A Tool for Visual Exploration and Analysis of Solar **Photovoltaic Module Data**

Vassilis Stamatopoulos, Stavros Maroulis, Konstantinos Kozanis, Ioannis Psarros, George Papastefanatos, Giorgos Giannopoulos and Manolis Terrovitis

ATHENA Research Center, Greece

Abstract

Solar photovoltaic (PV) modules are a popular source of clean energy, and effective monitoring and optimization of their performance requires the ability to explore and analyze the measurement data from their sensors. In this work, we introduce a tool for the visual analysis of such data stored in raw measurement files, offering efficient interactive visualization directly on the files. The tool includes a soiling detection module, and novel UI modules that combine time series data visualization with advanced solar panel analytics presentation.

Keywords

Time series Visualization, Visual Analytics, Solar energy, solar panels, soiling

1. Introduction

The use of solar photovoltaic modules is becoming increasingly widespread as a way of generating clean energy. Effective monitoring and optimization of the performance of these modules require analyzing and visualizing the data collected from their sensors. This data, which may often be stored in raw data files (e.g. CSV files), contains timestamped measurements of various numeric variables such as power output and solar irradiance. Such data are usually visualized on time series line charts and intuitive user operations are provided to assist their analysis (e.g. panning, zooming, filtering).

While there are many tools that provide generic time series visualization functionality, they do not address the specific needs of solar PV module data analysis. In particular, the detection of significant events such as rain washes, and the determination of optimal times for manual cleaning to improve panel performance, are critical objectives in the analysis of this data. Integrating these capabilities directly into the visualization would greatly enhance users' ability to analyze the data and optimize the performance of their solar PV systems. Moreover, most visualization tools require the deployment of a fully fledged data pipeline, in which data are first collected from sensors, transferred, loaded and indexed into a central data management system, and finally, prepared for

© 0000-0002-9044-796X (V. Stamatopoulos); 0000-0003-2816-4368 (S. Maroulis); 0000-0003-3983-6449 (K. Kozanis); 0000-0002-5079-5003 (I. Psarros); 0000-0002-9273-9843 (G. Papastefanatos); 0000-0002-8252-9869 (G. Giannopoulos);

0000-0003-0784-8402 (M. Terrovitis) 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 40 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)



visual processing before users can start the visual analysis. This increases the deployment complexity and costs, and limits the efficiency and the capabilities of ad hoc interactive analysis directly over raw data files, that are collected by sensor devices.

Contribution. The contributions of this demo paper are as follows: (1) We present a tool for visual analysis of sensor measurement data from solar photovoltaic modules stored in time series files. (2) It offers scalable visual analytics capabilities without support of an underlying database, through an in-memory multi-level aggregation tree that is built when the user starts a session, and adjusts to user interaction for speeding up performance. (3) It offers various advanced solar panel analytics including a soiling detection visualization functionality. (4) Novel UI modules have been developed to combine time series data visualization with advanced solar panel analytics presentation.

Related Work. Several techniques have been proposed for visually analyzing raw data files without the need for loading or indexing in a DBMS [1, 2, 3, 4]. Similarly, in previous work, aggregation structures have been proposed for the efficient multi-level exploration of numeric and temporal data [5]. Here, we focus on the interactive visualization of multivariate time series from raw files, and propose a multi-level aggregation tree structure that is dynamically constructed and enriched based on user interaction. Our goal is to reduce data file access costs and improve the visualization performance. In addition, we have integrated a solar detection module into our tool, specifically designed for the analysis of photo-voltaic (PV) modules. This module, based on the approach proposed in [6], outperforms current state-of-the-art methods (e.g. [7, 8, 9]) in the measurement of soiling losses in PV modules. Unlike other approaches, this method does

Proceedings of the 6th International Workshop on Big Data Visual Exploration and Analytics co-located with EDBT/ICDT 2023 Joint Conference (March 28-March 31, 2023), Ioannina, Greece



Figure 1: Architecture Overview. PV module measurements are sent in batches to the file system; Users perform visual operations at the Frontend which are evaluated over the aggregation tree which gets enriched with missing data read from the files. Soiling Detection may also be performed and the visualization results are sent back to the front-end.

not require labeled data or generic analytical formulas, but instead relies on a minimal set of available measurements to estimate soiling with high accuracy. Although there are many tools for the visualization and analysis of time series data, to the best of our knowledge, this is the first visual analytics tool specific for the visualization of PV module measurement data that also integrates soiling detection visualization functionality.

