SolarPACES 2023, 29th International Conference on Concentrating Solar Power, Thermal, and Chemical Energy Systems

Solar Resource Assessment

https://doi.org/10.52825/solarpaces.v2i.778

Published: 28 Aug. 2024

Statistical Analysis of Annual DNI Distribution and its Impact on Bankability Assessment for Concentrated Solar Power Plants

Is the Annual DNI Consistent With a Weibull Distribution?

M .Larrañeta[1]*, J. A. López-Álvarez[1], E. Pérez-Aparicio[1], J M. Delgado-Sanchez[2], J. B. Alonso-Montesinos, [3], and M A. Silva-Pérez.[1]

[1] Dpt. Of Energy Engineering, University of Seville. Spain

[2] Dpt. Of Applied Physics, University of Seville. Spain

[3] Dpt. Of Chemistry and Physics, University of Almería. Spain

*Correspondence: Larrañeta, M, mlarraneta@us.es

Abstract. The Weibull distribution is commonly accepted as the most suitable model for describing the annual distribution of Direct Normal Irradiance (DNI). However, when the annual DNI is assumed to follow a Normal distribution instead of a Weibull distribution, there is a notable increase in the cumulative annual values for unfavourable and worst-case scenarios. In this research, we assess the suitability of different statistical indexes for goodness-of-fit by applying them to the annual cumulative DNI and Global Horizontal Irradiance (GHI) data recorded at six locations with varying climates. Our observations reveal that, across all locations, the Representative Solar Year (RSY) aligns with the 50% Probability of Exceedance (PoE50) of a Normal distribution fitted to the observed data. We quantify that assuming the annual DNI conforms to a Weibull distribution, as opposed to a Normal distribution, results in a substantial decrease of approximately 7% in the annual cumulative value for the worst-case scenario. Based on our analysis, we find no compelling evidence to reject the hypothesis that the annual DNI follows a Normal distribution.

Keywords: DNI, Distribution, Bankability Assessment

1. Introduction

Analyzing the yearly distribution of direct solar radiation can provide valuable insights into the patterns and trends of solar radiation at a specific location. The financing of solar power systems typically relies on a statistical assessment of the solar resource, which involves evaluating average, adverse, and worst-case scenarios [1], expressed as the Probability of Exceedance at 50% (PoE50), 90% (PoE90), and 99% (PoE99) respectively. These scenarios, which complement percentile scenarios, are crucial for assessing the bankability of Concentrated Solar Power (CSP) projects.

Understanding the variability in annual solar radiation is crucial for a wide range of applications, including solar energy generation, agriculture, and climate modeling. However, there is currently no consensus in the scientific community regarding the annual distribution of

Direct Normal Irradiance (DNI). While the Weibull distribution is generally accepted as the best fit for the annual DNI distribution, further examination is warranted. Fernandez-Peruchena et al. [2] conducted an extensive evaluation of the statistical properties of annual DNI and Global Horizontal Solar Irradiation (GHI) data from 13 locations. They employed a battery of four goodness-of-fit tests [3-6] to assess how well the DNI data conformed to a Weibull distribution and the GHI to a Normal distribution. The results of these statistical tests were reported in terms of p-values. Fernandez-Peruchena et al. did not reject any null hypotheses for any of the locations because all the obtained p-values exceeded 0.05, suggesting that annual DNI adheres to a Weibull distribution and the annual GHI adheres to a Normal distribution. Based on this finding, the Spanish standard UNE 206013:2017 [7] was developed, assuming that annual DNI follows a Weibull distribution. This standard proposes a method for calculating adverse and worst-case scenarios (PoE90-PoE99) where the annual DNI data is fitted to a Weibull distribution. The cumulative annual values for the bad and worst-case scenarios are subsequently computed based on this fit.

