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

KIRSTEN PERRY

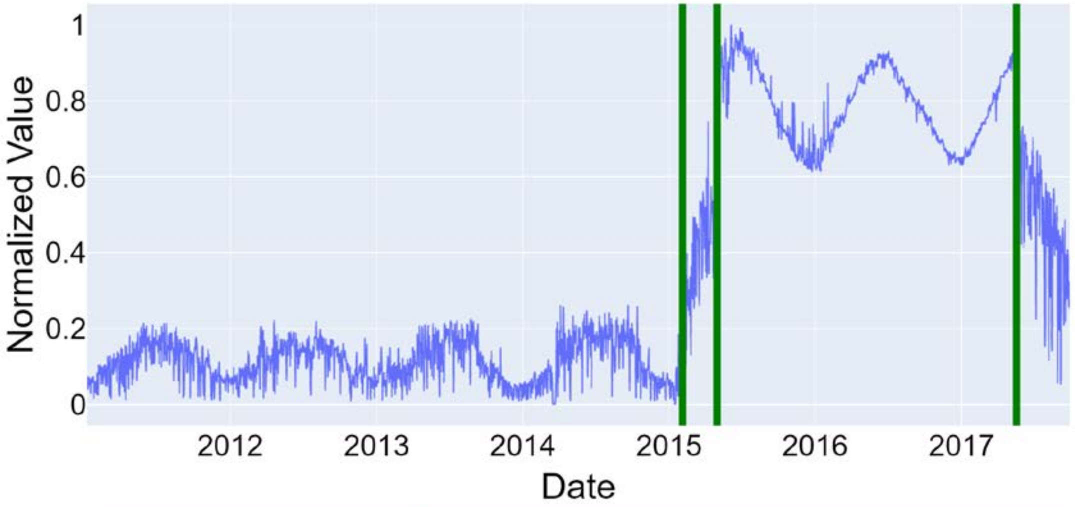
MATTHEW MULLER

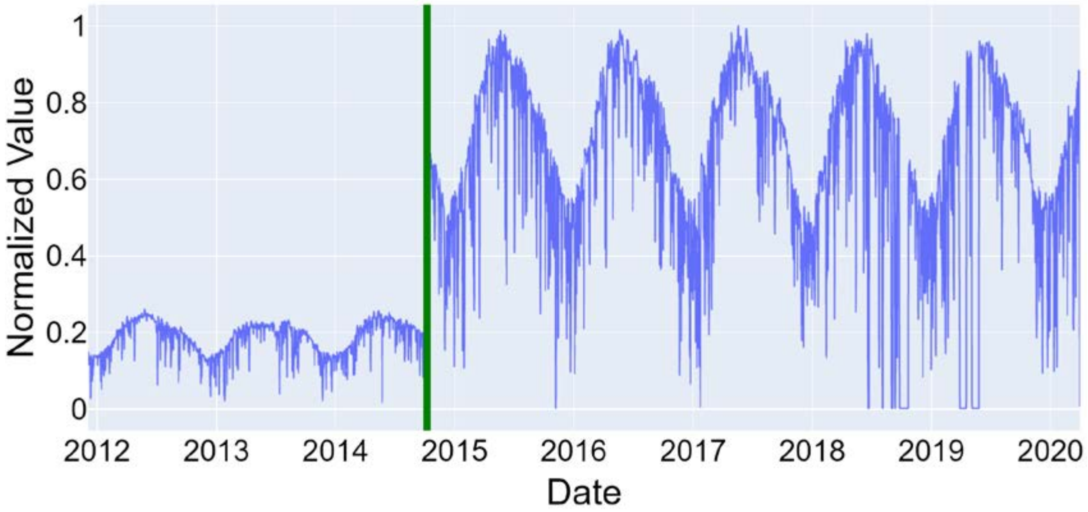
NATIONAL RENEWABLE ENERGY LABORATORY

NREL/PR-5K00-82079

Problem Statement

- Data shifts, like the ones shown to the right, are 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]
- Including data with these shifts in degradation analysis can lead to inaccurate results
- Previous research focusing on 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]
- **Our goal:** Develop a solution that automatically identifies data shift points in PV time series data, and filters out these periods for degradation analysis