2. Tool Overview

Exploration Scenario. The scenario we are examining involves a user who wants to visually analyze monitoring data from multiple PV modules in a solar park. The data consists of multivariate time-series that measure various metrics, including power output, irradiance, and temperature of PV modules, as well as weather parameters such as precipitation. This information is periodically collected and stored in data files ¹, with each file containing measurements from a single module over a specific time period. The user wants to explore and visually analyze data for one or more PV modules using appropriate visualization methods, as well as interact with the visualization by panning, zooming, and filtering. Additionally, they intend to perform more advanced analysis to detect subpar performance of solar panels due to dust accumulation and schedule washings to improve their performance.

Architecture Overview. Fig. 1 illustrates the tool's

architecture. Measurements from each PV module are received in batches and stored in separate folders in the file system **①**. The user selects a module and defines a time period to analyze its measurements 2. The data is parsed and stored in memory, in a tree-based structure, which is partially constructed based on the user request. During the construction of the tree, the results of the first user request are also evaluated 3. The user interacts with the visualization UI through a sequence of visual operations (4), which are evaluated using the tree (5). If necessary, missing measurements are fetched and the tree is updated 6. If the user wants to perform soiling detection analysis, the sensor measurements are read and used to train a regression model and evaluate the soiling index 7. The results are then sent back to the front-end for visualization.

Multi-level Aggregation Tree. The in-memory data structure is a multi-level aggregation tree which keeps at each level aggregated time series data from multiple files based on their timestamps. Each tree level represents a different temporal granularity (e.g. year, month, day, hour), coarser than the granularities of the levels below, and the finest level of granularity in the hierarchy is limited by the sampling frequency. Based on its path from the root, each node corresponds to a specific time period and aggregates the data points within it, storing aggregate values (e.g. mean, min, max, sum) over the time series variables. Since the measurements in the data file are ordered based on their timestamp, each node also stores the file offset of its first data point and the total number of points to fetch the data efficiently, without having to scan the entire file.

¹In our implementation, we consider row-based file formats and specifically CSV files.

Data Preparation and User Operation Evaluation. A user requests to visualize a PV module's measurements by selecting the folder (where the files of the PV modules reside) and providing an initial time period of interest.

To speed up initial processing, we partially build the tree with data from the requested period and enrich it as the user interacts with the data. First, we read the first and last record from each file to determine their temporal range, and then only parse the files with measurements within the time period requested to construct a partial tree. The first request also determines the maximum level of temporal granularity of the initial tree, as well as the level of details for different objects in the data. For instance, if the user requests data for every minute of the last 24 hours, those measurements will be aggregated up to the minute level, while the remaining data will be aggregated at coarser levels of detail.

After initializing the tree, user operations are evaluated by traversing the tree up to the granularity level specified in the query, considering only nodes overlapping with the query time interval. If a node is found that corresponds to the user's requested frequency, its timestamp and aggregate values are returned. If parts of the time series are missing or not pre-aggregated for the requested frequency, the required measurements are read from the corresponding files, aggregated, and the tree is enriched with finer granularity sub-trees.