In this study, we aim to quantify the impact of the assumption that annual DNI adheres to a Weibull distribution rather than a Normal distribution when calculating bad and worst-case scenarios following the Spanish standard UNE 206013:2017. Additionally, we revisit the conclusions drawn by Fernandez-Peruchena et al. by adopting a different approach to the use of goodness-of-fit tests. Instead of conducting normality tests for GHI and Weibull distribution tests for DNI, we evaluate both sets of tests for both GHI and DNI time series

2. Methodology

2.1 The selected goodness of fit tests

We have chosen three tests from the existing literature:

- Shapiro-Wilk Test: This test assesses normality in a dataset. It is selected because of its robustness, even with relatively small series sizes. It is important to note that the observed series are relatively short (20-25 values), whereas the synthetic series are significantly longer (1000 values). The test does not guarantee exact normality but provides evidence of whether it is reasonable to assume normality for analytical purposes [8].
- Lilliefors Normal Test: This test is a variation of the Kolmogorov-Smirnov test, specifically designed to evaluate whether a dataset follows a normal distribution. It is particularly well-suited for small sample sizes, typically fewer than 50 values [9].
- Anderson-Darling Test: The Anderson-Darling test is commonly employed to determine if a dataset adheres to a specific distribution, such as the Weibull distribution [10].

2.2 Observed and synthetically generated databases

Goodness-of-fit tests can be influenced by the length of the databases used for analysis. In our study, we worked with observed data collected from six diverse locations, each characterized by different climates (Table 1) [11]. The data was sourced from Solargis climData professional, covering a time span of 20-25 consecutive years. It's important to note that all Solargis parameters are rigorously validated using high-precision meteorological equipment deployed globally for quality control. Table 1 provides essential geographical and climatological details for the chosen sites.

Table 1. *Characterization of selected sites, period of years available and their Köppen-Geiger classification climate.*

Location (ID)	Country	Latitude (ºN)	Longitude (ºE)	Years	Köppen classification climate
Brasilia (BRB)	Brazil	-15.60	-47.71	1999-2020	Tropical Monsoon (Aw)
Boulder (BOU)	United States	40.13	-105.24	1999-2020	Dry Semi-Arid Cold (Bsk)
Tamanrasset (TAM)	Algeria	22.79	5.53	1994-2020	Dry Arid Desert Hot (Bwh)
Goodwin Creek (GCR)	United States	34.25	-89.87	1999-2020	Temperate No dry season hot summer (Cfa)
Toravere (TOR)	Estonia	58.26	26.46	1994-2020	Continental No dry season Warm Summer (Dfb)
Seville (SEV)	Spain	37.41	-6.01	2000-2022	Temperate Dry Hot Summer (Csa)

Additionally, we generated synthetic data for our analysis. This involved fitting both Weibull and Normal distributions to the observed annual GHI and DNI datasets, which allowed us to determine the key parameters defining each distribution. For the Normal distribution, these parameters are the mean (μ) and standard deviation (σ), while for the Weibull distribution, they are the shape (k) and scale (λ) parameters. Subsequently, we randomly generated 1000 values using these parameters, thereby creating an extended dataset of annual GHI and DNI values for the selected locations, fitting both Normal and Weibull distributions. Table 2 provides a summary of the observed and synthetically generated databases.

Table 2. *Observed and synthetically generated databases.*

Variable	Type of data	Fit	Length
GHI	Observed	-	20 to 25 values
	Synthetic	Normal	1000 values
	Synthetic	Weibull	1000 values
DNI	Observed	-	20 to 25 values
	Synthetic	Normal	1000 values
	Synthetic	Weibull	1000 values

3. Results and Evaluation

3.1 Impact of the distribution assumption on the calculation of bad and worst-case scenarios

In this section, we computed the Representative Solar Year (RSY) for the Direct Normal Irradiance (DNI) based on the Spanish standard UNE 206011:2014 [12]. Additionally, we determined the bad and worst-case scenarios in line with the Spanish standard UNE 206013:2017, which assumes that the annual DNI adheres to a Weibull distribution. We calculated the Probability of Exceedance at 90% (PoE90) and 99% (PoE99) in two scenarios, one where the annual DNI is assumed to follow a normal distribution and the other assuming a Weibull distribution. The results of these calculations are presented in Table 3.

