



**HAL**  
open science

## Experimental validation of a novel methodology for fast an accurate analysis of solar energy yields based on cluster analysis

Luis Guerreiro, Carlos M Fernández-Peruchena, Afonso Cavaco, Martín  
Gastón, Manuel Collares Pereira

### ► To cite this version:

Luis Guerreiro, Carlos M Fernández-Peruchena, Afonso Cavaco, Martín Gastón, Manuel Collares Pereira. Experimental validation of a novel methodology for fast an accurate analysis of solar energy yields based on cluster analysis. Proceedings of the ISES EuroSun 2016 Conference, Oct 2016, Palma de Mallorca, Spain. 10.18086/eurosun.2016.09.03 . hal-02380050

**HAL Id: hal-02380050**

**<https://hal.science/hal-02380050v1>**

Submitted on 26 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## ISES EuroSun 2016

### Experimental validation of a novel methodology for fast and accurate analysis of solar energy yields based on cluster analysis

Luis Guerreiro<sup>1</sup>, Carlos M. Fernández-Peruchena<sup>2</sup>, Afonso Cavaco<sup>1</sup>, Martín Gastón<sup>2</sup> and Manuel Collares Pereira<sup>1</sup>

<sup>1</sup> Renewable Energy Chair, University of Evora, Palacio do Vimioso, P - 7002 Evora

<sup>2</sup> CENER, Department of Solar Thermal Energy, E – 41092 Sevilla

#### Abstract

The design and optimization of solar power systems requires a detailed knowledge of the dynamic behavior of the meteorology at the site of interest, usually assessed by one typical meteorological year (TMY). Even with today's technology, the computational effort to simulate solar energy system performance with one year of data at high frequency (as 1-min) may become colossal if a multivariable optimization has to be performed. This work evaluates a methodology based on cluster analysis for selecting number of individual days able to represent the long-term performance of a solar energy system. This procedure permits to drastically reduce computational effort related to the calculation of a solar power plant energy yield by simulating a reduced number of days from a TMY, facilitating a fast and optimal design of the plant.

Keywords: Cluster, Day selection, STE (Solar Thermal Electricity), Optimization

---

#### 1. Introduction

Concentrating Solar Thermal Power (CSTP) technologies collect and concentrate the Direct Normal Irradiance (DNI) incident on the earth's surface to heat a working fluid and afterward producing electricity through a thermodynamic cycle. The mean power produced by a CSTP system often determines the viability of the Solar Power Project, which is calculated from energy yield simulation computer programs requiring weather data input to drive the optical and thermal models within the simulation tools.

The feasibility analysis of a CSTP plant project requires the estimation of the amount of energy that it will be able to deliver to the grid during its lifetime. A detailed knowledge of the Direct Normal Irradiation (DNI) is a critical point to perform an economical feasibility analysis of these plants, as it is the most determining variable in the final energy yield. This knowledge comprises a deep understanding of the availability and dynamics of the solar resource over different temporal scales (intraday, day, month, year) (Vignola et al., 2012). To achieve this, long time series of meteorological data are usually condensed into shorter series that typify the meteorological conditions at a specific location with the aim of data volume reduction and speeding-up of energy system simulations, considered as representative of the location of interest.

Currently, the typical year approach is by far the most widely used among the scientific and industry communities for the design and optimization of solar power plants. These datasets hold 8,760 hourly meteorological values that typify conditions at a specific location throughout a longer period. Accordingly, simulations are usually performed using one hour simulation steps, either with steady state models or somehow

taking into account system dynamics to more accurately simulate the performance of the plant. Also, worst cases of solar resource availability are often analyzed in order to ensure the viability of the project. Frequently, these worst cases are evaluated by probability of exceedance scenarios, which provide an estimate of electric generation output that the solar resource consultant expects to be exceeded in any given year over the life of the debt with 90% confidence (P90) and 99% confidence (P99) (C.M.F. Peruchena et al., 2016).

