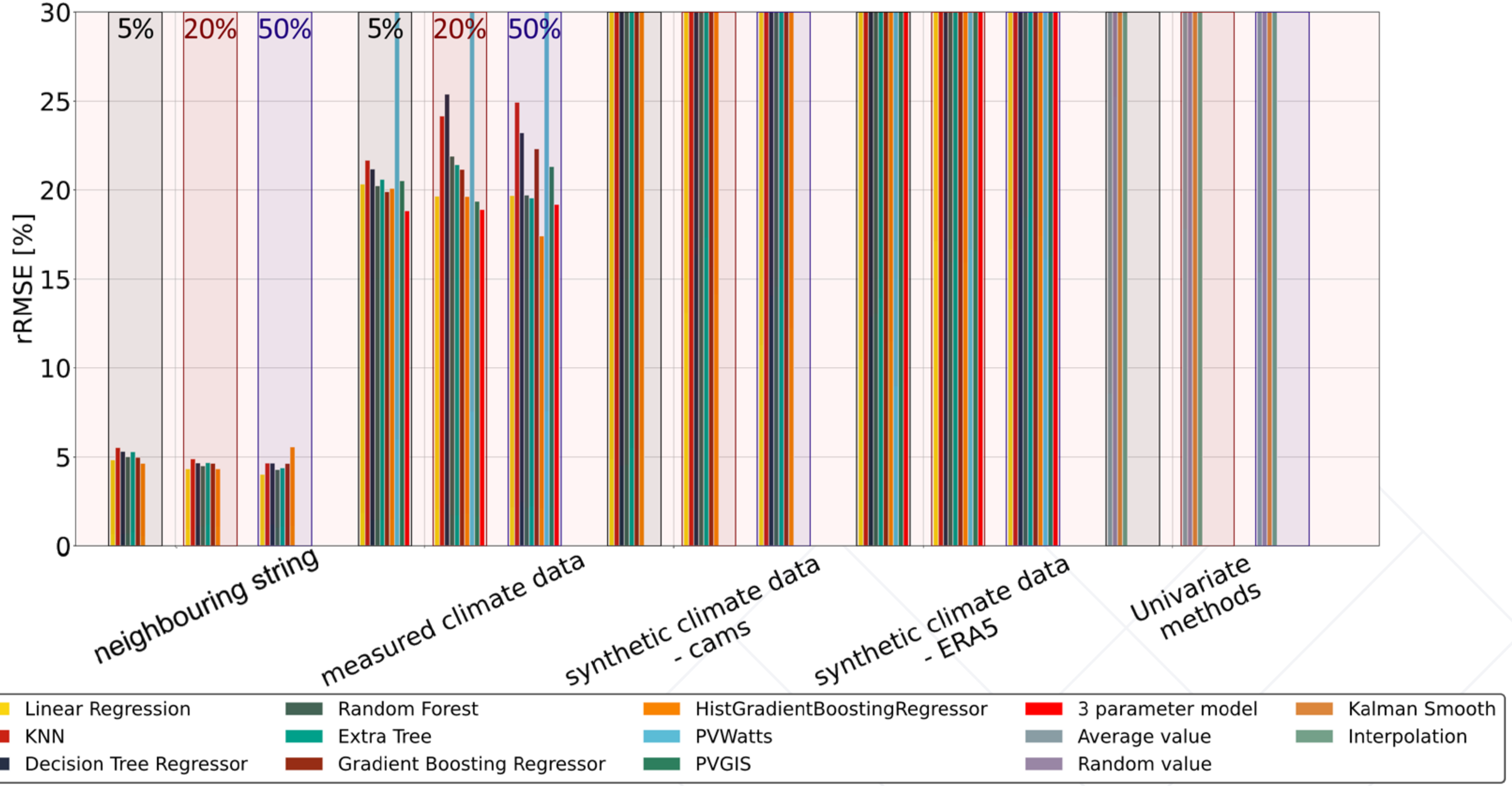
Results – rRMSE – 24 hours



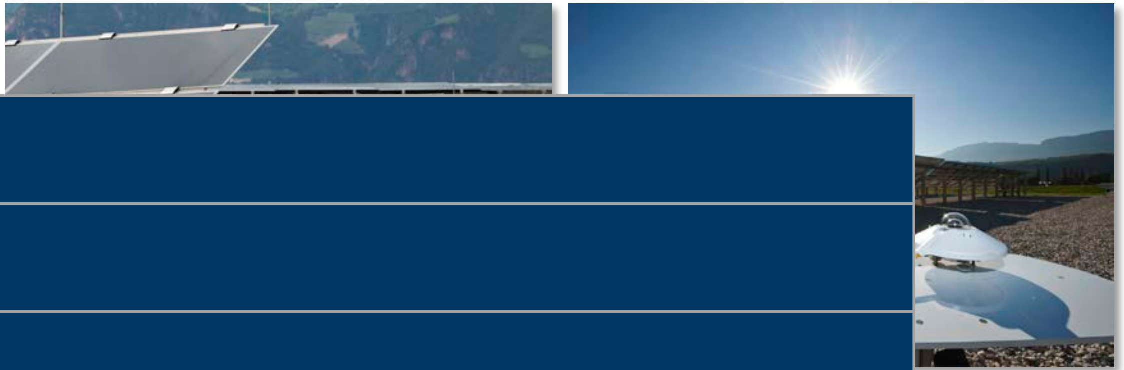
Results – rRMSE – 24 hours



Results – rRMSE – 240 hours



Summary



PREDICTORS

- Mea
  pred
- Clim
  univ

Type of missing data

Amount of missing data

Ratio training set / test set

Predictor availability

Imputation models

TRAIN

- with
  ratio
  accuracy

IMPUTATION MODELS

- Smaller data gaps: simpler ML models and power predictive models
- Bigger data gaps: more complex ML models

simpler ML & empirical models

more complex ML models

Outlook

- Extend study with more datasets
  - ➢ Study climate reanalysis data in other locations
- Economic sensitivity analysis of imputation methods


NOI TechPark Bolzano. Credits: NOI spa

Thank you for your attention

www.eurac.edu/
Sascha.Lindig@eurac.edu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under GA. No. 952957.

TRUSTPV
SOLAR PV, PERFORMANCE & RELIABILITY

https://trust-pv.eu/