Table 3. *RSY and PoE values of the annual DNI obtained for the selected locations applying the said Spanish standards assuming either a Normal and a Weibull distribution.*

Annual DNI (kWh/m^2)		Normal fit			Weibull fit		
Location	**RSY**	**PoE 50**	**PoE 90**	**PoE 99**	**PoE 50**	**PoE 90**	**PoE 99**
Sevilla	2107	2107	1977	1871	2119	1936	1713
Toravere	1012	1012	906	820	1021	880	731
GoodwinCreek	1672	1672	1573	1492	1680	1540	1381
Boulder	2079	2079	1961	1865	2092	1954	1795
Brasilia	1923	1923	1798	1695	1935	1768	1580
Tamanrasset	2441	2441	2322	2225	2451	2274	2072

It's important to note that in all the locations, the Representative Solar Year (RSY) aligns with the Probability of Exceedance at 50% (PoE50) when the data is fitted to a Normal distribution. However, when the data is fitted to a Weibull distribution, the PoE50 is higher than the RSY for all the locations. This discrepancy in the distribution type results in considerably lower values for PoE90 and PoE99 when assuming a Weibull distribution. Table 4 displays the percentage difference in the annual cumulative values compared to the reference values obtained in Table 3, where the normal fit was used as a reference.

Table 4. *Percentage difference in the annual cumulative values when fitting data to a Normal distribution instead to a Weibull.*

PoE difference (%)			
Location	**PoE 50**	**PoE 90**	**PoE 99**
Sevilla	0.54%	-2.07%	-7.49%
Toravere	0.81%	-2.93%	-10.83%
GoodwinCreek	0.50%	-2.07%	-7.39%
Boulder	0.65%	-0.33%	-3.75%
Brasilia	0.58%	-1.65%	-6.78%
Tamanrasset	0.41%	-2.05%	-6.88%

Given that all the locations exhibit similar trends, we will focus our graphical evaluation on the Seville location. Figure 1 illustrates the cumulative distribution function (CDF) of the annual DNI for Seville over the period 2000-2022. Additionally, we provide the PoE50, PoE90, and PoE99 values for both, Weibull and Normal fits, as well as the RSY calculated following the aforementioned Spanish standards.