TMY approach seems to be appropriate for rather detailed simulations of a specific solar power plant. Notwithstanding, DNI exhibits a great variability at high-frequency due to the dynamic effects of passing clouds, which is a key issue in the design and performance analysis of a CSTP system, given its nonlinear response to DNI caused by various thermal inertias due to their complex response characteristics (Fernández-Peruchena et al., 2015). DNI series of sub-hourly resolution (in the range of 1 to 10 minutes) permit an accurate modeling and analysis of transient processes in some CSTP technologies. For example, parabolic troughs with direct steam generation could be particularly sensitive to the cloud transients (Eck and Hirsch, 2007; Montes et al., 2009), and, therefore, better analyzed using DNI series with frequency resolutions much lower than 1-h (Meyer, 2010). This kind of simulations provides more accurate representation of the plant performance at the cost of larger computational effort leading to longer simulation times. Thanks to the powerful computers currently available, these transient simulations are being used not only for detailed simulations of the short term performance of the plant to improve the design and operation of specific aspects of the system, but also for the energy yield estimation over long periods of the plant (e.g. year simulation or even multi-year simulations) for a specific final plant design.

During the optimization of the large number of plant parameters before a final design is reached, a huge number of simulations are required. The computational effort to simulate a CSTP plant performance with 1-year series at high frequency (as 1-min) may become colossal if a multivariable optimization has to be performed. To overcome this problem, several approaches have been presented for reducing or condensing the data, as the Short Reference Year (SRY) (Hallgreen, 1983), which divides the calendar year into periods. However, in this approach, the selection of the representative days is left to the designer expertise, or they are selected emblematic days like the solstices. More recently, a novel approach based on cluster analysis has been proposed to optimally select individual days able to represent meteorological conditions at a site (Carlos M. Fernández Peruchena et al., 2016). The goal of this contribution is to apply this methodology in a location with high-quality and long term validated measurements, in order to evaluate if selected days can represent the performance of a CSTP plant throughout the complete year.

## 2. Experimental data and method

### 2.1. Experimental data

A 12-year series of hourly solar irradiance and meteorological parameters measured in Évora, Portugal by the University of Évora has been selected for this study. Table 1 below presents details about the location of the station, its pyrheliometer and the climatic conditions of the station location, according to the Köppen-Geiger classification (Peel et al., 2007).

**Tab. 1: Radiometric station selected for this study.**

Location (Country)	Coordinates (°)	Altitude (m)	Pyrheliometer	Period	Climate (Köppen classification)
Évora (Portugal)	38.567 N 7.911 W	290	Kipp&Zonen CHP1	2003-2015	Mediterranean (Csa)

The measured data set covers a 12-years period. The data has been recorded every 5 seconds and averaged every 10 minutes. DNI is measured with a Kipp&Zonen Pyrheliometer (model CHP1), a WMO First Class Pyrheliometer with an associated estimated uncertainty at daily scale of <1%. The pyrheliometer is calibrated every 2 years according to the standard ISO9059:1990.

## 2.2. Energy yield simulations

A 50 MWe Fresnel type power plant without thermal-energy storage and with a primary mirror area of 193.000 m<sup>2</sup> has been designed for Évora, Portugal. The software TRNSYS V.17 ([www.trnsys.com](http://www.trnsys.com)) has been used for the estimation of hourly values of the energy produced by this system. TRNSYS is a validated, time-series simulation program that can simulate the performance of photovoltaic, concentrating solar power, water heating systems, and other renewable energy systems using hourly resource data (Wagner, J. M. and Gilman, P., 2011).

## 2.3. Day selection

In this work, a TMY series is generated from measured series available at the location. For selecting specific days of this yearly series, the following parameters have been calculated to characterize individual days: daytime mean temperature; daily cumulative DNI value; sum of the product of DNI · cos (SZA), being SZA the solar zenith angle; a wind-dependent variable that is zero if maximum daytime wind speed is lower than 12 m/s, 2 if percentile 90 of daytime wind speed is higher than 12 m/s, and 1 in any other case.

Cluster analysis comprises a wide variety of techniques for delineating natural groups (clusters) in data sets, so that observations within each group are similar to one another with respect to variables or attributes of interest and the groups themselves stand apart from one another. This technique is useful for summary and discovery:

- *Summary* refers to the reduced representation of the full data;
- *Discovery* refers to finding out patterns into the original data series.

