“J’veux du soleil” Towards a decade of solar irradiation data (La Réunion Island, SW Indian Ocean)

Mathieu Delsaut[1], Patrick Jeanty[1], Béatrice Morel[1], and Dominique Gay[2]

[1] LE2P-EA4079, Université de La Réunion, France
[2] LIM-EA2525, Université de La Réunion, France
firstname.name@univ-reunion.fr

Abstract. This paper aims at presenting years of solar irradiation data together with meteorological data acquisition localized in the French region of La Réunion Island (SW Indian Ocean). The publicly available data take the form of multivariate time series data with one-minute sampling rate over eight years – with still ongoing acquisition. We also present typical analytics tasks that are related to solar energy application domain as well as general time series analytics tasks that are suitable for these data. Thus, we aim at drawing the attention of the time series data mining community to these valuable data.

Keywords: solar irradiation data, multivariate time series data, open data

Preamble. This paper is a *resource track paper*. Its aim is mainly to describe an innovative data set to *(i)* support research on the solar irradiation topic; *(ii)* to potentially suggest novel evaluation tasks; *(iii)* to encourage novel methods and/or algorithms. The concerned data set is already available under reasonably liberal terms and we hope sufficiently well-documented. We also suggest some open research and valuable applications.

1 Introduction

The European Union long-term climate strategy aims to be climate-neutral by 2050, i.e., an economy with net-zero greenhouse gas emissions [3]. As one of the most remote regions of EU, La Réunion island, a French overseas department, has implemented a EU-consistent multiannual energy program which main goal is *electric energy self-sufficiency* in the horizon 2030 [6]. Indeed, as an island (SW Indian Ocean), its isolated position prevents from being interconnected with the metropolitan power grid – thus leading to high dependency from fossil fuels for electricity production.

However, as a tropical island, La Réunion presents a high potential of renewable energies: besides biomass and wind exploitation, solar resource has attracted much attention in the past decade. In order to estimate the solar energy potential of the island, a local research team has led a solar resource research programme for the use of solar resource as a stable source of energy and to ensure its management in a reliable and efficient way for its integration into an electrical power grid [4].

In this paper, we present the data acquired during the last decade through the research programme. Sensors used for solar irradiation and meteorological parameters measurement are presented in Section 2. Section 3 is dedicated to the full description of the publicly available data. We discuss the main domain applications using the available data in Section 4 before concluding.

2 Data Acquisition

In order to obtain a representative view of the solar energy potential, several measurement stations has been spread out over the island. More precisely, six stations have been installed, mainly on EDF [3] power plants sites located along the leeward coast of the island which also concentrates most of the inhabitants (see Figure 1):

1. Moufia (*reference station* of the Université de La Réunion, Saint-Denis)
2. Bois de Nèfles (EDF site, Saint-Denis)
3. Saint-André (EDF site, Saint-André)
4. Port-Est (EDF site, Le Port)
5. Saint-Leu (EDF site, Saint-Leu)
6. Saint-Pierre (EDF site, Saint-Pierre)

Fig. 1. Localisation of the 6 measurement stations over La Réunion.

[3] Electricité De France

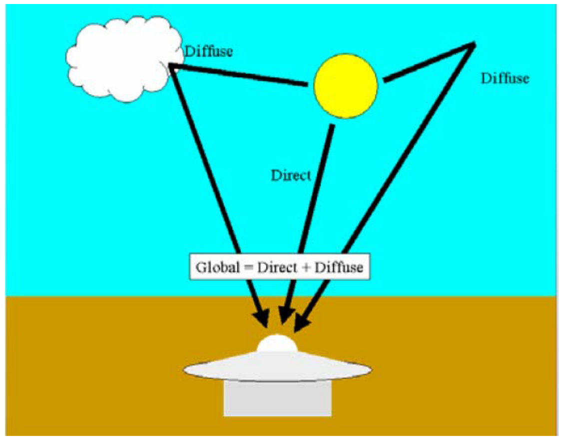


Fig. 2. Global, direct and diffuse components of solar irradiance.