Soiling Detection Module. The Soiling Detection module uses as input the PV module measurements within the user-specified time period and needs to fetch the corresponding data points from the disk. It may also take as input dates during which the solar park was manually cleaned as potential cleanings, along with raining periods. The module first determines effective cleanings, i.e. the ones with positive effect on energy production. To that end, it checks each potential cleaning for the existence of a changepoint in the behavior of power output as a function of irradiance and PV module temperature. This is implemented as follows: a regression model is trained before the potential cleaning and tested after the potential cleaning; prediction errors are then compared to determine a change of behavior. We apply Ridge Regression with polynomial features. Our choice satisfies a two-fold objective: i) good accuracy and ii) fast fitting time. The latter is vital in our method which fits one model for each potential cleaning. The method proceeds by training a regression model on periods following the effective cleanings. This model aims to capture the optimal expected performance, i.e., power output when the solar panels are clean. Finally, the Soiling Index is returned, which is a performance index calculated by dividing the actual power output by the expected power output for the whole time period.

able on GitHub ², features a front-end developed with TypeScript and React, designed as a single-page application. The back-end, written in Java 11, interacts with the front-end through a REST API.

3. User Interface

The tool's user interface is shown in Fig. 2. The UI layout is a Dashboard where the user can visually explore and analyze the time series sensor data of one or more PV modules. The basic features include:

PV Module Sensor Data Visual Exploration. The tool allows users to select a specific PV module by selecting it from a list of available modules and visually explore its data on the chart (a). They can analyze the sensor data in a sampling frequency of their choice (a) and select to visualize different sensor variables from the Measures Panel on the left (c). Also, they can temporally explore the data by panning left or right, or zooming in and out on the chart. To speed up the visualization process on large data, we utilize the in memory tree described in Section 2, whenever possible, instead of reading the raw data files.

Chart Control Panel. From this panel (), the user is able to perform several practical operations on the data depicted on the chart. These operations are enumerated from left to right. **a**) Manual selection of the visualized date range; **b**) compare sensor data with another PV module in the farm. For example, the user can visualize the power of two different PV modules, on the same chart, to detect potential under-performance issues; **c**) select between different chart types (line, candle); Finally, **d**) the user can change the chart view from stacked, where the sensor variables are visualized on different charts; to overlay, where sensor variables are visualized in the same chart.

Soiling Detection. The PV Dashboard offers a toolkit that provides the user with a set of tools, to help them detect interesting patterns in the sensor data. One such tool, is the Soiling Detection tool (**b**), which includes a toggle to visualize the manual washes on the chart or the potential rains that acted as washing events, and were detected by the Soiling Detection Module. The user can also specify in the chart intervals in which they know that a certain washing event took place. The various washing event intervals are highlighted with different colors based on their type as shown in Fig. 2 (rains are colored blue, manual washing events with green, and user-highlighted washes with yellow).

To help users analyze washing events detected by the algorithm, a slider can be used to filter events based on

Implementation Details. This open-source tool, avail-

²https://github.com/MORE-EU/more-visual



Figure 2: Solar PV module Dashboard

their score. When the user enables event visualization, the visualization changes and displays the power and precipitation sensor measurements. The Soiling Detection Module can then be enabled with input parameters including the washing events and the time range the panels are considered clean after each wash. Finally, the Soiling Index is displayed on top of the chart, allowing users to view estimated power loss per measurement or aggregated power loss in a specific time range.

Filtering. By utilizing this tool, users can specify range filters for the sensor variables, and easily filter out any data points that fall outside those ranges **①**. When a filter is applied to a certain variable, the corresponding measurements are filtered from the visualization of all variables. For example, a user may want to visualize the power of a PV module, where the irradiance in a specified period was above a certain threshold. This way, they can easily distinguish outliers, i.e. data points with low power despite the high irradiance. This, in combination with the Soiling Index visualization, can help them spot anomalies in the PV module's activities.

4. Demonstration Outline

In this section, we outline our demonstration scenario. The tool is available at:

http://visualize.more2020.eu/visualize/solar/eugene.

The attendees will be able to interact with the tool and analyze Solar Panel sensor data. The main focus of the demonstration is the Soiling Detection module and the visualization of its results as described in Section 3.

The datasets used for the visualization are provided in [10]. These real-world data come from PV modules in two locations, namely Cocoa, Florida, USA and Eugene, Oregon, USA. The sampling interval of the data is 5 minutes, and measurements are provided for all hours of daylight. The sensor variables provided for visualization include, the maximum power of the module in watts, the amount of solar irradiance in watts per square meter received on the module's surface, the modules back-surface temperature and the accumulated daily total precipitation.