In this work, the objective of applying cluster analysis is summarizing the data, by means of selecting a small number of days to represent the complete year.

Finally, the clustering method used in this work is the K-medoids algorithm, which selects points analyzed as the center of the clusters. Also, the gap statistic method (Tibshirani et al., 2001) is used for estimating the number of clusters (groups) in the daily series. The idea behind this technique is to find a benchmark (i.e., a *null distribution*) and to compare the observed value of  $W$  (the within-cluster variability) with the expected value under the null distribution (being large deviations from the mean evidence against the null hypothesis).

## 2.4. Method relevance

The aim of this method, using groups of data as clusters is to enable a faster analysis of a wide range of options when designing a Solar Plant. With today's computer processing capability this might seem outdated. However, engineers that design big plants usually do not have plenty of time to run multiple scenarios and in multiple locations. For that reason, reducing the computing time for one set of data for a specific location is an advantage specially in the pre-design phase. With the cluster analysis the computing effort in time is reduced in a magnitude of 15 to 20 times, thus enabling many other alternatives and configurations to be studied.

### 3. Results

#### 3.1. Optimal number of days

In fig. 1 it is represented the gap statistics for the TMY at Évora, where 3 number of clusters cases are highlighted: 15 days (blue), 18 days (green) and 22 days (orange). These cases represent the maximum value of the gap functions in 5-clusters intervals (10-15 clusters; 15-20 clusters; 20-25 clusters), and they are candidates to be the optimal number of clusters to represent the TMY.

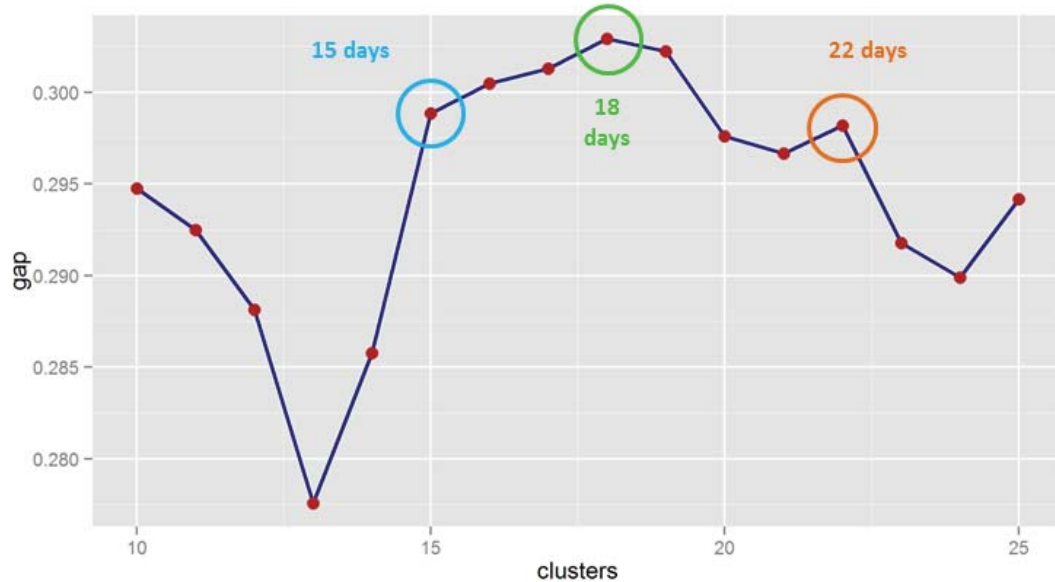


Fig. 1: Gap statistics for day selection

In the analysis carried out, the medoid of each cluster is selected as the most representative days, and also each of the 365 days of the TMY are assigned to one of these clusters. Table 2 shows the 18 selected days selected for representing the TMY at the site (this is expected to be the optimal case, as it reaches the highest Gap value), their DNI value and the number of days represented. This table also shows the maximal and average dissimilarity (max\_diss and av\_diss, respectively) between the observations in the cluster and the cluster's medoid, as well the *isolation* of the clusters: the maximal dissimilarity between the observations in the cluster and the cluster's medoid, divided by the minimal dissimilarity between the cluster's medoid and the medoid of any other cluster.

