NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency and Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC.

Survey of time shift detection algorithms for measured PV data

K. Perry[1], B. Meyers[2], M. Muller[1]

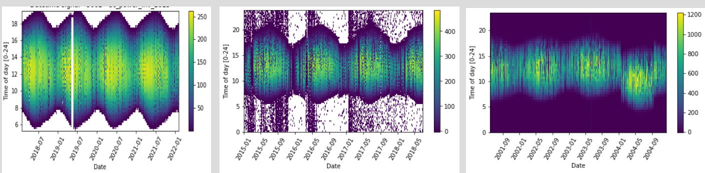
(1) National Renewable Energy Laboratory, Golden, CO 80401, USA

(2) SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

Introduction

Measured PV data from fielded systems may be incorrectly labeled by time zone, erroneously contain daylight savings time (DST), or have time shift issues. Correcting this data is essential for future analysis, including degradation analysis.

This research aims to benchmark three open-source Python automated time shift detection algorithms–two from PVAnalytics and one from Solar-Data-Tools—on ground truth data sets with known time shifts.



AC power time series data heat-mapped by hour of the day. The first time series contains no issues, the second contains DST, and the third contains a random time shift.

Methods

Data Sets

- 8 systems, comprising 19 unique AC power and irradiance streams free of time shifts, were identified in the NREL PVDAQ database
- Data stream were resampled at 1-, 15-, 30-, and 60-minute frequencies
- Time shift issues were randomly inserted into each data stream, including:
  - Erroneous DST (full series or partial series)
  - Random time shifts ranging from 15 minutes off to 120 minutes off
  - Incorrect time zone labeling
- Inserted issues recorded for the purpose of algorithm benchmarking

PVAnalytics-Changepoint Detection (CPD) Approach Steps

- Data pre-processing: removal or flatlined data, negative data, outliers
- Data is run through a day-night masking algorithm to get sunrise and sunset times
- Modeled sunrise and sunset values for the location are determined using PVLib
- The midday time between sunrise and sunset is calculated for modeled and measured data. The measured midday point is subtracted from the modeled midday point, creating a midday time difference series
- The midday time series is run sequentially through a Binary Segmentation CPD algorithm to find time shift points and amounts. Each segment mean is rounded to the nearest 15-minute period (settable)
- Currently being adapted into PVAnalytics (https://github.com/pvlib/pvanalytics)

PVAnalytics-Optimization-Based Signal Decomposition (OSD) Approach Steps

- Same pre-processing steps as PVAnalytics CPD algorithm
- Instead of using a CPD algorithm to find time shifts, an OSD algorithm is used (https://github.com/cvxgrp/signal-decomposition). No rounding used here.

Solar-Data-Tools (SDT) Algorithm Approach Steps

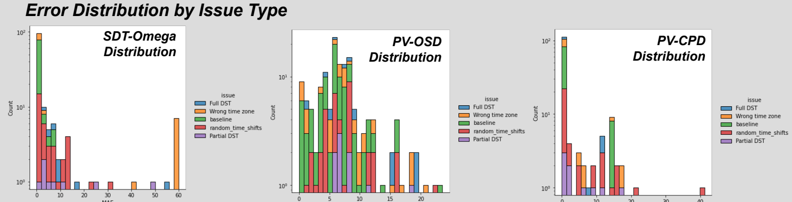
- Preprocessing/cleaning using SDT pipeline (DOI: 10.1109/PVSC45281.2020.9300847)
- Estimate solar noon as avg of sunrise and sunset times
- Apply signal decomposition (SD) with three component classes (DOI: 10.1561/2000000122)
  - Smooth and periodic
  - Piecewise constant
  - Residual term (sum-square small)
- Piecewise constant component is the detector of the shift and provides the estimate of the locations and amounts

Results

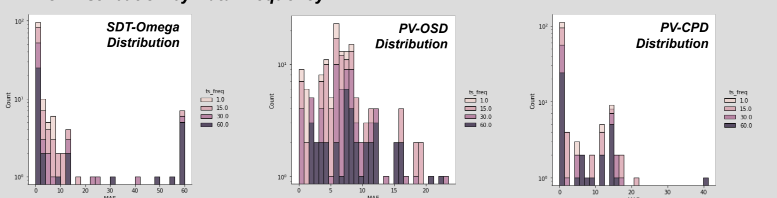
Mean absolute error (MAE) in Minutes by Issue Type

Issue	SDT-Alpha Avg MAE	SDT-Omega Avg MAE	PV-OSD Avg MAE	PV-CPD Avg MAE
Full DST	35.32	13.77	9.57	3.56
Partial DST	22.27	12.76	8.78	5.35
Wrong time zone	18.47	17.84	7.65	1.83
Baseline (no issues)	0.12	0.39	6.35	1.55
Random time shifts	14.02	5.80	7.31	5.03

Error Distribution by Issue Type

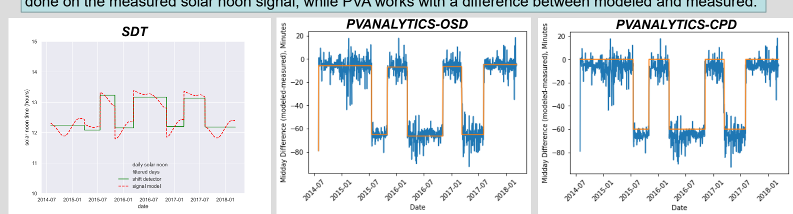


Error Distribution by Data Frequency



PARTIAL DST EXAMPLE: VISUALIZING TIME SHIFTS

Note the differences in the y-axis between the SDT results and PVAnalytics results. SDT shift detection is done on the measured solar noon signal, while PVA works with a difference between modeled and measured.



- ***Differences between algorithm approaches***
  - PV-CPD algorithm rounds to the nearest 15-minute period (settable), SDT and OSD methods don't currently round and can take on any values (this could be altered).
  - For PVAnalytics methods, seasonality in the solar noon signal is removed by utilizing lat and lon to calculate the equation of time (EoT) directly
  - SDT does not require lat/lon as an input and handles the seasonality due to the EoT through an additional component in the SD model

Conclusions

- PV-CPD highest overall performer, but results may be biased due to time series rounding
- All algorithms were marginally less accurate when estimating time shifts in lower frequency data sets (60-minute sampled)
- Preliminary analysis: We plan to create a more advanced benchmarking data set with field-representative issues for validation in the PVInsight Data Hub
- Later testing to include algorithm run time analysis on common compute architecture
- Data sets are generalizable while still allowing us to appropriately improve algorithms

This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08GO28308 and with the National Renewable Energy Laboratory. Associated SETO numbers are 38258 and 38529.

NREL/PO-5K00-85699

NREL PV Module Reliability Workshop, Golden CO, Feb.28-Mar.2, 2023