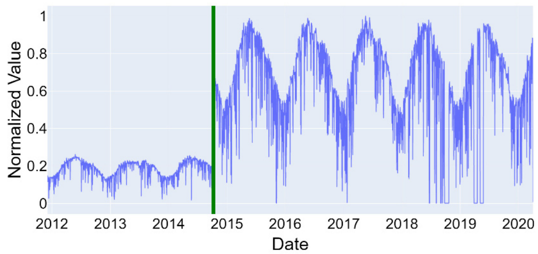
Automated Shift Detection in Sensor-Based PV Power and Irradiance Time Series

Kirsten Perry, Matthew Muller
NREL, Golden, CO, 80401, USA

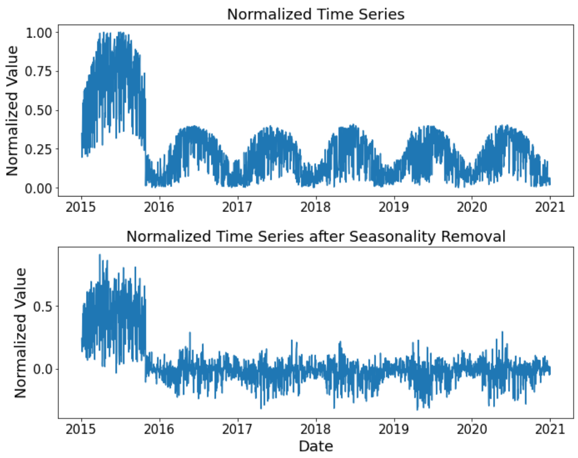
Introduction

- Large, abrupt data shifts can be induced unintentionally in sensor-based PV power and irradiance time series. Common reasons for data shifts include replacing hardware or performing software configuration changes [1].
- These shifts do not reflect an actual change in system performance but are usually a result of data acquisition issues. For example, by changing the scale factor for a particular data output, or by accidentally converting an AC energy data stream to an AC power data stream
- Including data with these shifts in PV analysis can lead to inaccurate results
- Previous research focusing on identifying and eliminating data shifts from time series has relied on manual identification of data shift periods, or assumed that degradation is linear when making corrections [1, 2, 3]
- Goal: to develop a solution that automatically identifies data shift points in PV time series data, with the option to filter out shifted periods for future PV analysis.

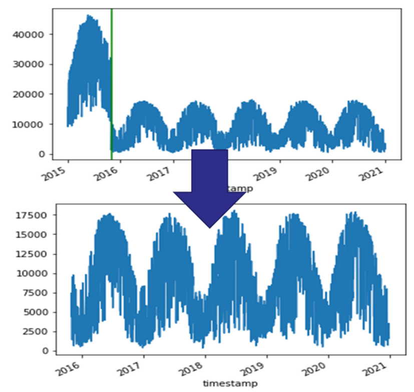


Methods

- Data sets: To benchmark algorithm performance when detecting data shifts, 101 sensor-based PV power and irradiance data streams were manually labeled
- Data pre-processing: The following pre-processing steps were performed on each labeled time series:
  - Erroneous data, such as negative/stale data and outliers, was removed
  - The time series was min-max normalized
  - For time series longer than 2 years in length, seasonality was removed
- Changepoint Detection: Changepoint detection was run on the processed time series (Ruptures Python package was used)
  - Grid Search: search method (Window, Binary Seg, Bottom-Up, PELT), cost function (rbf, L1, L2), penalty (ranges between 10-100)



Automated detection of data shifts via the PVAnalytics detect_data_shifts() function



Segmentation of longest time series sequence free of data shifts, via the PVAnalytics get_longest_shift_segment_dates() function

Results

- Each model configuration was run on the 101 labeled data sets, and model performance was benchmarked.
  - F1-Score: Measures the ability of each model to correctly identify data shift points within 30 days of their labelled occurrence.
  - Average run time: Taken across all 101 of the data sets
- Model results shown in the table below
- Best model for seasonality-removed data: Bottom-Up model. Fastest but still performant
- Best model for normalized-only data: Window-based model

Data	Model	Penalty	F1	Run Time (s)
Seasonality-removed	PELT	40	.767	50.81
Seasonality-removed	Binary Seg	50	.763	2.24
Seasonality-removed	**Bottom-Up**	**40**	**.760**	**.26**
Normalized Only	**Window**	**30**	**.745**	**.2**

PVAnalytics Package & Further Research

- All models developed in this research are publicly available via the Python PVAnalytics package (https://github.com/pvlib/pvanalytics/pull/142).
- The data shift pipeline includes the option to filter out the longest continuous time series segment free of data shifts, for future analysis (see below left image)
- In future research, we plan to hone our logic for selecting the "best" segment to run analysis on, by comparing segment data quality and availability
- Plan to investigate whether data shifts caused by scaling issues or similar can be identified and corrected, without compromising the overall quality of the time series and biasing future analyses

Acknowledgements

This work was authored in part by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Numbers 38258.

References

[1] D. C. Jordan, C. Deline, S. R. Kurtz, G. M. Kimball, and M. Anderson, "Robust pv degradation methodology and application," IEEE Journal of Photovoltaics, vol. 8, no. 2, pp. 525–531, 2018.
[2] D. C. Jordan and S. R. Kurtz, "Analytical improvements in pv degradation rate determination," 2010 35th IEEE Photovoltaic Specialists Conference, pp. 002 688–002 693, 2010.
[3] S. Lindig, A. Louwen, D. Moser, and M. Topic, "Outdoor pv system monitoring—input data quality, data imputation and filtering approaches," Energies, vol. 13, no. 19, 2020.

NREL/PO-5K00-83200