Tab. 2: Cluster parameters for the TMY representation

Specific days selected of TMY (1-365)	Daily DNI (kWh/m <sup>2</sup> )	Number of days represented	max_diss	av_diss	Isolation
301	1.55	23	2.60	1.56	0.98
2	2.97	28	2.45	1.31	0.93
328	6.23	25	4.00	1.45	1.76
17	4.71	20	6.34	1.88	1.64
8	0.97	15	2.38	1.21	0.68
326	5.81	28	2.64	1.17	1.16
276	3.88	20	2.84	1.54	1.18
60	7.89	19	2.98	1.56	1.09
282	2.46	11	2.70	1.96	0.76

103	10.34	16	3.15	1.92	0.69
86	1.72	7	2.36	1.03	0.98
289	6.76	23	2.91	1.65	1.13
186	9.72	31	2.54	1.38	1.13
287	7.22	17	2.83	1.33	1.10
165	11.34	20	3.14	1.38	1.40
228	4.50	14	3.36	1.97	0.67
237	9.22	26	4.35	1.93	1.24

In fig. 2 it is plotted the selected days, representing its own typical DNI value varying along a 24h period. Clusters are numbered from 1 to 18, and also color-coded. As expected in a Mediterranean climate, most days of summer are represented by clusters 14-18 (characterized by high DNI values as well as long days), and most of winter days are represented by clusters 1-5 (characterized by shorter days and higher DNI variability).

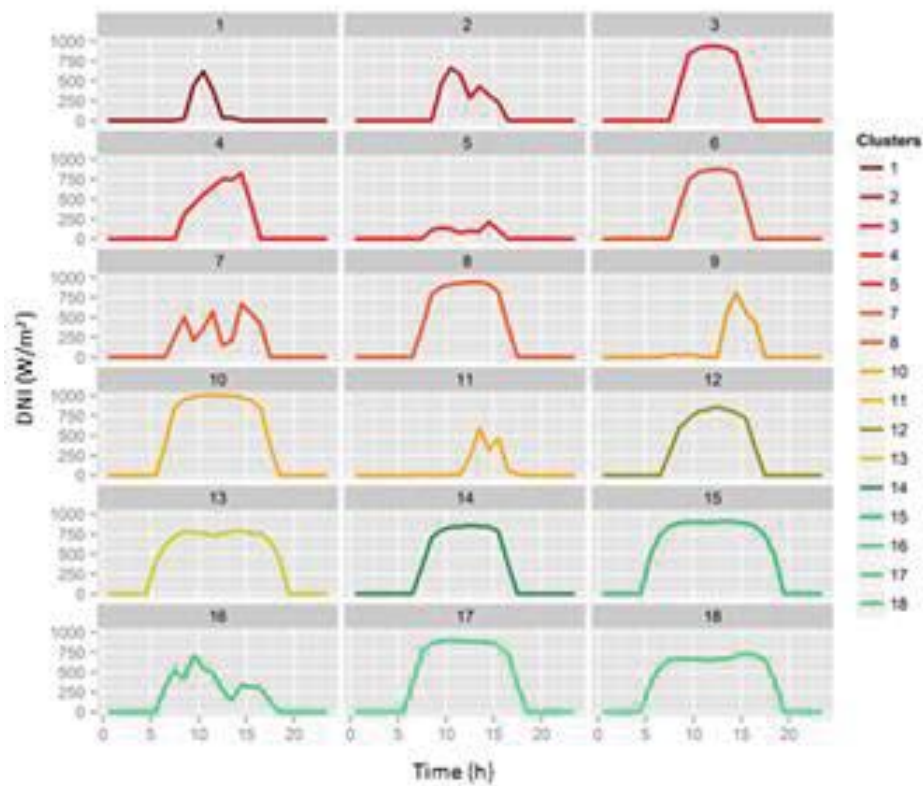


Fig. 2: left, days selected representative of the 18 clusters calculated for the TMY at Évora

This variability across the year can be observed in fig. 3 where which of the 18 typical days are represented throughout the whole year.