Users will be able to interact with the prototype and perform several operations such as: **a**) Interact with the chart to pan, zoom in/out to find areas of interest and filter the visualized datapoints. **b**) Select different measures (sensor data) from the Measures Panel. **c**) Visualize Washing Events and apply the Soiling Detection module's algorithm on these events.

Users will also perform specific analytical tasks to better understand the tool's capabilities. For instance, if a solar park operator wants to determine the best time to perform a manual wash of a solar panel to minimize power loss due to soiling, they can perform Soiling Detection on one of the solar panel datasets. Initially, they can visualize the relevant period of interest on the chart by performing pan and zoom operations, and select the types of washing events to view in the Soiling Detection tool. Then, they can enable Soiling Detection to visualize the Soiling Index above the chart. Using this information, they can analyze how each washing event affected the power loss of the panel. For example, if the Soiling Index increases after a washing event, the user can infer that the washing was effective.

Subsequently, the user can filter the detected washing events to keep only the most significant rains, and detect their parameters, such as duration and periodicity, that contributed to the cleaning of the panel. This information can be used to schedule future manual washes of the panel. For example, if the latest Soiling Index is low and a similar washing event is not expected in the near future, the operator can schedule a manual washing of the panel.

Acknowledgement: The authors were partially supported by the EU's Horizon 2020 Research and Innovation programme, under the grant agreement No. 957345: "MORE".

References

- S. Maroulis, N. Bikakis, G. Papastefanatos, P. Vassiliadis, Y. Vassiliou, Resource-aware adaptive indexing for in situ visual exploration and analytics, The VLDB Journal (2022) 1–29.
- [2] N. Bikakis, S. Maroulis, G. Papastefanatos, P. Vassiliadis, In-situ Visual Exploration over Big Raw Data, Information Systems 95 (2021).
- [3] S. Maroulis, N. Bikakis, G. Papastefanatos, P. Vassiliadis, Y. Vassiliou, Adaptive indexing for in-situ visual exploration and analytics, in: DOLAP, 2021.
- [4] G. Papastefanatos, G. Alexiou, N. Bikakis, S. Maroulis, V. Stamatopoulos, Visualfacts: A platform for in-situ visual exploration and real-time entity resolution., in: EDBT/ICDT Workshops, 2022.
- [5] N. Bikakis, G. Papastefanatos, M. Skourla, T. Sellis, A hierarchical aggregation framework for efficient multilevel visual exploration and analysis, Semantic Web 8 (2017) 139–179.
- [6] A. Kalimeris, I. Psarros, G. Giannopoulos, M. Terrovitis, G. Papastefanatos, G. Kotsis, Data-driven

soiling detection in pv modules, IEEE Journal of Photovoltaics (2023) 1–6. doi:10.1109/JPHOTOV. 2023.3243719.

- [7] A. Massi Pavan, A. Mellit, D. De Pieri, The effect of soiling on energy production for large-scale photovoltaic plants, Solar Energy 85 (2011) 1128–1136. doi:10.1016/j.solener.2011.03.006.
- [8] F. A. Mejia, J. Kleissl, Soiling losses for solar photovoltaic systems in california, Solar Energy 95 (2013) 357–363. doi:10.1016/j.solener.2013.06.028.
- [9] M. Muller, L. Micheli, A. A. Martinez-Morales, A method to extract soiling loss data from soiling stations with imperfect cleaning schedules, IEEE Explore, 2017, p. 2881–2886. doi:10.1109/PVSC. 2017.8366214.
- [10] B. Marion, A. Anderberg, C. Deline, J. del Cueto, M. Muller, G. Perrin, J. Rodriguez, S. Rummel, T. J. Silverman, F. Vignola, R. Kessler, J. Peterson, S. Barkaszi, M. Jacobs, N. Riedel, L. Pratt, B. King, New data set for validating pv module performance models, IEEE Xplore, 2014, p. 1362–1366. doi:10.1109/PVSC.2014.6925171.