2.1 Measuring irradiation and meteorological parameters

Each station has roughly the same set-up (hardware and software) that can be split into three main parts: the measuring tools, the control system and the supporting structure. Here, we focus on the measuring tools while more technical setup are available in appendix A.

Measuring tools - Ground-based measurements come from two sensors:

– 1 **pyranometer** SPN1 (manufacturer: Delta-T Devices) which simultaneously measures:
  - the **GHI** (Global Horizontal Irradiance in W/m2)
  - and the **DHI** (Diffuse Horizontal Irradiance in W/m2).

  These two components of solar irradiation are illustrated in Figure 2.
  Finally, the **BHI** (Beam Horizontal Irradiance in W/m2) may be easily obtained by difference of the previous two.

$$GHI = BHI + DHI$$

  One can also compute the **DNI** (Direct Normal Irradiance in W/m2) by introducing the zenith angle Θ:

$$DNI = \frac{BHI}{\cos \Theta}$$

– 1 **weather transmitter** WXT520 (manufacturer: Vaisala) which measures five meteorological parameters:
  - air temperature ($^{\circ}C$),
  - atmospheric pressure (Pa),
  - relative humidity (%),
  - wind speed (m/s)
  - and wind direction ($^{\circ}$)

Fig. 3. Main measurement tools: (left) SPN1 pyranometer and (right) WXT520 weather transmitter.

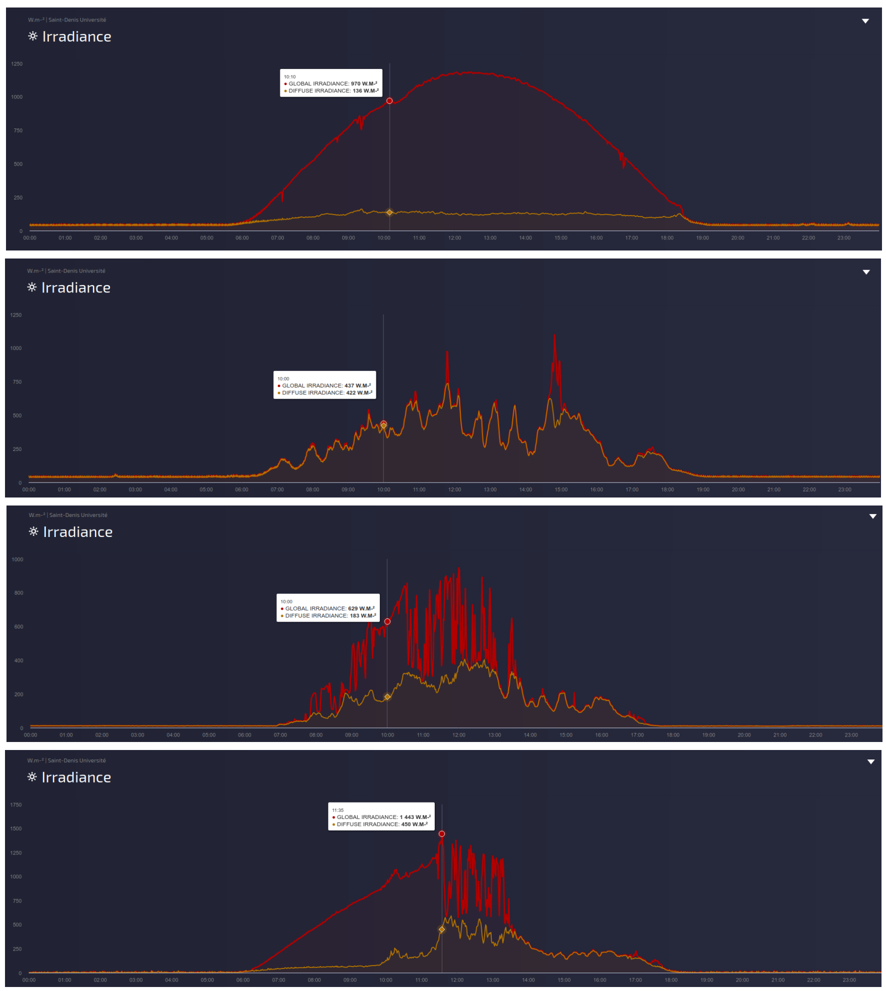
3 Available Data

The six stations have been installed in 2012. Thus, eight years of historical data are now available. Considering the seven parameters, (GHI and DHI, plus 5 meterological parameters), the data takes the form of multivariate time series (MTS). As the sampling rate is one minute, the MTS data contains more than 150 millions data points. A short extract in csv format is shown in Figure 4.