From fig. 3, it is clear that from November to February the representation is mainly obtained with clusters 1 to 6, while March, April and October are transition months with clusters 7 to 12 representing most of the period.

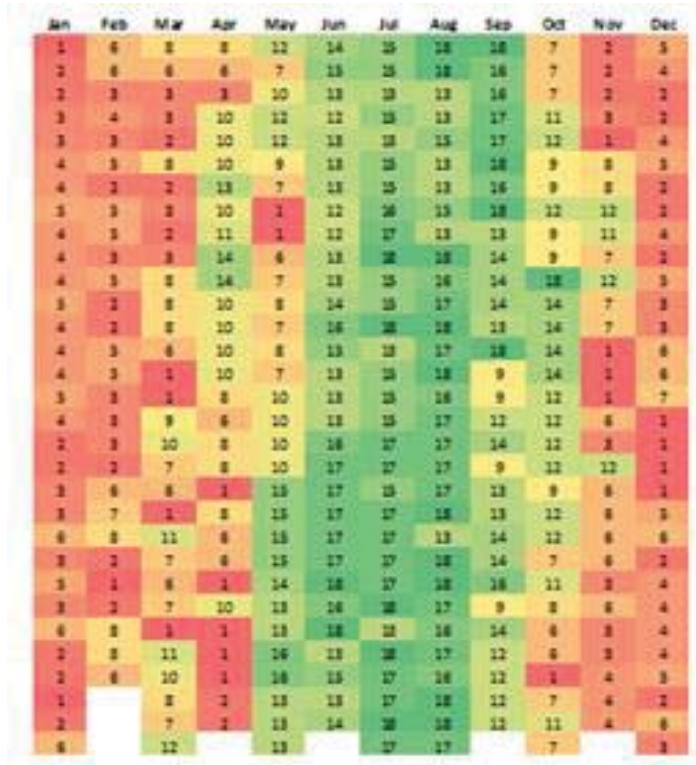


Fig. 3: Days throughout the year (January to December) represented by the selected days (#18 days)

The comparison of the annual energy yield provided by the days selected, multiplied by the number of days represented by each one of the cluster days, has been compared with the solar energy yield of the 365 days of the TMY, and their differences are shown in fig. 4 as a function of the number of clusters selected (from 10 to 183 days). These differences tend to decrease at increasing number of days selected, as expected. Very low differences (<1%) are found even for the selection of few days (15 or less).

Peaks are found in almost the whole range analyzed, as expected as a trend, the amplitude decays with an increasing number of days.

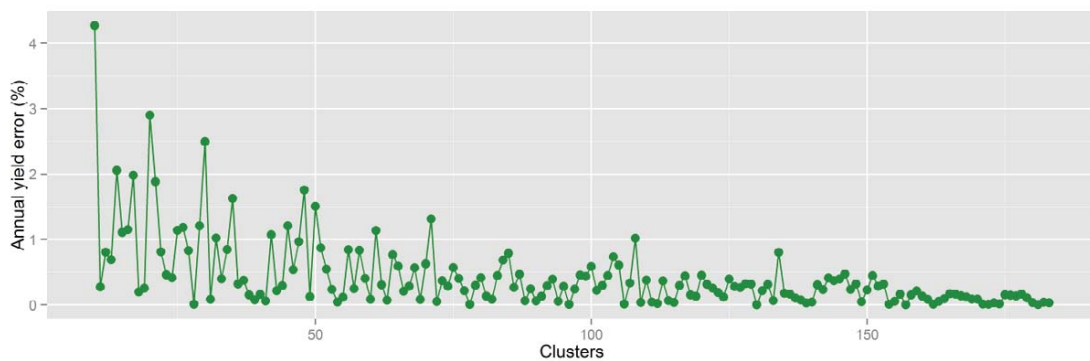


Fig. 4: Difference between annual energy yield calculated with clusters and TMY, as a function of the number of clusters selected

Even if a univocal relation between Gap function and accuracy of annual energy yield cannot be fully established, it has been found a more clear correlation by grouping clusters into intervals. Table 3 shows the optimal number of days selected as a function of the maximum number of days allowed in the cluster analysis, in 5-days intervals. It is shown the difference between TMY and days-selected electric yield, as a function of the maximum number of clusters allowed, showing a clear correlation between Gap function and accuracy in yield estimation.