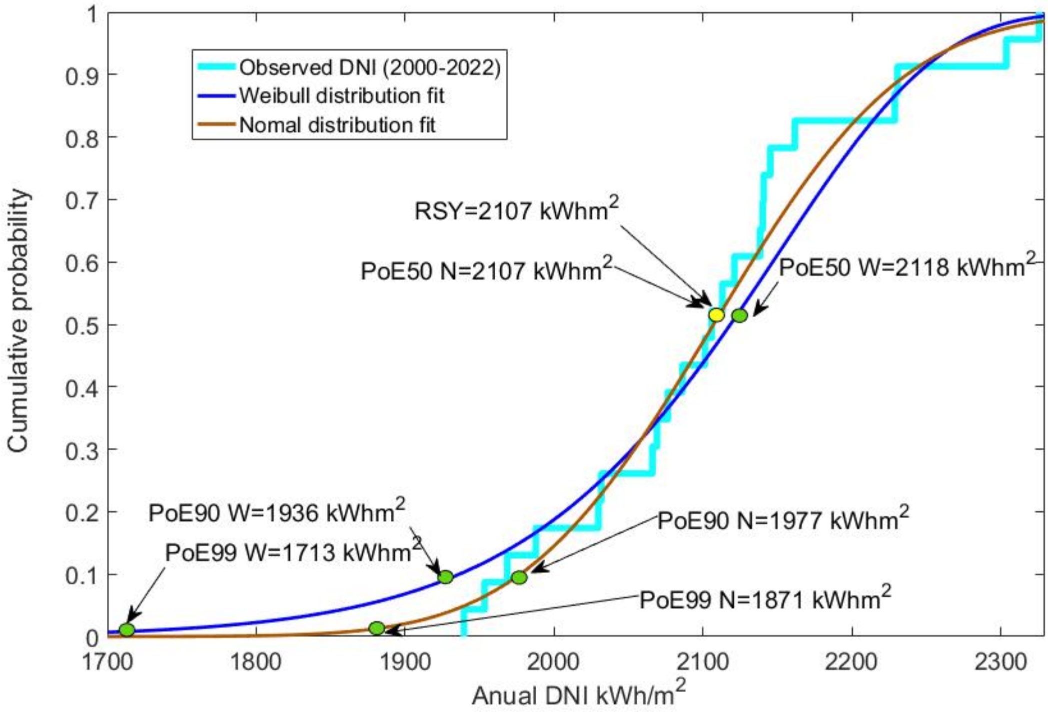

Figure 1. *CDF of the annual DNI for the location of Seville (period 2000-2022). The plot shows in cian the observed data, in blue the Weibull fit and in red the Normal Fit. We also present the P50, P90 and P99 for both fits and the RSY calculated following [12]*

The primary distinction in the cumulative distribution functions (CDFs) arises in the tail of the distribution, resulting in larger differences as the Probability of Exceedance (PoE) increases. For instance, in Seville, the disparities in PoE90 are around 2%, while the differences in PoE99 expand to approximately 7.5%, with the Weibull fit yielding lower values.

3.2 Goodness of fit evaluation

We utilize both h and p-values to quantitatively assess the goodness of fit. These values offer an objective measure of the strength of evidence supporting the null hypothesis, which is the assumption that the annual solar irradiation series conform to a specific distribution (Normal or Weibull).

- Low p-values indicate that the sample provides substantial evidence to reject the null hypothesis for the entire population. In simpler terms, low p-values suggest that the annual solar radiation series do not conform to a particular distribution. In statistics, it is customary to reject the null hypothesis when the p-value is less than 0.05. When the null hypothesis is rejected, the h-value becomes 1.
- High p-values, on the other hand, imply weak evidence against the null hypothesis. In other words, high p-values suggest that the annual solar radiation series do conform to a particular distribution. When the null hypothesis is not rejected, the h-value becomes 0.

Table 5 displays the obtained h and p-values resulting from the selected goodness of fit tests for the six locations with varying climates.

Table 5. *Obtained h and p-values result of the selected goodness of fit tests for the six selected locations with different climates. Calculations are performed for DNI and GHI*