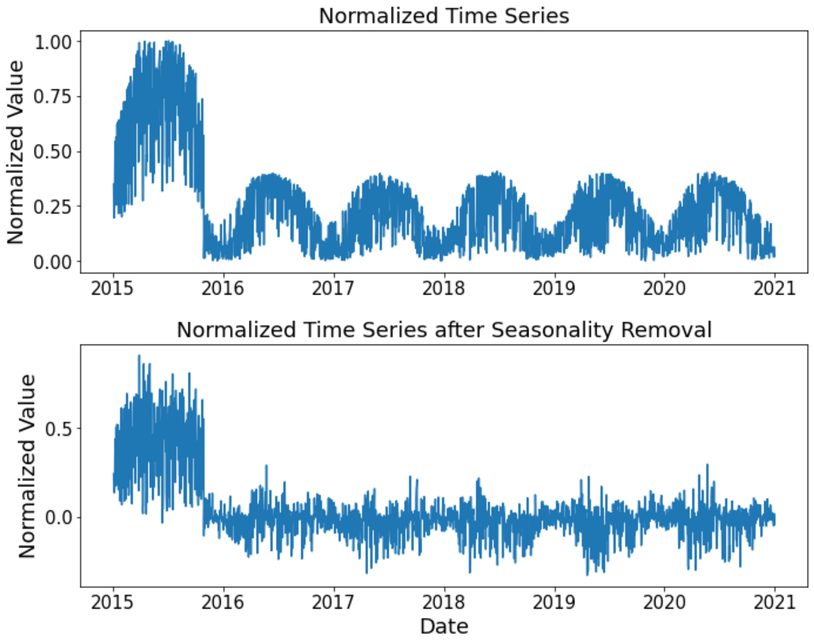




[1] D. C. Jordan, C. Deline, S. R. Kurtz, G. M. Kimball, and M. Anderson, "Robust PV degradation methodology and application."
[2] D. C. Jordan and S. R. Kurtz, "Analytical improvements in PV degradation rate determination."

Methodology

- **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 (see right).
        - A daily median value for each day of the year in the data set was calculated. This value was then subtracted from the normalized time series, based on the day of the year.
- **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
- Run analysis on the longest continuous time series segment without data shifts



Normalized time series pre- and post-seasonality removal (above and below graphics, respectively).

Results

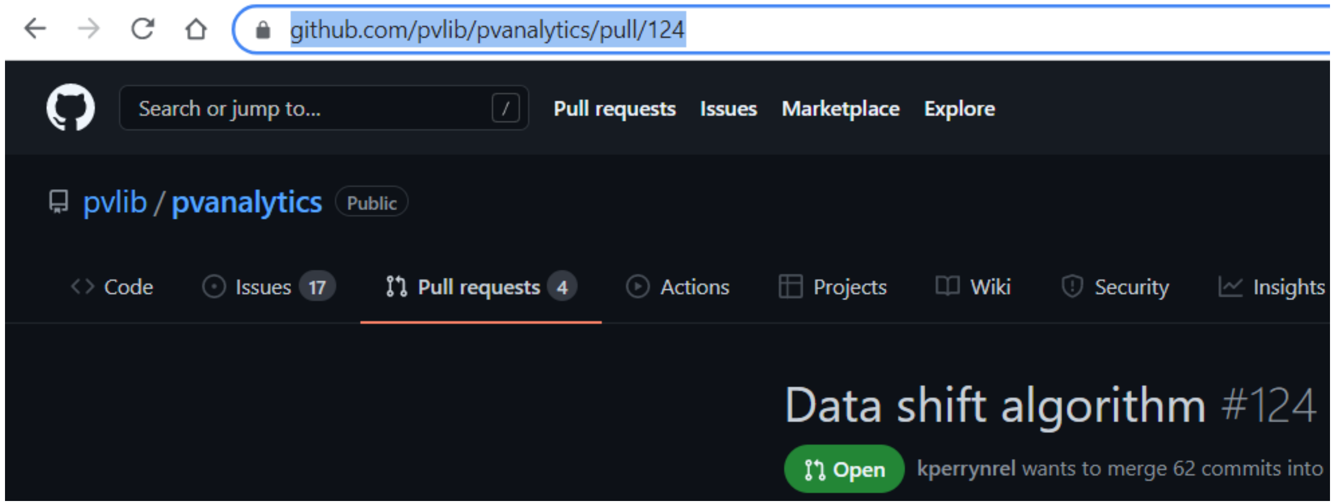
- Each model configuration was run on the 101 labeled data sets, and model performance was benchmarked.
  - F1-Score: Measures the ability of each model to correctly identify data shift points within 30 days of their labelled occurrence.
  - Average run time: Taken across all 101 of the data sets
- Model results shown in the table to the right
  - Best model or 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**

Performance by Model Configuration, using the Labeled Data Sets

Further Research

- Model integration into the PV Analytics package
  - Current PR: https://github.com/pvlib/pvanalytics/pull/124
- Releasing the labeled data sets on the DuraMAT DataHub
- Performing this same process with labeled performance index (PI) data
  - Labeling data sets from the PV Fleets Initiative
  - Tuning and benchmarking changepoint model performance using labeled PI data



Thank you!

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 34348. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government.