**Tab. 3: Results of cluster analysis as a function of the maximum number of clusters allowed**

Maximum number of days allowed for the cluster	15	20	25
Optimal number of days selected	15	18	22
Gap function value	0.2959222	0.2996370	0.2949644
Difference between electric yield production using TMY data and using cluster selected days	1,1%	0,2%	0,8%

#### 4. Conclusions

The methodology presented in this work allows a very significant reduction in the computational effort related to the calculation of a STE plant energy yield by simulating a reduced number of days from a high frequency TMY (90% of computing time can be reduced if 18 days are selected for representing the TMY). These days are the medoids of clusters selected according to gap statistic.

The energy yield produced by these selected days facilitates a fast and accurate, evaluation of the annual energy production of a STE plant. Concurrently, in the optimization phase of a solar plant design, many more configurations can be explored using these 18 days than using the whole TMY production analysis (with the same computing cost / time).

The computational effort of this evaluation depends on the uncertainty allowed: a tradeoff between inaccuracy (which is given at low numbers of clusters selected) and computational effort (which is given at low numbers of clusters) must be carried out in each specific analysis as it was shown for the specific location of Evora, Portugal where a STE research facility has been built, the EMSP (Evora Molten Salt Platform).

#### 5. Acknowledgments

The authors would like to thank to ICT-Uni.Evora (Prof. Paulo Canhoto and Prof. Heitor Reis) for enabling the usage of solar radiation collected data, to project DNI approved by IPES (Portuguese Solar Energy Institute), as well as to the H2020 Program which funded Project PreFlexMS, the trigger for the study presented.

#### 6. References

Eck, M., Hirsch, T., 2007. Dynamics and control of parabolic trough collector loops with direct steam generation. *Sol. Energy* 81, 268–279. doi:10.1016/j.solener.2006.01.008



- Fernández-Peruchena, C.M., Blanco, M., Gastón, M., Bernardos, A., 2015. Increasing the temporal resolution of direct normal solar irradiance series in different climatic zones. *Sol. Energy* 115, 255–263. doi:10.1016/j.solener.2015.02.017
- Hallgreen, L., 1983. Short Reference Year, SRY, in: Palz, W. (Ed.), *Solar Radiation Data, Solar Energy R&D in the European Community Series F*. Springer Netherlands, pp. 40–48.
- Meyer, R., 2010. Recommendations for bankable meteorological site assessments for solar thermal power plants, in: *Proceedings of the SolarPACES Symposium*. Perpignan, France.
- Montes, M.J., Abánades, A., Martínez-Val, J.M., 2009. Performance of a direct steam generation solar thermal power plant for electricity production as a function of the solar multiple. *Sol. Energy* 83, 679–689. doi:10.1016/j.solener.2008.10.015
- Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrol Earth Syst Sci* 11, 1633–1644. doi:10.5194/hess-11-1633-2007
- Peruchena, C.M.F., García-Barberena, J., Guisado, M.V., Gastón, M., 2016. A clustering approach for the analysis of solar energy yields: A case study for concentrating solar thermal power plants, in: *AIP Conference Proceedings*. Presented at the SOLARPACES 2015: International Conference on Concentrating Solar Power and Chemical Energy Systems, AIP Publishing, p. 70008. doi:10.1063/1.4949155
- Peruchena, C.M.F., et al, 2016. A statistical characterization of the long-term solar resource: Towards risk assessment for solar power projects. *Sol. Energy* 123, 29–39. doi:10.1016/j.solener.2015.10.051
- Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63, 411–423. doi:10.1111/1467-9868.00293
- Vignola, F., Grover, C., Lemon, N., McMahan, A., 2012. Building a bankable solar radiation dataset. *Sol. Energy, Progress in Solar Energy* 3 86, 2218–2229. doi:10.1016/j.solener.2012.05.013
- Wagner, J. M., Gilman, P., 2011. *Technical Manual for the SAM Physical Trough Model*.