Solar Curtailment Dataset Information

# Monthly D-PV Time Series Dataset

## Summary

|  |  |
| --- | --- |
| **File format name** | processed\_unsw\_YYYYMM\_data\_raw.csv |
| **Sample file name** | processed\_unsw\_201907\_data\_raw.csv |
| **Number of columns** | 7 |
| **Number of rows (exluding column name)** | 22,243,579 for the sample file. Could be a bit more or less |
| **File size** | 1.2 GB so pd.read\_csv will take around 1-2 minutes |
| **Available data** | 201907 until 202004 |

## Columns

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Name** | **Type** | **Description** |
| 0 | c\_id | int | circuit id, there are 500 circuit’s data in this dataset. Circuit means D-PV System |
| 1 | utc\_tstamp | str | universal time stamp, eg ‘2019-07-11 05:11:55.000’. Needs to be converted to local time (Adelaide time) because all circuit sites are in Adelaide (GMT +9:30) |
| 2 | energy | int | energy produced, but will not be used in curtailment analysis because we will calculate energy from integrating the power data |
| 3 | power | float | average power in watt |
| 4 | reactive\_power | int | average reactive power in VVAr |
| 5 | voltage | float | average voltage in volt |
| 6 | duration | int | duration between consecutive timestamps, in seconds. Possible values are 5s and 60s in this dataset. |

## Other Notes

* This dataset is from Solar Analytics, a company in Australia which is a partnership in this project.
* The size is big because one file contains a time series data with 60 or 5 seconds resolution for all 500 sites
* For the daily solar curtailment analysis, we will filter this dataset for a certain site for a certain date

# Site Details

## Summary

|  |  |
| --- | --- |
| **File name** | unsw\_20190701\_site\_details.csv |
| **Number of columns** | 7 |
| **Number of rows (exluding column name)** | 500 |
| **Size** | 27.5 KB |

## Columns

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Name** | **Type** | **Description** |
| 0 | site\_id | int | site id |
| 1 | s\_postcode | int | site postcode location |
| 2 | pv\_install\_date | str | date of pv system installation. After certain date, it is mandatory for a pv system to have vvar and vwatt response according to AS standards. |
| 3 | ac\_cap\_w | float | inverter ac rating in watt, which is the maximum power produced by the inverter |
| 4 | dc\_cap\_w | float | pv array power rating in watt, which is the total wattpeak of all modules installed. If it is too much higher than the ac\_cap\_w, the power could be curtailed. |
| 5 | inverter\_manufacturer | str | inverter brand |
| 6 | inverter\_model | str | inverter model |

## Other Notes

* The D-PV data from Solar Analytics are obtained by installing a monitoring tool after the PV system has been installed. However, some of the tools are installed with a reversed polarity (installation issue). So, the power and reactive power polarity are reversed. In that case, polarity = -1. Normal installation has polarity = 1.
* The inverter should absorb the reactive power of the grid in the afternoon, so if the reactive power value is positive, it is most likely due to wrong polarity (polarity = -1)
* This data will be merged with circuit\_details data to provide a complete information about the circuit. Most relevant data for the curtailment analysis include the site\_id, c\_id, and ac\_cap\_w.

# Circuit Details

## Summary

|  |  |
| --- | --- |
| **File name** | unsw\_20190701\_circuit\_details.csv |
| **Number of columns** | 4 |
| **Number of rows (exluding column name)** | 500 |
| **Size** | 15.8 KB |

## Columns

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Name** | **Type** | **Description** |
| 0 | site\_id | int | site id |
| 1 | c\_id | int | customer id |
| 2 | con\_type | str |  |
| 3 | polarity | int | the polarity of the monitoring tool. Can be 1 (normal) or -1 (reversed) |

## Other Notes

* The D-PV data from Solar Analytics are obtained by installing a monitoring tool after the PV system has been installed. However, some of the tools are installed with a reversed polarity (installation issue). So, the power and reactive power polarity are reversed. In that case, polarity = -1. Normal installation has polarity = 1.
* The inverter should absorb the reactive power of the grid in the afternoon, so if the reactive power value is positive, it is most likely due to wrong polarity (polarity = -1)

# Unique c\_id and site\_id

## Summary

|  |  |
| --- | --- |
| **File name** | UniqueCids.csv |
| **Number of columns** | 3 |
| **Number of rows (exluding column name)** | 499 |
| **Size** | 11.8 KB |

## Columns

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Name** | **Type** | **Description** |
| 0 | Unnamed: 0 | int | just an index, 0-498 |
| 1 | c\_id | int | circuit id number |
| 2 | site\_id | int | site id number |

## Other Notes

* In this dataset, each c\_id corresponds with one unique site\_id, kind of like bijective function (one-one). It is not really clear why we don’t just use either only c\_id or only site\_id. Probably it is just an anticipation for case, where one site can have more than 1 circuit, which never happens in this case.
* This data is used to map from c\_id to site\_id or vice versa.