Seville DNI	**Observed Series**		**Normal Distribution fit**		**Weibull Distribution fit**	
	h-value	**p-value**	**h-value**	**p-value**	**h-value**	**p-value**
Shapiro-Wilk (N)	0	0.50	0	0.68	1	0.01
Lilliefors (N)	0	0.31	0	0.66	1	0.01
Anderson D. (W)	0	0.05	1	0.00	0	0.16
Seville GHI	**Observed Series**		**Normal Distribution fit**		**Weibull Distribution fit**	
	h-value	**p-value**	**h-value**	**p-value**	**h-value**	**p-value**
Shapiro-Wilk (N)	0	0.67	0	0.46	1	0.01
Lilliefors (N)	0	0.28	0	0.31	1	0.01
Anderson D. (W)	0	0.22	1	0.00	0	0.18
Tamanrasset DNI	**Observed Series**		**Normal Distribution fit**		**Weibull Distribution fit**	
	h-value	**p-value**	**h-value**	**p-value**	**h-value**	**p-value**
Shapiro-Wilk (N)	0	0.40	0	0.11	1	0.01
Lilliefors (N)	0	0.48	0	0.15	1	0.01
Anderson D. (W)	1	0.02	1	0.00	0	0.35
Tamanrasset GHI	**Observed Series**		**Normal Distribution fit**		**Weibull Distribution fit**	
	h-value	**p-value**	**h-value**	**p-value**	**h-value**	**p-value**
Shapiro-Wilk (N)	0	0.32	0	0.29	1	0.01
Lilliefors (N)	0	0.71	0	0.55	1	0.01
Anderson D. (W)	1	0.01	1	0.00	1	0.02
Toravere DNI	**Observed Series**		**Normal Distribution fit**		**Weibull Distribution fit**	
	h value	**p value**	**h value**	**p value**	**h value**	**p value**
Shapiro-Wilk (N)	0	0.56	0	0.74	1	0.01
Lilliefors (N)	0	0.49	0	0.37	1	0.01
Anderson D. (W)	0	0.16	1	0.00	0	0.55
Toravere GHI	**Observed Series**		**Normal Distribution fit**		**Weibull Distribution fit**	
	h value	**p value**	**h value**	**p value**	**h value**	**p value**
Shapiro-Wilk (N)	0	0.06	0	0.95	1	0.01
Lilliefors (N)	0	0.10	0	0.54	1	0.01
Anderson D. (W)	1	0.00	1	0.00	0	0.09
GoodwinCreek DNI	**Observed Series**		**Normal Distribution fit**		**Weibull Distribution fit**	
	h-value	**p-value**	**h-value**	**p-value**	**h-value**	**p-value**
Shapiro-Wilk (N)	0	0.88	0	0.30	1	0.01
Lilliefors (N)	0	0.97	0	0.75	1	0.01
Anderson D. (W)	0	0.31	1	0.00	0	0.65
GoodwinCreek GHI	**Observed Series**		**Normal Distribution fit**		**Weibull Distribution fit**	
	h-value	**p-value**	**h-value**	**p-value**	**h-value**	**p-value**
Shapiro-Wilk (N)	0	0.75	0	0.37	1	0.01
Lilliefors (N)	0	0.61	0	0.29	1	0.01
Anderson D. (W)	0	0.36	1	0.00	0	0.06
Brasilia DNI	**Observed Series**		**Normal Distribution fit**		**Weibull Distribution fit**	
	h value	**p value**	**h value**	**p value**	**h value**	**p value**
Shapiro-Wilk (N)	0	0.65	0	0.20	1	0.01
Lilliefors (N)	0	0.29	0	0.09	1	0.01
Anderson D. (W)	0	0.26	1	0.00	0	0.51

Brasilia GHI	Observed Series		Normal Distribution fit		Weibull Distribution fit	
	h value	p value	h value	p value	h value	p value
Shapiro-Wilk (N)	0	0.68	0	0.22	1	0.01
Lilliefors (N)	0	0.78	0	0.10	1	0.01
Anderson D. (W)	0	0.30	1	0.00	0	0.29
Boulder DNI	**Observed Series**		**Normal Distribution fit**		**Weibull Distribution fit**	
	h value	**p value**	**h value**	**p value**	**h value**	**p value**
Shapiro-Wilk (N)	0	0.63	0	0.51	1	0.00
Lilliefors (N)	0	0.31	0	0.58	1	0.00
Anderson D. (W)	0	0.92	1	0.00	0	0.74
Boulder GHI	**Observed Series**		**Normal Distribution fit**		**Weibull Distribution fit**	
	h value	**p value**	**h value**	**p value**	**h value**	**p value**
Shapiro-Wilk (N)	0	0.15	0	0.60	1	0.00
Lilliefors (N)	1	0.02	0	0.33	1	0.00
Anderson D. (W)	0	0.53	1	0.00	0	0.70

4. Conclusions

This study yields three primary conclusions:

- The Probability of Exceedance (PoE) at the 50th percentile (PoE50) calculated by fitting observed annual Direct Normal Irradiance series to a normal distribution aligns with the value of the Representative Solar Year (RSY), equivalent to the Typical Meteorological Year (TMY) but accounting solely for DNI. In contrast, when fitting the DNI series to a Weibull distribution, PoE50 results in annual values that are, on average, 0.6% higher than the RSY.
- Assuming that DNI follows a Weibull distribution instead of a Normal distribution leads to extreme scenarios with values that are, on average, up to 2% lower for PoE90 and 7% lower for PoE99. This shift in distribution negatively impacts the financing conditions for stakeholders in solar energy systems.
- We don´t find evidence to support the rejection of the null hypothesis, which posits that the annual DNI follows neither a Weibull nor a Normal distribution. In other words, there is no compelling evidence to suggest that the annual DNI conforms more to a Weibull distribution than to a Normal distribution.

Data availability statement

The datasets generated during and/or analyzed during the current study are not publicly available due to the terms of use signed with the satellite data provider but are available from the corresponding author on reasonable request.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Miguel Larrañeta and Jose López-Álvarez. Algorithm improvement and test was performed by Miguel Larrañeta, Jose M Delgado-Sanchez, and Elena Pérez-Aparicio. The first draft of the manuscript was written by Miguel Larrañeta and Manuel A. Silva Pérez and reviewed by Jose M Delgado-Sanchez, and Joaquín Alonso-Montesinos. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Grant RYC2021-032300-I, funded by the Ministry of Science and Innovation/State Research Agency/10.13039/501100011033 and by the European Union "NextGenerationEU/Recovery, Transformation and Resilience Plan."

Acknowledgement

The authors are grateful to Solargis for providing the satellite-derived databases.

References

1. Fitch Ratings, 2011.Rating Criteria for Solar Power Projects, Utility-Scale Photovoltaic and Concentrating Solar Power. Fitch Ratings.
2. Fernández Peruchena CM, Ramírez L, Silva-Pérez MA, Lara V, Bermejo D, Gastón M, et al. A statistical characterization of the long-term solar resource: Towards risk assessment for solar power projects. Sol Energy 2016; 123: 29–39. https://doi.org/10.1016/j.solener.2015.10.051.
3. Chandra, M., Singpurwalla, N.D., Stephens, M.A., 1981. Kolmogorov statistics for tests of fit for the extreme Value and Weibull distributions. J. Am. Stat. Assoc. 76 (375), 729–731.
4. Liao, M., Shimokawa, T., 1999. A new goodness-of-fit test for type-I extreme-value and 2-parameter Weibull distributions with estimated parameters. J. Stat. Comput. Simul. 64 (1), 23–48.
5. Darling, D.A., 1957. The Kolmogorov–Smirnov, Cramer–Von Mises tests. Ann. Math. Stat., 823–838
6. Vaquera-Huerta, P.P., Villasenor-Alva, J.A., 2009. A Goodness-of-fit test for the Gumbel distribution based on Kullback–Leibler information. Commun. Statistics Theory Methods 38, 842–855.
7. UNE 206013:2017 “Solar thermal electric plants. Procedure for the generation of solar radiation percentiles years”
8. Shapiro, S. S., & Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)." Biometrika, 52(3-4), 591-611.
9. Lilliefors, H. W. (1967). "On the Kolmogorov-Smirnov test for normality with mean and variance unknown." Journal of the American Statistical Association, 62(318), 399-402.
10. Darling, D.A., 1957. The Kolmogorov–Smirnov, Cramer–Von Mises tests. Ann. Math. Stat., 823–838
11. Jiménez-Valero P., Larrañeta M., López-García E., Moreno-Tejera S., Silva-Pérez M.A., Lillo-Bravo I. Synthetic generation of plausible solar years for long-term forecasting of solar radiation (2022) Theoretical and Applied Climatology, 150 (1-2), pp. 649 – 661. https://doi.org/0.1007/s00704-022-04163-9
12. UNE 206011:2014 “Solar Thermal Electric Plants. Procedure for Generating a Representative Solar Year.”