Timestamp	FD_Avg	FG_Avg	Patm_Avg	RH_Avg	Text_Avg	WD_MeanUnitVector	WS_Mean
...							
01/05/14 10:20	72.32	687.6	980.2	58.26666	26.78333	308.8434	0.5833333
01/05/14 10:21	71.7	688.1	980.2	56.83333	26.9	330.7189	0.9333333
01/05/14 10:22	71.36	673.7	980.2	56.76666	26.91666	310.8201	0.6833333
01/05/14 10:23	72.95	666.3	980.1333	57.01666	27.0	345.3426	2.216666
01/05/14 10:24	75.5	672.0	980.0999	56.66667	26.93333	355.3273	2.233333
01/05/14 10:25	78.67	680.8	980.0999	57.56667	26.76666	340.6307	1.216667
01/05/14 10:26	76.8	683.2	980.0999	57.96666	26.7	318.039	0.6166667
01/05/14 10:27	75.27	689.0	980.0999	57.75	26.71667	23.57079	1.083333
01/05/14 10:28	74.08	692.7	980.0166	58.25	26.8	23.79196	0.65
01/05/14 10:29	72.89	710.5	980.0333	58.4	26.8	24.55521	0.9
01/05/14 10:30	72.1	709.6	980.0	57.46666	26.8	341.5136	1.416667
...							

Fig. 4. Multivariate data sample in csv format for a short 10-minutes period on may 1st, 2014. Timestamp (date and time), FD_Avg (diffuse), FG_Avg (global), Patm_Avg (atmospheric pressure), RH_Avg (relative humidity), Text_Avg (Temperature), WD_MeanUnitVector (wind direction) and WS_Mean (wind speed).

Typical irradiation data shapes - Given the tropical climate of the island, at least four typical days might be observed when regarding GHI and DHI parameters. We illustrate these cases in Figure 5: (a) a sunny day is identified by a hill-form GHI curve (with the maximum around noon) and a high difference with DHI almost flat curve; (b) a cloudy day when GHI is confounded with DHI; (c) intermittency (e.g., frequent cloud pass) is characterized by a high variability in GHI; and (d) a sequence of the three previous phenomenons.



(a) (b) (c) (d)

Fig. 5. GHI (red) and DHI (orange) evolution during four typical days : (a) sunny, (b) cloudy, (c) intermittent then cloudy, (d) sequence of sunny, intermittency, cloudy periods.

4 Valuable Applications

Solar energy domain - As pointed out in [4], using solar resource as an electrical and stable source of energy is not an easy task. It raises three scientific and technical issues that are still open:

1. the prediction of solar irradiation despite its variability
2. the management of the solar photovoltaic production through storage systems to reduce the impact of the intermittent nature of the solar resource
3. the integration of the solar resource into a power grid in order to meet the local energy needs and to cope with the load fluctuations

The first problem, the prediction of solar irradiation, is particularly important as it is a precondition for the success of the two others. From a data science point of view, it may directly be formalized as a forecasting problem when considering either short, mid or long term prediction (e.g., 1 hour, 6 hour or a day ahead forecasting).