# Monthly GHI Data

## Summary

|  |  |
| --- | --- |
| **File format name** | sl\_023034\_YYYY\_MM.txt  Note: sI means solar irradiance, 023034 is the station number |
| **Sample file name** | sl\_023034\_2019\_01.txt |
| **Number of columns** | 36, but most are not relevant for curtailment analysis |
| **Number of rows (exluding column name)** | 40,320 data from the sample file. |
| **Size** | 11.1 MB |
| **Available data** | 2019\_01 until 2020\_07 |

## Relevant Columns

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Name** | **Type** | **Description** |
| 2 | Year Month Day Hours Minutes in YYYY | int | year |
| 3 | MM | int | month |
| 4 | DD | int | day |
| 5 | HH24 | int | hour |
| 6 | MI format in Local standard time | int | minute. The hour and minute is already in local time, so it is not needed to convert the timezone like the D-PV time series data |
| 7 | Mean global irradiance (over 1 minute) in W/sq m | str | the mean global irradiance in 1 minute long |

## Other Notes

* We use only this data, which includes 1 observation station only, for the Solar Curtailment analysis. It means, we assume all 499 sites from Solar Analytics data experience the same irradiation profile, which sometimes can be inaccurate. It is possible for a site to experience shading from cloud even though the ghi profile from this dataset is clear in the date because the observation station may be a bit far from the analyzed site.
* Global irradiance include the direct and diffuse irradiance, or we can say global irradiance is the total irradiance received by a certain site.
* We will use this ghi profile mainly to check whether a certain date is a clear sky day (without cloud) or not, and also showing the ghi plot.

# Sample D-PV data for a certain site for a certain date

## Summary

|  |  |
| --- | --- |
| **File format name** | data\_sample\_N.csv  Note: N is the sample number. |
| **Sample file name** | data\_sample\_1.csv |
| **Number of columns** | 7 |
| **Number of rows (exluding column name)** | 1433 for the sample file. Could be a bit more or less |
| **File size** | 78.5 KB |
| **Available data** | 14 samples are available on 14 Sept 2022 |

## Columns

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Name** | **Type** | **Description** |
| 0 | Timestamp | str | Timestamp already in Adelaide time (GMT +9.30), eg:  ‘2019-09-03 11:21:55+09:30’ |
| 1 | c\_id | int | circuit id, there are 500 circuit’s data in this dataset. Circuit means D-PV System |
| 2 | energy | int | energy produced, but will not be used in curtailment analysis because we will calculate energy from integrating the power data |
| 3 | power | float | average power in watt |
| 4 | reactive\_power | int | average reactive power in VVAr |
| 5 | voltage | float | average voltage in volt |
| 6 | duration | int | duration between consecutive timestamps, in seconds. Possible values are 5s and 60s in this dataset. |

## Other Notes

* This dataset is obtained by filtering the Monthly D-PV Time Series Dataset for a certain site and certain date. The timestamp, however, is already converted from utc to Adelaide time (GMT +9:30).
* We have made 14 samples, with some important samples:
  1. Tripping Curtailment – Non clear sky day: sample 1
  2. Tripping Curtailment – Clear sky day: sample 11
  3. VVAr Curtailment: sample 14
  4. VWatt Curtailment: sample 4
  5. Incomplete dataset: sample 5
  6. Clear sky day without curtailment: sample 9

# Sample GHI data for a certain date

## Summary

|  |  |
| --- | --- |
| **File format name** | ghi\_sample\_N.csv  Note: N is the sample number. |
| **Sample file name** | ghi\_sample\_1.csv |
| **Number of columns** | 36, but most are not relevant for curtailment analysis |
| **Number of rows (exluding column name)** | 1440 data from the sample file. |
| **Size** | 416 KB |
| **Available data** | 14 samples are available on 14 Sept 2022 |

## Relevant Columns

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Name** | **Type** | **Description** |
| 2 | Year Month Day Hours Minutes in YYYY | int | year |
| 3 | MM | int | month |
| 4 | DD | int | day |
| 5 | HH24 | int | hour |
| 6 | MI format in Local standard time | int | minute. The hour and minute is already in local time, so it is not needed to convert the timezone like the D-PV time series data |
| 7 | Mean global irradiance (over 1 minute) in W/sq m | str | the mean global irradiance in 1 minute long |

## Other Notes

* We use only this data, which includes 1 observation station only, for the Solar Curtailment analysis. It means, we assume all 499 sites from Solar Analytics data experience the same irradiation profile, which sometimes can be inaccurate. It is possible for a site to experience shading from cloud even though the ghi profile from this dataset is clear in the date because the observation station may be a bit far from the analyzed site.
* Global irradiance include the direct and diffuse irradiance, or we can say global irradiance is the total irradiance received by a certain site.
* We will use this ghi profile mainly to check whether a certain date is a clear sky day (without cloud) or not, and also showing the ghi plot.