General Machine Learning / Data Mining - Besides daily clustering or short, mid, long term forecasting of solar irradiation, which correspond to core ML/DM tasks for MTS, one can benefit from this large-scale MTS data source for the evaluation of other *classical* ML/DM tasks like, e.g.:

- correlation analysis between irradiation parameters and meteorological parameters (more generally, between the various dimensions of MTS)
- outlier detection / extreme values analysis for MTS
- MTS missing values imputation,
- MTS data compression
- similarity-based query optimisation
- advanced visualisation techniques for large-scale MTS data

Solar irradiation data: a challenging data set - A two-year piece of these data (2014-2015) has been suggested as an open challenge data set for the annual data mining challenge of the 2018 French data mining and management conference (EGC 2018) [2]. For this open challenge, two selected and award-winning papers have addressed the irradiation data clustering and prediction problems:

- Per, Dalleau and Smail-Tabbone [5] has explored the challenge data in multiple ways: Prior exploratory data analyses have enabled the statistical comparison of characteristics of cities with respect to the measured weather variables (diffuse and overall solar fluxes, atmospheric pressure, moisture, temperature, wind speed and direction). Data was preprocessed and univariate time-series and multivariate time-series aggregated over hours or days were analyzed in order to build simple and effective prediction models. A classical clustering approach was performed. Groups of days sharing weather parameters in common were found by two biclustering algorithms. The characterisation of found biclusters and their succession displayed in a calendar-based visualization tool have helped assess their interest.
- Bruneau, Pinheiro and Didry [1] focus on short-term prediction, i.e., the prediction of solar irradiance one hour ahead. The authors have tested the value of using recently observed data as input for prediction models, as well as the performance of models across sites. After a data cleaning and normalization pre-processing step, they combine a variable selection step based on AutoRegressive Integrated Moving Average (ARIMA) models, to end up with general purpose regression techniques such as neural networks and regression trees.

5 Conclusion

In this data paper, we have presented eight years of solar irradiation data together with meteorological data. The raw data takes the form of multivariate times series with one-minute sampling rate. While the data acquisition is still active, we think the already available data is a valuable support for advanced analytical tasks such as daily solar irradiance clustering or forecasting, as well as other general analytical tasks. Thus, we hope to see many papers using these data in the future.

Acknowledgments - Publicly available data come from recent successive research projects, RCI-GS and GeoSun, with financial support from Europe, Regional Reunion Island Council and the French government through the ERDF (European Regional Development Fund).

Permanent link to La Réunion Island solar irradiation Data -

`https://doi.org/10.5281/zenodo.3898530`

References

1. Bruneau, P., Pinheiro, P., Didry, Y.: Prédiction du rayonnement solaire par apprentissage automatique. In: Lebbah, M., Largeron, C., Azzag, H. (eds.) Extraction et Gestion des Connaissances, EGC 2018, Paris, France, January 23-26, 2018. RNTI, vol. E-34, pp. 439–450. Éditions RNTI (2018)
2. Défi EGC 2018 : Un défi sous le soleil de l'île de la Réunion (2018), `https://www.egc.asso.fr/manifestations/defi-egc/defi-egc-2018.html`, *(in french)*
3. European Union – Climate strategies and targets – 2050 long-term strategy (2020), `https://ec.europa.eu/clima/policies/strategies/2050_en`
4. Goujon Person, M., Delage, O., Bessafi, M., Chabriat, J.P., Jeanty, P.: Solar energy research and development program on the exploitation of the solar resource on the Réunion island and its integration into an electrical power grid. In: Third Southern African Solar Energy Conference (SASEC 2015). Skukuza, South Africa (May 2015)
5. Per, Y., Dalleau, K., Smaïl-Tabbone, M.: Exploration et analyses multi-objectifs de séries temporelles de données météorologiques. In: Lebbah, M., Largeron, C., Azzag, H. (eds.) Extraction et Gestion des Connaissances, EGC 2018, Paris, France, January 23-26, 2018. RNTI, vol. E-34, pp. 427–438. Éditions RNTI (2018)
6. Région Réunion - multiannual energy program – programmation pluriannuelle de l'énergie (ppe) (2020), `http://www.reunion.developpement-durable.gouv.fr/programmation-pluriannuelle-de-l-energie-ppe-r336.html`, *(in french)*

A Technical setup

Data logging - The control system contains the following equipment (see Figure 6):

- 1 waterproof IP64 enclosure (manufacturer: Legrand), size 400 mm x 300 mm x 250 mm, in fiberglass or coated metal.
- 1 datalogger CR1000 (manufacturer: Campbell Scientific), a programmable device that handles sensors measurements, drives communication and stores data and programs.
- 1 Ethernet interface NL 115 (manufacturer: Campbell Scientific) with memory module.
- 1 CompactFlash memory card 500 Mb

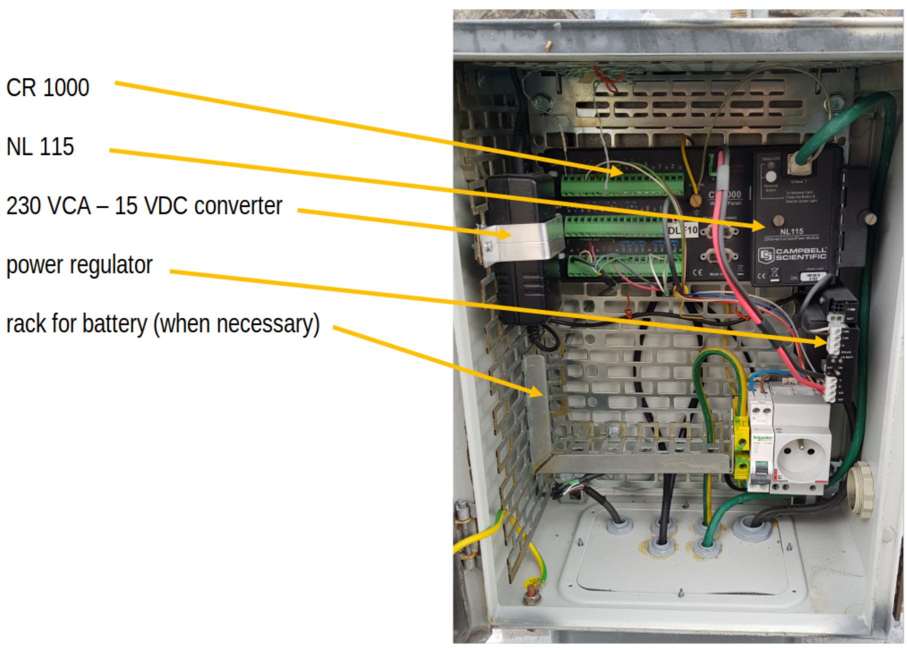


Fig. 6. Equipment of a control system unit.

The station may be powered through 230 VCA grid or by renewable energy using solar panel and battery. Communication between station and server may be ensured by a GPRS link (in that case, a modem with a SIM card needs to be added in the local cabinet in connection with the CR1000 logger).

Supporting structure - Ground-based stations network hosts two types of stations depending of environmental conditions. In EDF power plant sites, a 10-meter high foldable mast (see Figure 7) is used as well as a metallic enclosure.

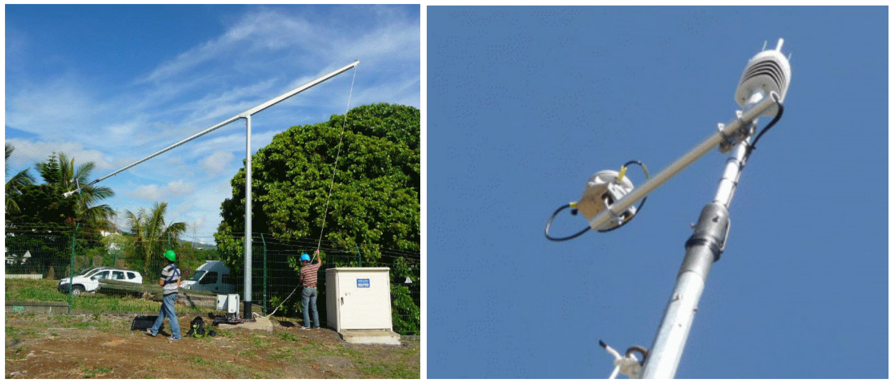


Fig. 7. 10-meter high foldable mast supporting a measurement station.

In other places, a compact and easily transportable station is used (see Figure 8). In any case, the whole station is designed and fabricated to bear cyclones winds up to 250 km/h.

Fig. 8. A compact and easily transportable station setup.